An automated framework for high-throughput predictions of NMR chemical shifts within liquid solutions

Rasha Atwi1, Ying Chen2, Kee Sung Han2, Karl T. Mueller2, Vijayakumar Murugesan2 and Nav Nidhi Rajput1

Identifying stable speciation in multi-component liquid solutions is fundamentally important to areas from electrochemistry to organic chemistry and biomolecular systems. Here we introduce a fully automated, high-throughput computational framework for the accurate prediction of stable species in liquid solutions by computing the nuclear magnetic resonance (NMR) chemical shifts. The framework automatically extracts and categorizes hundreds of thousands of atomic clusters from classical molecular dynamics simulations, identifies the most stable species in solution and calculates their NMR chemical shifts via density functional theory calculations. Additionally, the framework creates a database of computed chemical shifts for liquid solutions across a wide chemical and parameter space. We compare our computational results to experimental measurements for magnesium bis(trifluoromethanesulfonyl)imide Mg(TFSI)2 salt in dimethoxyethane solvent. Our analysis of the Mg2+ solvation structural evolutions reveals key factors that influence the accuracy of NMR chemical shift predictions in liquid solutions. Furthermore, we show how the framework reduces the performance of over 300 13C and 600 1H density functional theory chemical shift predictions to a single submission procedure.

Liquid solutions are critical components of various chemical, materials science, engineering and biological applications, such as batteries1-3, fuel4, the food industry5 and drug discovery6-8. Optimizing the performance of these technologies requires taking into careful account transport and structural features, along with the thermodynamic stability of chemical compounds comprising the solution. More specifically, developing a fundamental understanding of the correlations between functional properties and the underlying atomistic interactions is necessary for advancing the rational design of liquid solutions. In this regard, NMR spectroscopy stands out as a powerful and widespread technique for studying the three-dimensional (3D) organization of matter and associated structural and dynamical properties9,10. Over the years, technological advances in NMR spectroscopy have substantially improved the operational ease and spectral resolutions obtainable from non-traditional nuclei (such as 17O and 25Mg)11. However, NMR spectroscopy is limited by the temporal scale and low sensitivity, making it difficult to speculate structural patterns that are often driven by electrostatic interactions, reactivity, temperature, compositional variance and pressure12,13.

In such complex scenarios, computational NMR studies are necessary to decipher experimental results and better understand different chemical and physical effects whose interplay determines the overall spectrum. For example, ab initio molecular dynamics (AIMD) simulations have been used to capture structural evolutions and associated chemical shifts14,15. However, the computational cost associated with large systems (>100 atoms) and simulation time scales (~10 ps) imposes severe restrictions for tests of liquid solutions across a wide chemical space. Density functional theory (DFT) calculations have also provided valuable insights into chemical shifts even in complex multi-component liquid solutions and endogenous (temperature and pressure) and endogenous (pH and composition) conditions. In addition, gaps in knowledge between systems examined in situ or ex situ and those modelled in silico still exist. For example, NMR DFT studies are often focused on singular phenomena, for example, magnetic shielding tensor. Recently, an automated framework16 and a machine learning approach17 were implemented to predict the 13C/1H NMR chemical shift for organic molecules. However, a generalized approach to identify complex structures in multi-component solutions and accurately predict NMR chemical shift especially for non-traditional nuclei remains a great challenge. On the other hand, NMR experiments can reveal much more information about the chemical system, such as details of chemical exchange, correlation times or energetics for rotational and translational dynamics. Even for the singular focus on chemical shift calculations, the possible molecular structure(s) are traditionally built manually based on chemical intuition, trial and error, and/or results reported in the literature16,18-20. This approach of providing the initial guesses is fraught with bias, is time-consuming, can be challenging to automate fully and leaves behind many persistent metastable configurations of fundamental importance for interpreting experimental results. To overcome these challenges, we designed an automated computational framework that allows accurate prediction of NMR chemical shifts even in complex multi-component liquid solutions and guide experiments to identify stable speciation in solution.

The paper is composed of two sections. First, we discuss the details of our high-fidelity and robust computational tool that seamlessly integrates classical molecular dynamics (CMD) simulations with DFT calculations. Although the developed tool is general enough to be applied to a wide range of applications, we consider magnesium bis(trifluoromethanesulfonyl)imide (Mg(TFSI)2) salt in dimethoxyethane (DME) solvent as an illustrative example.
The chosen electrolyte formulation has received considerable attention in battery literature but reported findings regarding the speciation and the exact solvation structure of the Mg cation are under contention. More specifically, experimental work reported the formation of solvent separated ion pairs (SSIPs), while contact ion pairs (CIPs) were observed in previous computational research16,18,21. The average coordination number between the cation and anion is zero to one for SSIPs and between one and two for CIPs29. We note that we chose a system in which complexities in the solvation phenomena arise due to the multivalent nature of the cation, providing an example to demonstrate that the developed framework can be applied to other simpler systems. We report a detailed comparison between computed and experimental NMR chemical shifts for 25Mg, 13C and 1H nuclei in this electrolyte. We build upon the benchmarking results to elucidate the solvation structures of lithium bis(trifluoromethanesulfonyl)imide (LiTFSI) in DME and correlate the computed 7Li chemical shifts to the experimental ones. We also demonstrate the high-throughput capability of the workflow by accurately predicting more than 300 13C and 600 1H NMR chemical shifts from a set of 100 organic molecules from the spectral database for organic compounds (SDBS)23 and a previous experimental study23. In addition, we use the workflow to fully assign 13C and 1H NMR chemical shifts of different stereoisomers of penam β-lactams, each of which is modelled as a family of conformers. In the second section, we address the fundamental challenge of how to accurately predict NMR chemical shift of liquid solutions by associating the framework with a benchmarking study.

Results

Overview of the automated framework. We construct an NMR computational framework using MISPR (Molecular Informatics for Structure–Property Relationship), our high-throughput and scalable infrastructure that allows automatic handling of thousands of materials science simulations and multiple systems with a strong focus on data provenance. Functionalities of MISPR span from processing and manipulating molecular structures, preparing and executing DFT and CMD simulations on supercomputing resources, parsing and analysing output data, and creating output computational databases. To manage the heterogeneous data that DFT and CMD workflows output and allow for flexible and complex queries, MISPR employs MongoDB for data storage. MongoDB is a document-oriented NoSQL database that stores data as JSON-formatted documents with flexible schema. Force field parameters and derived properties are saved in their collections with auxiliary information like molecular metadata and input parameters, making it easy to reproduce and query computational results.

