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

Rasha Atwi
Stony Brook University
Ying Chen
Pacific Northwest National Laboratory
Kee Sung Han
Pacific Northwest National Laboratory
Karl Mueller
Pacific Northwest National Laboratory
Vijayakumar Murugesan
Pacific Northwest National Laboratory
Nav Nidhi Rajput (navnidhi.rajput@stonybrook.edu)
Stony Brook University

Article

Keywords: nuclear magnetic resonance, classical molecular dynamics, chemical shifts

Posted Date: October 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-893249/v1

Version of Record: A version of this preprint was published at Nature Computational Science on February 1st, 2022. See the published version at https://doi.org/10.1038/s43588-022-00200-9.

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

Rasha Atwi[1], Ying Chen[2], Kee Sung Han[2], Karl T. Mueller[2], Vijaykumar Murugesan[2], Nav Nidhi Rajput*[1]

[1]Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States

[2]The Joint Center for Energy Storage Research (JCESR), Pacific Northwest National Laboratory, Richland, WA 99352, USA

* To whom correspondence should be addressed:

Nav Nidhi Rajput; Email: navnidhi.rajput@stonybrook.edu

ABSTRACT

Identifying stable speciation in multicomponent liquid solutions is of fundamental importance to areas ranging from electrochemistry to organic chemistry and biomolecular systems. However, elucidating this complex solvation environment is a daunting task even when using advanced experimental and computational techniques. Here, we introduce a fully automated, high-throughput computational framework for the accurate and robust prediction of stable species present in liquid solutions by computing the nuclear magnetic resonance (NMR) chemical shifts of molecules. The framework automatically extracts and categorizes hundreds of thousands of atomic clusters from classical molecular dynamics (CMD) simulations to identify the most stable speciation in the solution and calculate their NMR chemical shifts via DFT calculations. Additionally, the framework creates an output database of computed chemical shifts for liquid solutions across a wide chemical and parameter space. This task can be infeasible experimentally and challenging using conventional computational methods. To demonstrate the capabilities of our framework, we compare our computational results to experimental measurements for a complex test case of magnesium bis(trifluoromethanesulfonyl)imide $Mg(TFSI)_2$ salt in dimethoxyethane (DME) solvent, which is a common electrolyte system for Mg-based batteries. Our extensive benchmarking and analysis of the Mg^{2+} solvation structural evolutions reveal critical factors such as the effect of force field parameters that influence the accuracy of NMR chemical shift predictions in liquid solutions. Furthermore, we show how the framework reduces the efforts of performing and managing over 300 ^{13}C and 600 ^{1}H DFT chemical shift predictions to a single submission procedure. By enabling more efficient and accurate high-throughput computations of NMR chemical shifts, our approach can accelerate theory-guided design of liquid solutions for various applications.

INTRODUCTION

Liquid solutions are critical components of various chemical, materials science, engineering, and biological applications such as batteries[1-3], fuel[4], food industry[5], and drug discovery[6,7]. 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, nuclear magnetic resonance (NMR) spectroscopy stands out as a powerful and widespread technique for studying the 3D organization of matter and associated structural and dynamical properties[8-10]. Over the years, technological advances in NMR spectroscopy have significantly improved the operational ease and spectral resolutions obtainable from non-traditional nuclei (such as ^{17}O, ^{25}Mg, etc.), leading to a comprehensive and atomistic view of liquid solutions[11,12]. However, NMR spectroscopy is limited by the temporal scale and low sensitivity, making it difficult to speciate structural patterns that are often driven by electrostatic interactions, reactivity, temperature, compositional variance, and pressure[13-15].