A unique feature of MISPR is that it allows seamless and automated integration of DFT calculations with CMD simulations to capture structural and dynamical phenomena that span over wide spatial and temporal scales. It contains multiple preset DFT and CMD workflow templates that, from the outside, the user only needs to call in a single Python script with minimal required inputs to generate and run a comprehensive workflow.

The framework designed for automatic NMR chemical shift calculations in liquid solutions is outlined in Fig. 1. It takes as input the structures of molecules comprising a liquid solution of interest (details about supported structure formats are provided in Methods). Next, the framework runs an electrostatic partial charges (ESP) workflow that starts by converting the input structure formats into Gaussian input files. In this work, the ESP workflow used the 6-31+G* level of theory, and tight convergence criteria. It then runs three sequential steps: (1) a DFT geometry optimization, (2) a vibrational frequency calculation to ensure that there are no imaginary frequencies, and (3) a population analysis to assign atomic charges. The framework executes the ESP workflow for each component of the liquid solution.

We note that the framework can be applied to various complex liquid solutions at different conditions (for example, concentration, temperature and pressure). It requires, as a minimum, the

---

**Fig. 1** | Scheme of the computational framework used to calculate NMR chemical shifts in solution as implemented in the MISPR high-throughput infrastructure. Structure of individual components of the liquid solution are first input to an ESP workflow, in which optimization, frequency and partial charges calculations are performed. Optimized structures and computed charges are then fed to a CMD workflow to build the initial system configuration using the input concentrations (C), equilibrate it and derive its structural properties. Atomic clusters of interest are automatically extracted, categorized and sorted based on their frequency of occurrence in solution. Top clusters are then subjected to an NMR workflow, which consists of sequential optimization, frequency and NMR calculation steps. All data collected are saved to a computational database and compared with experimental measurements for error analysis. Examples of optional inputs include the equilibration temperature (T) in CMD simulations and the level of theory in DFT calculations.
Fig. 2 | Structural properties of Mg(TFSI)$_2$ in DME at 298.15 K using FF1 (GAFF), FF2 (non-polarizable oplS) and FF3 (polarizable oplS). a, b, RDF of Mg$^{2+}$–O (DME; a) and Mg$^{2+}$–O (TFSI; b). c, Coordination numbers with Mg$^{2+}$ with the corresponding type of structure: SSIPs, CIPs and AGGs. d, Corresponding types of coordination with oxygen atoms of DME and TFSI$^{-}$.

The framework then uses the generated LAMMPS trajectory files to compute the radial distribution function (RDF) between all possible pairs of particle types in the system or specific pairs specified as inputs. The RDF module is part of a standalone in-house suite of Python tools (MDPropTools) that we developed to extract a range of structural and dynamical properties from LAMMPS trajectory and output files. The RDF defines the probability of finding a particle at a distance $r$ from another particle. More details about the RDF calculations are provided in Methods.

Sampling solvation structures from the CMD step is a key component of the NMR framework. Traditional NMR calculations can be relatively inefficient at constructing initial guesses for molecular structures. Building molecular structures by manually placing a number of molecules in the solvation shell of the particle of interest can be very time consuming. By contrast, our framework passes the computed RDF from the previous step to perform sampling of the first solvation shell of a specific particle in a straightforward and automated manner. In the framework, the first solvation shell is defined by the cut-off distance $r_{\text{min}}$, corresponding to the position of the first minimum after the main peak of the RDF between the particle of interest and other coordinating particles in the solution. Thus, a cluster representing the solvation structure is defined as the group of species within $r_{\text{min}}$ of the particle. By ensemble averaging hundreds of thousands of clusters, we obtain a distribution of clusters corresponding to all the possible chemical environments surrounding the particle in the solution.

Next, the framework categorizes the extracted clusters into unique configurations based on the type and number of species surrounding the particle and their mode of coordination (more details are provided in Methods). Then, it calculates the probability of each configuration as the ratio of the number of clusters that belong to a specific configuration to the total number of extracted clusters. By default, the framework selects the top configurations whose probabilities sum to more than 90% of the total number of extracted clusters, but the user may also select the configurations as needed. The selection of the configurations is done to reduce the number of required DFT calculations and their associated computational cost. It is also important to select a representative cluster from each configuration since it is common that thousands of clusters with subtle geometrical differences (for example, bond lengths and orientation) belong to the same configuration. To this end, the framework performs a local minimization procedure on all the clusters from the selected top configurations using by default the universal force field (UFF) as implemented in the RDKit library. The framework
then feeds the lowest-energy conformer of each configuration to an NMR DFT workflow. Local minimization is only used as an energy evaluation strategy and a criterion to select one cluster representing each unique configuration, since computing all the clusters may be impractical and unnecessary. We note that the DFT calculations discussed later utilized the structures extracted from CMD simulations as starting geometries and not the locally minimized ones.

The NMR workflow relaxes the CMD clusters selected from the previous step, performs a vibrational frequency analysis, and calculates the magnetic shielding tensor on each atom. The framework then performs an analysis step that stores the calculation results in an NMR collection in the database or a local JSON file. Creating a local file allows the user to check outputs quickly, retrieve data without accessing the database and exchange data with other parties. An example of the structure of an NMR document is shown in Supplementary Fig. 1. Finally, results from the computational framework are compared to experimental NMR spectra.

In summary, the framework takes solvent effects into account using both an explicit approach where several solvent molecules surrounding the species are correctly placed in the first solvation shell, and an implicit approach, which approximates the bulk solvent effects using a dielectric continuum model. It is important to note that in the NMR workflow, a series of convergence checks are performed to ensure the results are as reliable as possible. More details of the solvation model, error handler and possible customization of the NMR framework are provided in Methods.

Demonstration of automated DFT calculations. Components of the NMR framework presented in Fig. 1 can be decoupled according to the needs of the user. For example, we used the NMR workflow as a standalone code to compute the $^{13}$C and $^1$H chemical shifts for a set of 100 organic molecules. Detailed information about the library is provided in Supplementary Table 1. The code snippet in Supplementary Fig. 2 demonstrates how to submit these calculations starting from structures defined in the XYZ file format. The workflow generated and managed over 600 input and output files and inserted more than 300 $^{13}$C and 600 $^1$H chemical shifts into the database via a simple one-shot script. We compared our predictions to experimental data from the SDBS$^{22}$ database and a previous study$^{21}$. Supplementary Fig. 3 and Supplementary Fig. 4 show parity plots of the computed chemical shifts and their associated error distribution, respectively. A positive correlation is observed between the workflow output and the experimental data. A further application of the framework is the study of challenging molecules like biologically active natural products and pharmaceutically relevant molecules with complex stereochemistry and flexible conformations. In these cases, elucidating the structure of a molecule is paramount to understanding its biological activity$^{31}$. The penam $\beta$-lactams shown in Supplementary Fig. 5 serve as a contemporary test case to assess the ability of the framework via the DFT workflow to distinguish each of the four stereoisomers using experimental data from Wiitala et. al$^{32}$. Namely, (2$S$,5$S$,6$S$)-, (2$S$,5$R$,6$R$)-, (2$S$,5$S$,6$R$)-methyl 6-(1,3-dioxoisindolin-2-yl)-3,3-dimethyl-7-oxo-4-thia-1-aza bicyclo[3.2.0]heptane-2-carboxylates (1 to 4) were considered. The chosen compound has moderately complex constitution, both in terms of size, functional groups and heteroatoms. For each $\beta$-lactam, we considered multiple conformers characterized by different dihedral angles, resulting in 46 conformers, similar to the approach followed by Wiitala et. al$^{32}$. The smallest mean absolute deviation (MAD) data between experimental and theoretical $^{13}$C and $^1$H NMR chemical shifts (Supplementary Table 2 and Supplementary Table 3) was used as a statistical criterion to assign a true structural match of $\beta$-lactam. The use of B3LYP/6-311+G(2d, p) leads to correct assignment of stereoisomers with $^1$H shifts, while oB97X$^+$/def2-TZVP incorrectly matches 4 with 2. The $^{13}$C MAD comparisons indicate that both levels of theory fail to distinguish between 1, 2 and 3 penam esters. The 3D structure, Gibb's free energy, and NMR chemical shifts of the conformers are provided in the accompanying dataset.

Factors affecting the accuracy of NMR chemical shifts. Reliably differentiating among different extracted solvation structures requires highly accurate NMR chemical shifts. The successful implementation of our framework necessitates adequate consideration of several important factors. First, a key question for the CMD component is the quality of the interatomic potentials since significant deviations in system properties have often been observed compared to experimental data$^{33}$. Second, the DFT level of theory comprising the density functional and basis set is critical for achieving well-converged chemical shieldings. In addition, the choice of the implicit solvent model is crucial for approximating the bulk solvent effect. A remarkable number of benchmarking studies have been done on quantum mechanical methods for predicting properties in complex multi-component battery electrolytes similar to the
test case here\textsuperscript{36,37}. However, parallel studies for NMR calculations for these systems are still in their infancy. Other factors include selecting an appropriate number of molecules in the chemical reference to account for intermolecular interactions and a representative conformer from each solvation environment. In the following sections, we report the role of these factors using results obtained by the framework for the Mg(TFSI)\textsubscript{2}/DME test case system.

**Role of the force field.** We benchmark the most commonly used and reliable force fields for liquid solutions, including GAFF (FF1), non-polarizable OPLS\textsuperscript{38} (FF2) and polarizable OPLS (FF3) force fields. FF1 and FF2 are computationally less expensive due to their non-polarizable nature and have been used extensively in the battery literature showing satisfactory agreement with experimental findings. FF3, built on top of FF2, allows for a polarizable response of molecules to an electric field using the Drude oscillator model\textsuperscript{39}. In this model, particles are added to each polarizable atom to mimic physical dipoles and model the corresponding distortion of electron density.

The simulation density ($\rho$), shown in Supplementary Table 4, agrees well with the experimental value. The lowest average error (1.2\%) is achieved with FF3. The RDFs between the cation and oxygen atoms of DME and TFSI\textsuperscript{−} are shown in Fig. 2a,b. FF1 results in the weakest cation–solvent [Mg\textsuperscript{2+}–DME] interaction and the most vital cation–anion [Mg\textsuperscript{2+}–TFSI\textsuperscript{−}] interaction, as evident from the sharp RDF peak between the cation and oxygen atoms of the anion. On the other hand, with FF3, little coordination occurs with the anion (inset of Fig. 2b), indicating that solvent molecules dominate the first solvation shell of the cation. Figure 2c shows Mg\textsuperscript{2+}–O (DME) and Mg\textsuperscript{2+}–O (TFSI\textsuperscript{−}) coordination numbers, calculated by integrating the corresponding RDF curves for the first solvation shell. FF3 results indicate that Mg(TFSI\textsubscript{2}) tends to form SSIPs in DME, while FF2 shows that the salt participates in forming CIPs. On the other extreme, FF1 results in the formation of aggregate solvates (AGGs), in which two or more anions coordinate with the cation. An example of the type of coordination represented by the RDFs is displayed in Fig. 2d. The tested force fields also result in different percentages of DME and TFSI\textsuperscript{−} that coordinate to Mg\textsuperscript{2+} with two oxygen atoms, that is, in bidentate configuration, as shown in Supplementary Fig. 6.

The top Mg\textsuperscript{2+} configurations identified by the framework are provided in Supplementary Figs. 7 and 8. Overall, we find substantial differences in the type and distribution of these structures among the tested force fields. For example, the most probable solvation structure predicted with FF1 involves one DME in bidentate configuration and four TFSI\textsuperscript{−} anions in monodentate configuration. In addition, rather than forming a single stable solvate like with FF3, the distribution of coordination environments for the cation with FF1 is much more heterogeneous and involves configurations dominated by the anion. With FF2, the electrostatic interaction with the anion is slightly suppressed, and the most probable solvation shell is composed of two DME solvates and TFSI\textsuperscript{−} anions participating in bidentate and monodentate configurations, respectively. FF3 results in a Mg\textsuperscript{2+} solvation shell dominated by three DME molecules participating in bidentate configuration. This configuration has been previously suggested based on experimental measurements of diffusion and Raman and NMR spectroscopy\textsuperscript{10,21}, and computationally by Kubisiak and Eilm\textsuperscript{39}e for a concentration range of 0.1–1 M.

Variations in the structural properties between the force fields are translated to the dynamical behaviour of the electrolyte. The distribution of diffusion coefficients (Supplementary Fig. 9) from FF3 indicates 2.58 slow DME molecules per Mg\textsuperscript{2+} cation. This result is in close agreement with the experimentally measured value of 3.0\textsuperscript{14} and is consistent with the computed structural properties. FF1 and FF2 predict 1.17 and 1.86 slow DME molecules per Mg\textsuperscript{2+}, respectively. The calculated ionic diffusion coefficients with FF3 are also in better agreement with experimental results (mean absolute error of 20\%), whereas those from FF1 and FF2 are underestimated by approximately 90\% and 30\%, respectively (Supplementary Fig. 10). More details of the dynamical analysis are provided in Methods and Supplementary Section 1.