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 the structural evolutions and associated chemical shifts[16-18]. 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 shift

trends[8,19,20]. However, they fail to fingerprint the temporal evolution of solvation structures under exogenous (temperature and pressure) and endogenous (pH and composition) conditions. In addition, gaps in knowledge between systems examined *in-situ* or *ex-situ* and those modeled *in silico* still exist. For example, NMR DFT studies are often focused on singular phenomena, e.g., magnetic shielding tensor. Recently, an automated framework[21] and a machine learning based approach[22] were implemented to predict the $^{13}C/^{1}H$ NMR chemical shift for organic molecules. However, a generalized approach to identify complexes 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, etc. Even for the singular focus on chemical shift calculations, the possible molecular structure(s) are built manually based on chemical intuition, trial and error, and/or results reported in the literature[20,23,24]. 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 multicomponent liquid solutions and guide experiments to identify stable speciation in the 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 through force field generation and information flow between the two length scales. The tool automates the entire process, starting from sampling hundreds of

91 thousands of possible configurations in solute-solvent systems to identifying the most stable
92 configurations and predicting and storing their NMR chemical shifts in a database. To the best of
93 our knowledge, an automatic derivation of NMR chemical shifts with explicit solvation has not
94 yet been implemented in any software infrastructure. Although the developed tool is general
95 enough to be applied to any liquid solution, we consider magnesium
96 bis(trifluoromethanesulfonyl)imide $Mg(TFSI)_2$ salt in dimethoxyethane (DME) solvent as an
97 illustrative example. The chosen electrolyte formulation has received considerable attention in
98 battery literature but reported findings regarding the speciation and the exact solvation structure
99 of the Mg cation are under contention. More specifically, experimental work reported the
100 formation of solvent separated ion pairs (SSIPs), while contact ion pairs (CIPs) were observed in
101 previous computational results[14,23,25,26]. A comprehensive molecular level understanding of the
102 speciation present in the solution can allow tuning the chemical structure to control the stability,
103 solubility, structural, and dynamical properties of liquid solutions. We note that we chose a system
104 in which complexities in the solvation phenomena arise due to the multivalent nature of the cation,
105 providing an example to demonstrate that the developed framework can be applied to other simpler
106 systems. We report a detailed comparison between computed and experimental NMR chemical
107 shifts for ^{25}Mg, ^{13}C, and ^{1}H nuclei in this electrolyte. We also demonstrate the high-throughput
108 capability of the workflow by accurately predicting more than 300 ^{13}C and 600 ^{1}H NMR chemical
109 shifts from a set of 100 organic molecules from the SDBS[27] database and a previous experimental
110 study[28]. In the second section, we address the fundamental challenge of how to accurately predict
111 NMR chemical shift of liquid solutions by associating the framework with a benchmarking study.
112 This study reveals several factors such as the choice of force field parameters that affect the
113 accuracy of predicted chemical shifts, which can be employed by a number of research

communities by increasing the accessibility to DFT-based chemical shifts for a wide variety of structures and liquid systems.

RESULTS AND DISCUSSION

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 computational materials science simulations and multiple systems with a strong focus on data provenance. MISPR automates many computational tasks that are typically performed manually. Its functionalities span from processing and manipulating molecular structures, preparing and executing DFT and CMD simulations on supercomputing resources, parsing and analyzing output data, and creating output databases that organize the results from individual calculations. To manage the heterogeneous data that DFT and CMD workflows output and allow for flexible and complex queries, MISPR employs MongoDB[29] for data storage. MongoDB is a document-oriented NoSQL database that stores data as JSON-formatted documents with flexible schema. 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 (e.g., molecular structure, the size and geometry of the system for CMD simulations, etc.) to generate and run a comprehensive workflow. We built MISPR on top of base libraries developed by the Materials Project, namely: (1) *pymatgen*[30] for structure representation and input/output files generation and handling, (2) *FireWorks*[31] for managing workflows over computing resources, and (3) *custodian*[32] for monitoring inevitable

errors during simulations and applying on-the-fly fixes. At the backend, MISPR uses Gaussian[33] electronic structure software for DFT calculations and LAMMPS[34] (https://www.lammps.org/) open-source code for CMD simulations. Examples of implemented DFT workflows include calculating binding energy, redox potentials, and bond dissociation energy. CMD workflows in MISPR allow executing CMD simulations in various ensembles and analyzing collected trajectories for structural and dynamical properties. Force field parameters and derived properties are saved in their collections with auxiliary information like molecular metadata (e.g., InChI representation, chemical formula, etc.) and input parameters, making it easy to reproduce and query computational results. More details about the MISPR infrastructure will be the subject of a future publication.