The discrepancies in the predicted properties are not particularly a problem of a specific force field or the Mg(TFSI)\textsubscript{2}/DME system, but rather due to a lack of accounting for the critical interactions in the non-polarizable simulations. The predicted properties using FF3 are the most consistent with previous experimental\textsuperscript{11,15} and computational\textsuperscript{40} studies among the tested force fields. However, the better performance of FF3 comes at the expense of its two- to threefold higher computational time compared to FF1 and FF2. Evaluating the quality of the force field used in the sampling process is shown to be a necessary primary step to obtain reliable structures for NMR computations. Here, we proceed with the FF3-predicted solvation structures to report results from the DFT component of the NMR framework.

**Role of the DFT level of theory.** We evaluate the performance of selected DFT functionals and basis sets in predicting chemical shifts of $^{25}$Mg, $^{13}$C, and $^{1}$H of the top configurations and the chemical shifts of $^{13}$C and $^{1}$H resonances in the bulk. The NMR framework (Fig. 1) is designed to be used in high-throughput mode to study speciation evolution in liquid solutions at variable conditions, for example, concentration and temperature. Therefore, the comparison is made not only based on accuracy but also on factors that are particularly important for high-throughput simulations (for example, computational cost and tendency to fail).

During the benchmark study, the most common failures encountered include failure to converge the geometry to a potential energy surface (PES) minimum in a finite number of optimization steps, difficulties in converging self-consistent field (SCF) calculations, and errors in internal coordinate transformations. Around 78\% of the total performed calculations were completed without error-correction procedures. Levels of theory primarily involved in the failed calculations include B3LYP/6-31+G* and PBE1PBE\textsuperscript{41} hybrid functional coupled with each of the 6-31+G* and 6-311++G** basis sets. The automatic error handler corrected around 86.3\% of the failed calculations, leading to an overall success rate of 97\%.

The $^{25}$Mg NMR results from the top-performing level of theory (oB97X/def2-TZVP) are shown in Fig. 3 along with the corresponding structure of the predicted species. A single broad peak is observed, indicating either a single solvation structure or a convolution of multiple structures with rapid exchange. The predicted $^{25}$Mg...
chemical shift in the most probable configuration is \(-0.809\) ppm, which is highly consistent with the experimental peak centre at \(-0.71\) ppm. Given the broad line width of the \(^{25}\text{Mg}\) peak, that is, the half peak height at 0.83 and \(-2.13\) ppm, the chemical shift of \(^{25}\text{Mg}\) in configuration 2 (Supplementary Table 5) is also considered to be in satisfactory agreement with experimental data. Therefore, multiple \(^{25}\text{Mg}\) structures that are entirely dissociated from the anion are possible in the solution. Excluding configuration 4, the increase in the ion–dipole interaction between \(^{25}\text{Mg}\) and TFSI\(^{-}\) in the following order: configuration 1 < configuration 2 < configuration 3 < configuration 6 < configuration 5 leads to the observed monotonic upfield shift in the corresponding \(^{25}\text{Mg}\) chemical shift. The presence of loosely packed clusters of [Mg(DME)]\(_n\) (n \(\leq 2\)), that is, configuration 4, is attributed to the high degree of freedom and structural flexibility of DME. This type of configuration has been reported to be favourable at lower concentrations due to lower electrostriction (reduced solvent volume in the \(^{25}\text{Mg}\) solution shell relative to the bulk) and diminished entropy loss\(^{15}\). On the contrary, higher concentrations (0.51 M) such as the one used here lead to closer distances between Mg\(^{2+}\) ions, resulting in stronger electrostatic interactions and dampened DME motion, thus favouring fully solvated clusters (n = 3, configuration 1). This behaviour is consistent with the low probability of configuration 4 and the predicted \(^{25}\text{Mg}\) chemical shift of this configuration, which is far from the experimental peak centre (Fig. 3). The benchmarking results for \(^{25}\text{Mg}\) chemical shift calculations are displayed in Supplementary Fig. 11.

On average, going from left to right moves most of the predicted chemical shifts within the bounds of the observed NMR spectrum. From top to bottom, significant variations are observed using the four functionals with 6-31+G*, while this difference is less clear with def2-TZVP. Different levels of theory can also lead to contradictory conclusions regarding the dominant species in solution.

Figure 4 shows \(^{13}\text{C}\) NMR shifts assigned to CH3 of DME existing in the bulk solution (labelled ‘free CH3’) and DME coordinated to Mg\(^{2+}\) (labelled ‘bound CH3’) from DFT predictions and experimental measurements. Similar plots for \(^{13}\text{C}\) shifts assigned to CH2 and \(^{1}\text{H}\) shifts assigned to CH3 and CH2 of both types of DME molecules are shown in Supplementary Figs. 12–14, respectively. While free and bound DME molecules are distinguishable from experimental and predicted \(^{13}\text{C}\) and \(^{1}\text{H}\) NMR chemical shifts, it is impossible to differentiate between bound DME at different configurations. The spectroscopic differences between the structures may be subtle (see, for example, Supplementary Table 5 for \(^{13}\text{C}\) and \(^{1}\text{H}\) chemical shifts in different configurations). On the contrary, \(^{25}\text{Mg}\) chemical shifts can be utilized for this purpose, whereby changes in charge density localization on different Mg\(^{2+}\) complexes directly alter the screening effects experienced by the \(^{25}\text{Mg}\) nucleus. As displayed in Fig. 4 and Supplementary Fig. 12, the highest deviation from experimental \(^{13}\text{C}\) shifts are obtained with 6-31+G* and 6-311++G** combined with any of the tested density functionals. The basis set from the ‘def2’ family of Alrichs and co-workers\(^{15}\), particularly in combination with \(\omega\)B97X, leads to \(^{13}\text{C}\) NMR chemical shift error that approaches the underlying uncertainty in experimental measurements (Supplementary Table 5). Supplementary Fig. 13 and Supplementary Fig. 14 indicate that for \(^{1}\text{H}\) chemical shifts, M06-2X/def2-TZVP outperforms the other levels of theory with absolute errors between 0.01 and 0.3 ppm (Supplementary Table 5).

We conclude that the choice of the basis set has the highest impact on the accuracy of NMR chemical shift predictions. The 6-31+G* basis set is ruled out as a suitable basis set for NMR calculations of complexes similar to those studied herein due to its degraded accuracy compared to other basis sets, despite its lower computational cost (see Supplementary Fig. 15 for timings). For \(^{25}\text{Mg}\) and \(^{13}\text{C}\) chemical shifts, the \(\omega\)B97X/def2-TZVP level of theory is recommended if computational resources are available. If computational resources are limited, M06-2X/6-311++G** is recommended for \(^{25}\text{Mg}\) shifts as its cost is not prohibitive while still predicting the correct Mg\(^{2+}\) solvation structure. Finally, M06-2X with def2-TZVP or 6-311++G** are recommended for \(^{1}\text{H}\) chemical shift predictions.

Role of conformer. Another consideration in the NMR framework is that it utilizes the lowest-energy conformer from each configuration to initialize the NMR DFT calculations. Previous NMR DFT studies have also reported findings on possible solvation structures based on a single conformer\(^{22,31,34}\). However, the measured shift is the weighted average of chemical shifts of all possible conformers in solution during the NMR acquisition time. The plots in Fig. 5 show the mean difference (including the 95% confidence interval) between the Boltzmann average NMR chemical shift for the entire ensemble, \(\delta\)\(_i\), and our initial chemical shift estimation, \(\delta^*\)\(_i\), as a function of the number of optimized conformers (\(i\)). Details of how \(\delta^*\)\(_i\) is computed are provided in Methods. Variable degrees of errors are obtained with each nucleus type, with the highest difference in the \(^{25}\text{Mg}\) chemical shift. In this electrolyte system, a maximum unsigned error of 1.2 ppm in the \(^{25}\text{Mg}\) chemical shift of configuration 1 upon excluding conformational sampling does not alter the interpretation of experimental findings in terms of the most probable solvation structure while saving 15\% computational resources.
Fig. 6 | Structural analysis of 1:18 Li(TFSI) in DME solution at 298.15 K. a. RDF between Li–O (DME) and Li–O (TFSI−). b. Predicted 7Li NMR chemical shifts for the top six Li+ configurations in the solution using the NMR computational protocol and the experimental NMR spectrum; the computed frequency of occurrence of each configuration is shown on the right y axis. c. The corresponding structure of the top configurations with the type of coordinating atoms are located underneath. DFT calculations are performed at the ωB97X/det2-TZVP level of theory using the PCM solvation model.

resources. Nevertheless, conformational sampling has a more pronounced impact on other less probable solvation structures like configuration 2, for which an error of 4.4 ppm is incurred if only the UFF global minimum is considered for calculations at the higher level of theory. Significantly lower errors are obtained for 13C chemical shifts of CH2 and CH3 groups, while 1H chemical shifts are insensitive to conformer sampling. Therefore, an evaluation of the impact of conformational sampling on DFT predictions should be done whenever possible to boost the confidence in NMR chemical shift predictions made by the framework.

Finally, results of the effect of geometry optimization, choice of chemical reference for 25Mg, 13C and 1H chemical shifts, and implicit solvent model are summarized in Supplementary Sections 2–4.

LiTFSI in DME test case. The NMR framework was utilized to elucidate the solvation structure of lithium bis(trifluoromethanesulfonyl) imide (LiTFSI) in DME at 298.15 K and a salt-to-solvent ratio of 1:18. RDF plots, 7Li experimental NMR spectrum and computational shifts, and the top six configurations identified from CMD simulations are displayed in Fig. 6. As evident from the Li–O (DME) and Li–O (TFSI−) RDF plots in Fig. 6a and the top configurations in Fig. 6c, the solvation structure of Li+ is dominated by solvent molecules. Compared to Mg2+, the first well-defined Li+ RDF peak occurs at a shorter distance, an observation that is consistent with previously reported experimental X-ray total scattering data25. This behaviour is attributed to the higher charge density of the divalent cation (Mg2+) that can contribute to restricted conformational flexibility and dynamics of the coordinating molecules. In configurations, 1, 2 and 4, Li+ is coordinated by four to six oxygen atoms originating from a total of three DME molecules, unlike Mg2+ which is mostly coordinated by six atoms with the high frequency of occurrence of Mg2+ in configuration 1. Configuration 5 includes three DME molecules, only two of which are contributing oxygen atoms towards Li+ coordination. Optimizing this structure leads to Li+ coordination similar to the one in configuration 1. In configuration 6, Li+ is coordinated by one TFSI− anion and two DME molecules in monodentate and bidentate configurations, respectively. However, coordination by TFSI− occurs at a much lower probability compared to the Mg(TFSI)2 electrolyte. The lack of contact ion pairs and aggregates in this solution could possibly explain the reported maximum in the ionic conductivity of the electrolyte as a function of concentration26.

The predicted 7Li chemical shift in configuration 1 is −1.215 ppm, which is highly consistent with the experimental peak centre located at −1.193 ppm. The 7Li shifts corresponding to configurations 2, 4 and 5 (−1.338, −1.158 and −1.176 ppm, respectively) are also in reasonable agreement with the experimental data, supporting the possibility of the presence of these structures in solution. Given the farther location of the 7Li shift from the peak centre for configuration 3 where Li+ is coordinated by two DME molecules each in bidentate configuration could correspond to a transition structure. The narrower 7Li peak compared to that of 25Mg (full width at half maximum of 0.026 ppm versus 2.961 ppm, respectively) can be attributed to the faster molecular exchange in the former...
electrolyte between molecules in the first and second solvation shells with those in the bulk phase. This behaviour is consistent with the higher average diffusion coefficient of DME in the Li+ -based electrolyte (1.24 × 10^−3 m^2 s^−1) compared to the Mg2+ -based electrolyte (0.746 × 10^−3 m^2 s^−1).

Discussion

We have developed and tested a computational framework that couples first-principle calculations with CMD simulations to robustly and efficiently calculate, analyse and store NMR chemical shifts from a variety of molecules in liquid solutions. The framework overcomes limitations in current NMR computational studies such as the Edislonian approach in selecting possible solvation structures and the substantial time required for manual file management, data collection and error handling. As with all computational frameworks, the current NMR framework is not without its own limitations. Though the computational framework is flexible to include any cut-off distance for the solvation shell, in the current work we explored the first solvation shell. The second solvation shell can play a critical role in chemical shift assignment especially in systems with long-range interactions and fast dynamics. In addition, the framework does not reduce the substantial computational resources required to calculate highly accurate chemical shifts. It rather eliminates human intervention, making it much more convenient to study wider chemical and parameter spaces. Special attention is also needed to ensure appropriate parameters (for example, interatomic potentials and DFT level of theory) are used for a specific system. This can be achieved by conducting a benchmarking study similar to the one presented here prior to the actual calculations.