The framework designed for automatic NMR chemical shift calculations in liquid solutions is outlined in Fig 1. The framework takes as input the structures of molecules comprising a liquid solution of interest. Many molecule formats are supported (e.g., XYZ file, PDB file, pymatgen molecule object, Gaussian output, etc.) via the OpenBabel[35] and pymatgen libraries. Besides 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. Next, the framework runs an electrostatic partial charges (ESP) workflow that first converts the input structure formats to pymatgen molecule objects. The ESP workflow uses this molecule object to generate a Gaussian input file with input parameters specified as optional inputs to the workflow. The workflow uses default values if these parameters are not provided. 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.

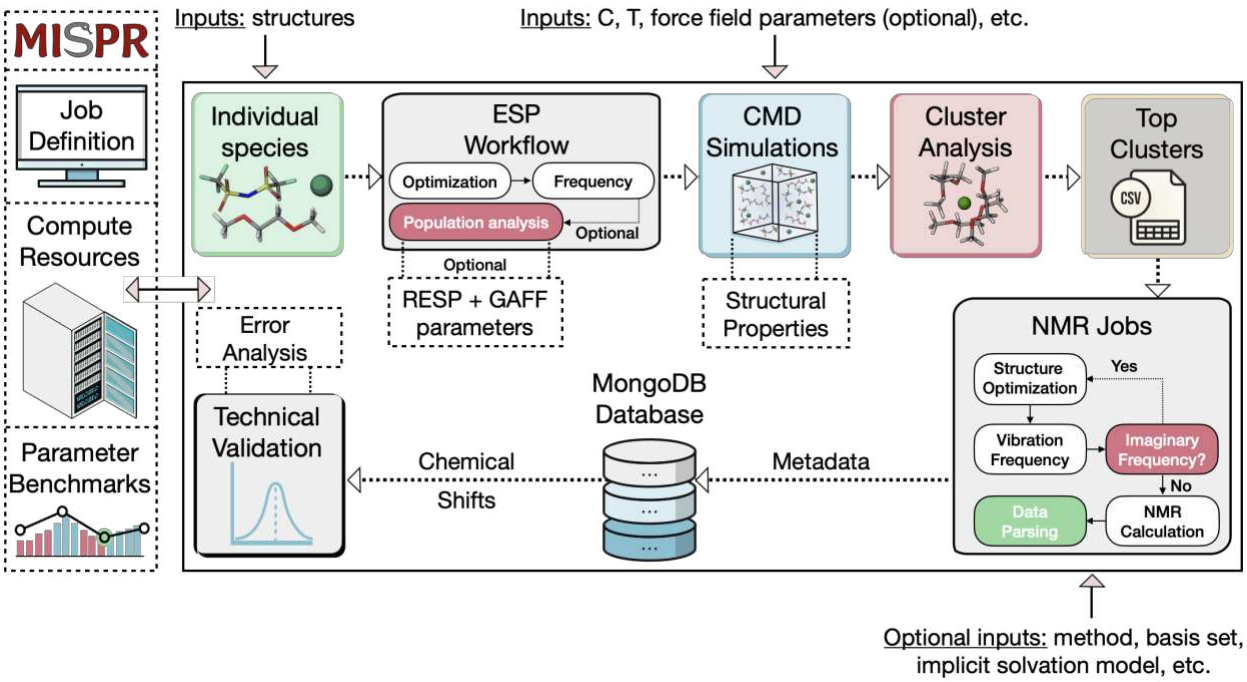

Fig. 1 Scheme of the computational framework used to calculate NMR chemical shifts in solution as implemented in the MISPR high-throughput infrastructure