Future directions will include the second solvation shell and coupling this strategy with a more detailed analysis of the exchange dynamics in the solution. In addition, support for performing automated polarizable CMD simulations using the thermalized Drude dipole method as implemented in LAMMPS will be added. The current and extended framework will be used to study other complex multi-component liquid solutions such as monovalent and multivalent electrolytes whose structure is not intuitive or when the chemical and parameter spaces are too large for human search using conventional methods. Data collected from the framework is expected to provide fingerprints to guide future experimental investigations of liquid solutions with optimal properties.

Methods

MISPR dependencies. We built MISPR on top of base libraries developed by the Materials Project, namely (1) pymatgen for structure representation and input/output files generation and handling, (2) FireWorks for managing workflows over computing resources, and (3) custodian for monitoring inevitable errors during simulations and applying on-the-fly fixes. At the backend, MISPR uses Gaussian electronic structure software for DFT calculations and LAMMPS (https://www.lammps.org/) open-source code for CMD simulations.

Supported structure formats. Many formats can be used to input structures in the NMR framework as treated by the project. For example, the framework supports XYZ, pymatgen molecule objects, Gaussian outputs and so on, via the OpenBabel and pymatgen libraries. In these formats, the framework can take query criteria to retrieve previously optimized structures from the database. It can also derive a structure on the fly by either attaching a functional group or linking two structures at a specific binding site.

Default operation of the NMR framework. The framework bypasses the ESP workflow if the ESP charges have been previously calculated or other force fields are directly provided either in a dictionary or in a database of force field parameters. The default CMD configuration involves four sequential steps: (1) energy minimization, (2) isothermal–isobaric (NPT) equilibration, (3) melting and quenching, and (4) canonical (NVT) production runs. However, the user may alter these set of protocols to run any series of LAMMPS calculations according to their own project needs. In addition, many of the input parameters to the NMR framework are optional and default values will be used if these parameters are not specified. For example, the length of the RDF and the bin size take default values of half the box length and 0.05 Å, respectively, but they can also be adjusted as needed. Our implementation is also flexible to compute different types of RDF; namely the atomic and molecular RDF as well as the RDF between an atom and the centre of mass of a molecule. By default, the framework computes the atomic RDF between all possible pairs of atoms in the solution, but this may be overridden by setting the appropriate submission parameters. In addition, the cut-off radius, rmax, is automatically extracted from the RDF, but the user may override this by providing rmax as an optional input. In the cluster analysis, the framework selects the top configurations whose probabilities sum to more than 90% of the total number of extracted clusters to use in the NMR calculations. However, the user may override this percentage or request to consider configurations whose individual percentage is above a certain threshold specified as input. The local minimization step performed using RDKit before optimizing the clusters at the higher level of theory uses by default the UFF, but it can also be done using any type of force field supported by RDKit. For the NMR DFT calculations, the framework gives the user the flexibility to skip the optimization step if a qualitative analysis is sufficient. This can be especially justified in cases where there is a systematic offset in the NMR chemical shifts like the case of Mg(TF3), in DME (Supplementary Fig. 16). The framework by default uses the b97Xd defer-TZVP level of theory for performing the sequential optimization, frequency, and NMR steps. Switching the functional, basis set and other Gaussian input parameters (for example, solvation model, numerical and algorithmic parameters) is straightforward and requires the user to input them in the form of a Python dictionary.

RDF analysis. The RDF is calculated by binning distances between two particles while taking periodic boundary conditions into account. The code to calculate the RDF uses the following equation:

\[
g_r(r) = \frac{V}{N_i N_j} \sum d(r_r - r) \delta(r_r - r)
\]

where \( g_r(r) \) is the RDF between particle types \( i \) and \( j \), \( V \) is the volume, \( N_i \) and \( N_j \) are the number of particles of type \( i \) and \( j \), respectively, \( r \) is the distance between the particles, and \( d \) is the width of the bins.

Diffusion analysis. The self-diffusion coefficient \( D \) for each species is calculated in MDPropTools using the Einstein relation:

\[
D_i = \lim_{t \to \infty} \frac{\text{MSD}(t)}{6t}
\]

where MSD(t) is the mean-square displacement of each species as a function of time in the diffusive regime, obtained from the production run.

Cluster analysis. For each extracted cluster, the framework identifies all neighbouring atoms to the site of interest within a sphere of radius, \( r_{max} \). It uses the structure of the individual molecules used to build the initial mixture to match their site pattern (for example, atom order) to the ones belonging to the cluster. Based on that, it then determines the parent species to which each neighbouring atom belongs. For example, if the mixture is composed of molecules of type A and type B, the framework compares each consecutive sequence of atomic sites in the cluster to those in A and B. If a match with A is found, it assigns the group of atoms to type A, otherwise, it continues the comparison until a match is identified. After all neighbouring atoms have been assigned to a parent species, the framework stores the information about the type and number of each component in the solvation shell as well as those of the coordinating atoms in a dataframe format. When all the clusters are analysed, the framework groups them into unique configurations based on the type and number of species surrounding the particle and their mode of coordination. For example, if a site of interest in one cluster is coordinated by two oxygen atoms from one DME molecule, while the same site in another cluster is coordinated by two oxygen atoms from two different DME molecules, the two clusters are considered different. Then, the framework calculates the probability of each configuration as the ratio of the number of clusters that belong to a specific configuration to the total number of extracted clusters. Configurations with the highest probability of occurrence correspond to persistent metastable solvation structures in the solution.

Solvation model. The framework allows computing NMR chemical shifts using an explicit, implicit or a mixed explicit/implicit solvation model. The user can select the solvation model by specifying the corresponding input parameters, for example, the solvent type and solvation model. If these parameters are not provided, implicit solvation is not accounted for in the DFT calculations. Using a mixed solvation model allows incorporating a thermodynamically stable and realistic chemical environment of species compared to the traditional approach, which relies on either implicit solvent models or manual prediction of the possible solvation structures. Since multiple configurations are considered, collected data result in various chemical shifts corresponding to different chemical environments experienced by the nucleus of interest. Therefore, predictions from this approach can be compared and fitted to the entire experimental NMR peak rather than...
just matching the peak centre, especially when peak broadening occurs due to intermediate exchange dynamics in solutions.

Error handling. In the NMR workflow, a series of convergence checks are performed to ensure the results are as reliable as possible. For example, we implemented checks for normal termination of DFT calculations and automatic inspection of the 3D structure resulting from optimization to confirm connectivity matches the input structure. Once each step of the NMR workflow has terminated, the output file is parsed for errors. An automatic error-correction procedure applies the rate of the calculation without relying on human intervention, which would be very difficult for handling large computational investigations. Examples of the errors addressed are SCF failure, geometry optimization convergence, error in internal coordinates and exceeded wall time limit. Note that in case of unstable geometries that are challenging to converge, the frameworks are the structures at a different level of theory to provide a possibly better initial guess before repeating the calculation. In some non-trivial number of cases, this error-correction procedure may still fail, in which case a more thorough human intervention becomes necessary.

DFT calculations. All DFT calculations are performed using Gaussian 16 Rev. C.01. NMR calculations for the organic molecules are performed in a chloroform solvent at the B3LYP/6-31+G(d,p) level of theory59 referred to the tetraphenylcyclophane (TMS). Our criteria for building the library of compounds included the availability of their experimental 1H and 13C chemical shifts in a chloroform solvent and their reference to TMS at room temperature. We also selected compounds that span a broad array of chemical classes and distribution of molecular sizes. The chemical shifts for β-lactam penam compounds were optimized in chloroform using the polarizable continuum model (PCM) solvation model, and TMS is used as a reference for the chemical shift of the B3LYP/6-311+G(2d,p) and ωB97X/def2-TZVP levels of theory, respectively. The number of conformers used for penam β-lactams was 12 for structure 1, 10 for structure 2, 12 for structure 3 and 12 for structure 4. Boltzmann averaging was performed for each family to determine the fractional population of the conformers based on their energy. The 46 conformers were optimized at each level of theory prior to NMR calculations. We note that no linear corrections were applied to the computed chemical shifts.

Magnetic shieldings are calculated for the extracted Li+ and Mg2+ clusters, ranging in size from 33 to 78 atoms. The benchmark study is performed with the conformations of four functional groups (B3LYP, M06-2X, PBE1PBE and ωB97X) and three basis sets (6-31+G, 6-311+G* and def2-TZVP), resulting in twelve combinations chosen due to their broad application in the NMR literature. An ultrafine integration grid is employed, and van der Waals interactions are treated using Grimme dispersion correction (D3)59 with the B3LYP, M06-2X and PBE1PBE methods. To extrapolate the explicit solvent model used in this work, bulk solvent effects are described using a continuum model, particularly the PCM59 or the solvation model based on density (SMD)60 in tetrahydrofuran. Following the optimization and frequency steps, magnetic response calculations are performed using the gauge-independent atomic orbital (GIAO)61 method at the same level of theory. The Li chemical shift calculations for the top six conformations are performed at ωB97X/def2-TZVP using the PCM solvation model in tetrahydrofuran to model bulk solvent effects. Chemical shifts are converted to the experimentally observed scale using δobs = σobs - σTMS where δobs and σobs are the chemical shift and the isotropic shielding constant of the nucleus of interest in a given cluster, respectively, and σTMS is the calculated isotropic shielding constant of the same nucleus in a suitable reference compound. We use an Li+ ion coordinated by five water molecules, Mg2+ ion coordinated octahedrally by six water molecules, dimethyl sulfoxide and water, as the chemical references for Li+ , Mg2+ , 1H and 13C, respectively. To reduce systematic errors, we use secondary references (TMS) by Aqvist62 and TMS is used as a reference to TMS at room temperature. We also selected compounds that span a broad array of chemical classes and distribution of molecular sizes.

DFT calculations. All DFT calculations are performed using Gaussian 16 Rev. C.01. NMR calculations for the organic molecules are performed in a chloroform solvent at the B3LYP/6-31+G(d,p) level of theory59 referred to the tetraphenylcyclophane (TMS). Our criteria for building the library of compounds included the availability of their experimental 1H and 13C chemical shifts in a chloroform solvent and their reference to TMS at room temperature. We also selected compounds that span a broad array of chemical classes and distribution of molecular sizes. The chemical shifts for β-lactam penam compounds were optimized in chloroform using the polarizable continuum model (PCM) solvation model, and TMS is used as a reference for the chemical shift of the B3LYP/6-311+G(2d,p) and ωB97X/def2-TZVP levels of theory, respectively. The number of conformers used for penam β-lactams was 12 for structure 1, 10 for structure 2, 12 for structure 3 and 12 for structure 4. Boltzmann averaging was performed for each family to determine the fractional population of the conformers based on their energy. The 46 conformers were optimized at each level of theory prior to NMR calculations. We note that no linear corrections were applied to the computed chemical shifts.

Magnetic shieldings are calculated for the extracted Li+ and Mg2+ clusters, ranging in size from 33 to 78 atoms. The benchmark study is performed with the conformations of four functional groups (B3LYP, M06-2X, PBE1PBE and ωB97X) and three basis sets (6-31+G, 6-311+G* and def2-TZVP), resulting in twelve combinations chosen due to their broad application in the NMR literature. An ultrafine integration grid is employed, and van der Waals interactions are treated using Grimme dispersion correction (D3)59 with the B3LYP, M06-2X and PBE1PBE methods. To extrapolate the explicit solvent model used in this work, bulk solvent effects are described using a continuum model, particularly the PCM59 or the solvation model based on density (SMD)60 in tetrahydrofuran. Following the optimization and frequency steps, magnetic response calculations are performed using the gauge-independent atomic orbital (GIAO)61 method at the same level of theory. The Li chemical shift calculations for the top six conformations are performed at ωB97X/def2-TZVP using the PCM solvation model in tetrahydrofuran to model bulk solvent effects. Chemical shifts are converted to the experimentally observed scale using δobs = σobs - σTMS where δobs and σobs are the chemical shift and the isotropic shielding constant of the nucleus of interest in a given cluster, respectively, and σTMS is the calculated isotropic shielding constant of the same nucleus in a suitable reference compound. We use an Li+ ion coordinated by five water molecules, Mg2+ ion coordinated octahedrally by six water molecules, dimethyl sulfoxide and water, as the chemical references for Li+ , Mg2+ , 1H and 13C, respectively. To reduce systematic errors, we use secondary references (TMS) by Aqvist62 and TMS is used as a reference to TMS at room temperature. We also selected compounds that span a broad array of chemical classes and distribution of molecular sizes.

Conformer sampling. We assess the sensitivity of DFT chemical shifts to conformer sampling by starting from the UFF energies of a total of ~270,000 conformers of conformer 1 and ~4,000 conformers of conformer 2 extracted from CMD simulations. From each configuration, 15 conformers spanning the entire energy range are selected to initialize full NMR calculations that include geometry optimization, frequency and chemical shift estimation at the ωB97X/ def2-TZVP level of theory using the PCM solvation model. Boltzmann averaging is done according to the equation shown in Fig. 5 to calculate the ensemble NMR chemical shift. Because optimization at the higher level of theory leads to the reordering of conformational energies, the results are reported relative to the UFF global minimum energy conformer. When computing δC, we considered all the possible combinations of i conformers from a pool of 15 structures, with a restriction that the UFF global minimum energy conformer is included in these combinations. The result is a chemical shift distribution of shifts for i, from which the mean is computed and shown in Fig. 5. To maintain statistical significance, we excluded i = 2, 14, 15 from this analysis, as each of them results in less than 30 possible conformer combinations.