We note that the framework is general enough to be applied to various complex liquid solutions at different conditions (e.g., concentration, temperature, pressure, etc.). It requires, at minimum, the concentration of species in the solution and the size and geometry of the system box to prepare the multicomponent system for CMD simulations. One of the most challenging aspects of running automated CMD simulations is selecting or generating accurate force field parameters. By default, the framework uses the output of the ESP workflow to derive the general amber force field (GAFF)[36] parameters for each species. The framework also supports other force fields allowing the user to test different physical models for a specific application or system. In this case, the user may input the force field parameters to the framework in the form of a Python dictionary or retrieve them from our in-house database. We note that the user may bypass the ESP workflow if the ESP charges have been previously calculated or other force fields are directly provided. The framework

then passes the optimized geometries, force field parameters, concentrations, and information about the geometry of the simulation box (e.g., lengths, shape, etc.) to the next step to build the system for LAMMPS simulations. Following this, the framework runs a CMD workflow to generate time trajectories of atomic positions and velocities. Configurations for common CMD procedures are encoded in a set of protocols that can be used directly or altered to run any series of LAMMPS calculations according to the user's needs. The default CMD configuration involves energy minimization, NPT equilibration at the desired temperature and pressure, melting and quenching, and NVT production runs.

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 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 the section 1 of the SI.

Sampling solvation structures from the CMD step is a key component of the NMR framework. Traditional NMR calculations are 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 is extremely time consuming[20,23,24]. In contrast, our framework passes the computed RDF from the previous step to perform sampling of the first solvation shell of a particle of interest in a straightforward and automated manner. In the framework, the first solvation shell is defined by the cutoff distance r_{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. In the default operation of the framework, r_{min} is automatically extracted from the RDF, but the user may override this by providing r_{min} as an optional input. Thus, a cluster representing the solvation structure is defined as the group of species within r_{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 of interest 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. 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. Configurations with the highest probability of occurrence correspond to persistent metastable solvation structures in the solution. 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 (e.g., bond lengths, orientation, etc.) 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 the MMFF94s force field[37] as implemented in the RDKit library[38]. The framework then feeds the lowest-energy conformer of each configuration to an NMR DFT workflow. We note that it would be infeasible to manually generate and categorize this large number of structural files and account for all the possible solvation structures using conventional

methods that rely on chemical intuition. This task is especially challenging for chemical systems that have not been previously explored in detail.

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 if a true potential energy surface (PES) minimum is reached. The framework by default uses the ωB97X[39]/def2-TZVP level of theory for performing these three sequential DFT steps. Switching the functional, basis set, and other Gaussian input parameters (e.g., solvation model, numerical and algorithmic parameters, etc.) is straightforward and requires the user to input them in the form of a Python dictionary to the framework. The framework then performs an analysis step that stores the calculation results in an NMR collection in the database or a local JSON file with all the necessary metadata for future reference. 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 Fig S1. Finally, results from the computational framework are compared to experimental NMR spectra to elucidate the solvation structures.

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 process is employed through well-defined rules via the custodian package if an error is detected. If possible, the error handler applies the appropriate remedy, generally by modifying the input parameters, writing a new Gaussian input file, and restarting the calculation. If no remedy has been implemented for a

240 particular error or the error handler cannot interpret the encountered error, the calculation is
241 allowed to fail. The error handler improves the success rate of the calculations without relying on
242 human intervention, which would be impossible for handling large computational investigations.
243 Examples of the errors addressed are SCF failure, geometry optimization convergence, error in
244 internal coordinates, and exceeded wall time limit.

245 The framework takes solvent effects into account by two approaches. It uses an explicit
246 approach where several solvent molecules surrounding the species are correctly placed in its first
247 solvation shell since the geometries are extracted directly from CMD simulations. Second, it
248 approximates bulk solvent effects using a dielectric continuum model. This approach allows
249 incorporating a thermodynamically stable and realistic chemical environment of species compared
250 to the traditional approach, which relies on either implicit solvent models or manual prediction of
251 the possible solvation structures. Since multiple configurations are considered, collected data result
252 in various chemical shifts corresponding to different chemical environments experienced by the
253 nucleus of interest. Therefore, predictions from this approach can be compared and fitted to the
254 entire experimental NMR peak rather than just matching the peak center, especially when peak
255 broadening occurs due to distribution of chemical shifts or intermediate exchange dynamics in
256 solutions.

257 Components of the NMR framework presented in Fig 1 can be decoupled according to the
258 needs of the user. For example, we used the NMR workflow as a standalone code to compute the
259 ^{13}C and ^{1}H chemical shifts for a set of 100 organic molecules. Detailed information about the
260 library is provided in Table S1. The calculations were performed in a chloroform solvent at the
261 ωB97X/def2-TZVP level of theory and referenced to tetramethylsilane (TMS). The code snippet
262 in Fig S2 demonstrates how to submit these calculations starting from structures defined in the

263 XYZ file format. Upon submitting the script, the calculations were added to a *FireWorks* database
264 and subsequently executed over computing resources. The workflow generated and managed over
265 600 input and output files and inserted more than 300 ^{13}C and 600 ^{1}H chemical shifts into the
266 database via a simple one-shot script. We compared our predictions to experimental data from the
267 SDBS[27] database and a previous study[28]. Fig S3 and Fig S4 show parity plots of the computed
268 chemical shifts and their associated error distribution, respectively. A good correlation is observed
269 between the workflow output and the experimental data with only minor deviations from the fitted
270 line. This example demonstrates how our high-throughput approach may be adapted for the
271 determination of accurate NMR chemical shifts.

272 **Factors affecting the accuracy of NMR chemical shifts**

273 Reliably differentiating among different extracted solvation structures requires high accuracy
274 NMR chemical shift predictions. The successful implementation of our framework necessitates
275 adequate consideration of several important factors. First, a key question for the CMD component
276 is the quality of the interatomic potentials since significant deviations in system properties have
277 often been observed compared to experimental data[40]. Second, the DFT level of theory comprising
278 the density functional and basis set is critical for achieving well-converged chemical shieldings.
279 Achieving this convergence for small molecules is relatively straightforward by combining DFT
280 or even coupled cluster calculations with large basis sets. However, this is much more challenging
281 with complexes consisting of multiple species. Therefore, there is a need to balance the cluster size
282 with the quality of the DFT level of theory. In addition, the choice of the implicit solvent model is
283 crucial for approximating the bulk solvent effect. A remarkable number of benchmarking studies
284 have been done on quantum mechanical methods for predicting properties in complex
285 multicomponent battery electrolytes similar to the test case here[41-43]. However, parallel studies for

286 NMR calculations for these types of systems are still in their infancy. Other factors include
287 selecting an appropriate number of molecules in the chemical reference to account for
288 intermolecular interactions and a representative conformer from each solvation environment. In
289 the following sections, we report the role of each of these factors using results obtained by the
290 framework for the $Mg(TFSI)_2$/DME test case system.

291 Role of the force field

292 The choice of the force field parameters used in CMD simulations can significantly influence
293 the speciation observed in solution, and thus the NMR chemical shift predictions. Therefore, we
294 benchmark the most commonly used and reliable force fields for liquid solutions, including GAFF
295 (FF1), non-polarizable OPLS[44] (FF2), and polarizable OPLS (FF3) force fields to compare their
296 performance in terms of the solution properties. FF1 and FF2 are computationally less expensive
297 due to their non-polarizable nature and have been used extensively in the battery literature showing
298 satisfactory agreement with experimental findings. FF3, built on top of FF2, allows for a
299 polarizable response of molecules to an electric field using the Drude oscillator model[45]. In this
300 model, particles are added to each polarizable atom to mimic physical dipoles and model the
301 corresponding distortion of electron density.