Experimental methods. Mg(TFSI)2 and LiTFSI (99.5%, Solvionic) were dried for 48 hours under vacuum at 180°C and the DME solvent (Battery-grade, Cotson) was further dried over activated 3 Å molecular sieves in a glovebox until its water content was determined to be below 10 ppm using a Karl-Fisher Titirator (Metrohm). Mg(TFSI)2/DME and LiTFSI/DME solutions were prepared inside a glovebox filled with nitrogen right before NMR measurements. The 1H and 13C NMR measurements were performed on a Varian DD2 spectrometer with a 17.0 T magnet and a broad-band (BBO) probe with 1H and 13C Larmor frequencies of 748.1 and 188.1 MHz, respectively. The 90° pulse widths were 16µs for 1H and 16µs for 13C. The 1H spectra were collected using 30 µs pulses with a transition number of 16 and a recycle delay of 20s with a coaxial tube holding Mg(TFSI)2/DME solution and an outer NMR tube holding D,O (99.9%, from Sigma-Aldrich) as an external reference at 4.77 ppm. The 13C spectra were collected using 30 µs pulses with averaging of 1,024 transients and a recycle delay of 12s using a thin-wall 5 mm NMR tube. The 1Mg and 1Li NMR spectra were collected on a 14.1 T magnet (Varian DDR spectrometer) with a 2Mg Larmor frequency of 36.7 MHz and a 1Li Larmor frequency of 232.94 MHz. The 90° pulse width was 20µs for 1Mg and 9µs for 1Li. For 1Mg, a small tip angle of 15° with a recycle delay of 5.1 s was used and 128,000 transients were acquired; for 1Li, 32 transients were acquired with a recycle delay of 1s and a tip angle of 30°. In order to minimize the spectrometer drift effect on chemical shift, DMSO-d6, 5M MgCl2, and 1 M LiCl were used to reference 1H at 39.52 ppm, 1Mg (0 ppm) and 1Li (0 ppm) respectively, right before each NMR measurement.

Data availability. The dataset used to generate the results in this work along with the optimized 3D structures in XYZ format and initial and final MD trajectories are available at GitHub (https://github.com/rashatwi/nmr-dataset) and Zenodo68. Source data are provided with this paper.
Harks, P., Mulder, F. & Notten, P. In situ methods for Li-ion battery research: A review of recent developments. *J. Phys. Chem. C*, 117, 7002–7014 (2013).

Veselý, F., Kratsch, S., Usov, A., Rollema, H. & Cerezo, A. H and 13C high resolution NMR spectroscopy of carragenans: application in research and industry. *Trends Food Sci. Technol.* 13, 73–92 (2002).

Shimada, I., Ueda, T., Kofuku, Y., Eddy, M. T. & Wuthrich, K. GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures. *Nat. Rev. Drug Discov.* 10, 3360 (2011).

Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236 (1996).

Lamoureux, G. & Roux, B. Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* 119, 3025–3039 (2003).

Kubiak, P. & Elimes, A. Solvation of Mg2+ ions in Mg(TFSI)2–dimethoxyethane electrolytes—a view from molecular dynamics simulations. *J. Phys. Chem. C* 122, 12615–12628 (2018).

Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* 110, 6158–6170 (1999).

Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305 (2005).

Zhang, F. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochromic kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 120, 215–241 (2008).

Andersen, A. et al. Structure and dynamics of polysulfide clusters in a nonaqueous solvent mixture of 1,3-dioxolane and 1,2-dimethoxyethane. *J. Phys. Chem. A* 121, 2308–2319 (2019).

Lapidus, S. H. et al. Solvation structure and energetics of electrolytes for multivalent energy storage. *Phys. Chem. Chem. Phys.* 16, 21941–21945 (2014).

Liyanara-Ararichi, T. P. et al. Polarizable molecular dynamics and experiments of 1,2-dimethoxyethane electrolytes with lithium and sodium salts: structure and transport properties. *J. Phys. Chem. B* 122, 8548–8559 (2018).

Iruru, K. Electrochemistry in Nonaqueous Solutions (Wiley, 2009).

Jain, A. et al. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency Computat. Pract. Experience* 27, 5037–5059 (2015).

O’Boyle, N. M. et al. Open Babel: an open chemical toolbox. *J. Cheminformatics* 3, 33 (2011).

Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132, 154104 (2010).

Miertuš, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a solvent, defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396 (2009).

Wolsin, K., Hinton, J. F. & Polanyi, P. R. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* 112, 8251–8260 (1990).

Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* 30, 2157–2164 (2009).

Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Atomic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Modell.* 25, 247–260 (2006).
Acknowledgements
High-performance computational resources for this research were provided by the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges and Comet computational resources (charge number: TG-DMR 190087), which is supported by National Science Foundation (NSF) grant number ACI-1548562. This work also used computational resources at the Stony Brook Institute for Advanced Computational Science (iACS). N.N.R. was supported by the startup funds from Stony Brook University. N.N.R. and R.A. acknowledge the startup funds from Tufts University. We acknowledge the iACS Junior Research Award to R.A. Experimental NMR research work was supported as part of the Joint Center for Energy Storage Research, an Energy Innovation Hub funded by the US Department of Energy, Office of Science, Basic Energy Sciences. The NMR experiments were performed using EMSL (grid.436923.9, Y.C., K.S.H., V.M., K.T.M.), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. PNNL is a multi-programme national laboratory operated for the DOE by Battelle Memorial Institute under contract DE-AC06-76RLO 1830 (Y.C., K.S.H., V.M., K.T.M.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions
R.A. developed the automated NMR framework and the underlying Python-based codes, performed all the necessary calculations and had primary writing responsibilities. Y.C. and K.S.H. carried the NMR experiments. V.M. and K.T.M. guided the experimental aspect of the project. N.N.R. guided and led the computational aspects of the project. All authors contributed to writing and reviewing the manuscript.

Competing interests
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-022-00200-9.

Correspondence and requests for materials should be addressed to Nav Nidhi Rajput. Peer review information Nature Computational Science thanks Yanfei Guan, Alexej Jerschow and Yoonjoo Kim for their contribution to the peer review of this work. Handling editor: Kaitlin McCardle, in collaboration with the Nature Computational Science team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022