302 The simulation density (ρ) using the three force fields, shown in Table S2, agrees well with
303 the experimental value. The lowest average error (1.2 %) is achieved with FF3. The RDFs between
304 the cation and oxygen atoms of DME and $TFSI^-$ are shown in Fig 2a and 2b, respectively. FF1
305 results in the weakest cation-solvent [Mg^{2+}-DME] interaction and the most vital cation-anion
306 [Mg^{2+}- $TFSI^-$] interaction, as evident from the sharp RDF peak between the cation and oxygen
307 atoms of the anion. On the other hand, with FF3, little coordination occurs with the anion (inset of
308 Fig 2b), indicating that solvent molecules dominate the first solvation shell of the cation. Fig 2c

shows Mg^{2+} - O (DME) and Mg^{2+} - O ($TFSI^-$) coordination numbers, calculated by integrating the corresponding RDF curves for the first solvation shell. FF3 results indicate that $Mg(TFSI)_2$ 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 $TFSI^-$ that coordinate to Mg^{2+} with two oxygen atoms, *i.e.,* in bidentate configuration, as shown in Fig S5.

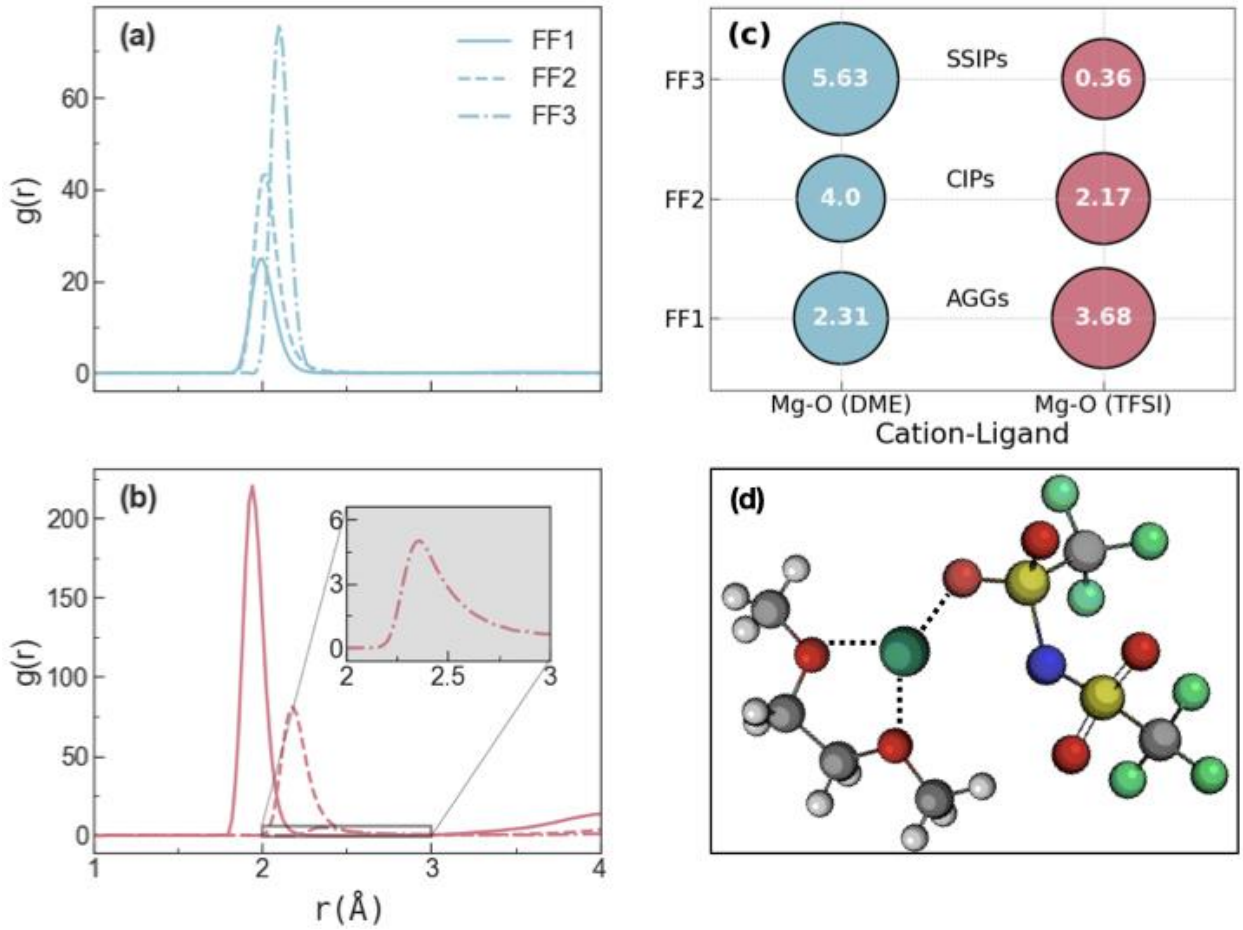

Fig. 2 Structural properties of $Mg(TFSI)_2$ in DME at 298.15 K using FF1 (GAFF), FF2 (non-polarizable OPLS), and FF3 (polarizable OPLS). RDF of (a) Mg^{2+} – O (DME) and (b) Mg^{2+} – O ($TFSI^-$), (c) coordination numbers with Mg^{2+} with the corresponding type of structure: solvent separated ion pairs (SSIPs), contact ion pairs (CIPs), and aggregates (AGGs), and (d) corresponding types of coordination with oxygen atoms of DME and $TFSI^-$

317 The top Mg^{2+} configurations identified by the framework are provided in Fig S6 and Fig S7.
318 Overall, we find significant differences in the type and distribution of these structures among the
319 tested force fields. For example, the most probable solvation structure predicted with FF1 involves
320 one DME molecule in bidentate configuration and four $TFSI^-$ anions in monodentate configuration.
321 In addition, rather than forming a single stable solvate like in the case of FF3, the distribution of
322 coordination environments for the cation with FF1 is much more heterogeneous and involves
323 configurations dominated by the anion. With FF2, the electrostatic interaction with the anion is
324 slightly suppressed, and the most probable solvation shell is composed of two DME solvents and
325 $TFSI^-$ anions participating in bidentate and monodentate configurations, respectively. FF3 results
326 in an Mg^{2+} solvation shell dominated by three DME molecules participating in bidentate
327 configuration, with only minor structures containing an anion. This configuration has been
328 previously suggested based on experimental measurements of diffusion and Raman and NMR
329 spectroscopy[23,26], and computationally by Kubisiak and Eilmes[46] for a concentration range of 0.1-
330 1 M.

331 Variations in the structural properties between the tested force fields are translated to the
332 dynamical behavior of the electrolyte. The distribution of diffusion coefficients (Fig S8) from FF3
333 indicates 2.58 slow DME molecules per Mg^{2+} cation. This result is in close agreement with the
334 experimentally measured value of 3.0[23] and is consistent with the computed structural properties.
335 FF1 and FF2 predict 1.17 and 1.86 slow DME molecules per Mg^{2+}, respectively. The calculated
336 ionic diffusion coefficients with FF3 are also in better agreement with experimental results (mean
337 absolute error of 20%), whereas those from FF1 and FF2 are underestimated by approximately
338 90% and 30%, respectively (Fig S9).

339 The discrepancies in the predicted properties are not particularly a problem of a specific force
340 field or the $Mg(TFSI)_2$/DME system, but rather due to a lack of accounting for the critical
341 interactions in the non-polarizable simulations. The predicted properties using FF3 are the most
342 consistent with previous experimental[23,26] and computational[46] studies among the tested force
343 fields. However, the better performance of FF3 comes at the expense of its 2-3 fold higher
344 computational time compared to FF1 and FF2. To summarize, the computational results for the
345 cation-anion motifs and the propensity of the salt to form ion aggregation in the solution are
346 strongly dependent on the type of the force field. Therefore, evaluating the quality of the force
347 field used in the sampling process is a necessary primary step to obtain reliable structures for NMR
348 computations. Here, we proceed with the FF3-predicted solvation structures to report results from
349 the DFT component of the NMR framework.

350 Role of the DFT level of theory

351 We evaluate the performance of selected DFT functionals and basis sets in predicting chemical
352 shifts of ^{25}Mg, ^{13}C, and ^{1}H of the top configurations and the chemical shifts of ^{13}C and ^{1}H
353 resonances in the bulk solution. All calculations presented in this section are performed using the
354 polarizable continuum model (PCM)[47-49]. The NMR workflow (Fig 1) is designed to be used in
355 high-throughput mode to study speciation evolution in liquid solutions at variable conditions, *e.g.,*
356 concentration and temperature. Therefore, the comparison is made not only based on accuracy but
357 also on factors that are particularly important for high-throughput simulations (*e.g.,* computational
358 cost and tendency to fail). Predictions from combinations of four commonly utilized DFT
359 functionals (B3LYP[50], M06-2X[51], PBE1PBE[52], and ωB97X[39]) and three Gaussian basis sets (6-
360 31+G*, 6-311++G**, and def2-TZVP) are compared with experimental NMR data.

361 During the benchmark study, the most common failures encountered include failure to
362 converge the geometry to a PES minimum in a finite number of optimization steps, difficulties in
363 converging SCF calculations, and errors in internal coordinate transformations. Around 78% of
364 the total performed calculations were completed without error correction procedures. Levels of
365 theory primarily involved in the failed calculations include B3LYP/6-31+G* and PBE1PBE
366 hybrid functional coupled with each of the 6-31+G* and 6-311++G** basis sets. Given that one
367 of our primary goals is to find a level of theory that is not likely to fail with complex
368 multicomponent clusters, these levels of theory are not considered the most appropriate for the
369 required task.

370 The ^{25}Mg NMR results from the top-performing level of theory (ωB97X/def2-TZVP) are
371 shown in Fig 3 along with the corresponding structure of the predicted species. A single broad
372 peak is observed, indicating either a single solvation structure or a convolution of multiple
373 structures with a rapid exchange. The predicted ^{25}Mg chemical shift in the most probable
374 configuration is -0.809 ppm, which is highly consistent with the experimental peak center located
375 at -0.71 ppm. Given the broad line width of the ^{25}Mg peak, *i.e.,* the half peak height at 0.83 and -
376 2.13 ppm, the chemical shift of ^{25}Mg in configuration 2 (Table 1) is also deemed to be in
377 satisfactory agreement with experimental data. Therefore, multiple Mg^{2+} structures that are
378 entirely dissociated from the anion are possible in the solution. Excluding configuration 4, the
379 increase in the ion-dipole interaction between Mg^{2+} and $TFSI^-$ in the following order: configuration
380 1 < configuration 2 < configuration 3 < configuration 6 < configuration 5 leads to the observed
381 monotonic upfield shift in the corresponding ^{25}Mg chemical shift. The presence of loosely packed
382 clusters of $[Mg(DME)_n]$ ($n \leq 2$), *i.e.,* configuration 4, is attributed to the high degree of freedom
383 and structural flexibility of DME. This type of configuration has been reported to be favorable at

lower concentrations due to lower electrostriction (reduced solvent volume in the Mg^{2+} solvation shell relative to the bulk) and diminished entropy loss[23]. On the contrary, higher concentrations (0.51 M) such as the one used in this study lead to closer distances between Mg^{2+} ions, resulting in stronger electrostatic interactions and dampened DME motion, thus favoring fully solvated clusters (n = 3, configuration 1). This behavior is consistent with the low probability of configuration 4 and the predicted ^{25}Mg chemical shift of this configuration, which is far from the experimental peak center (Fig 3).

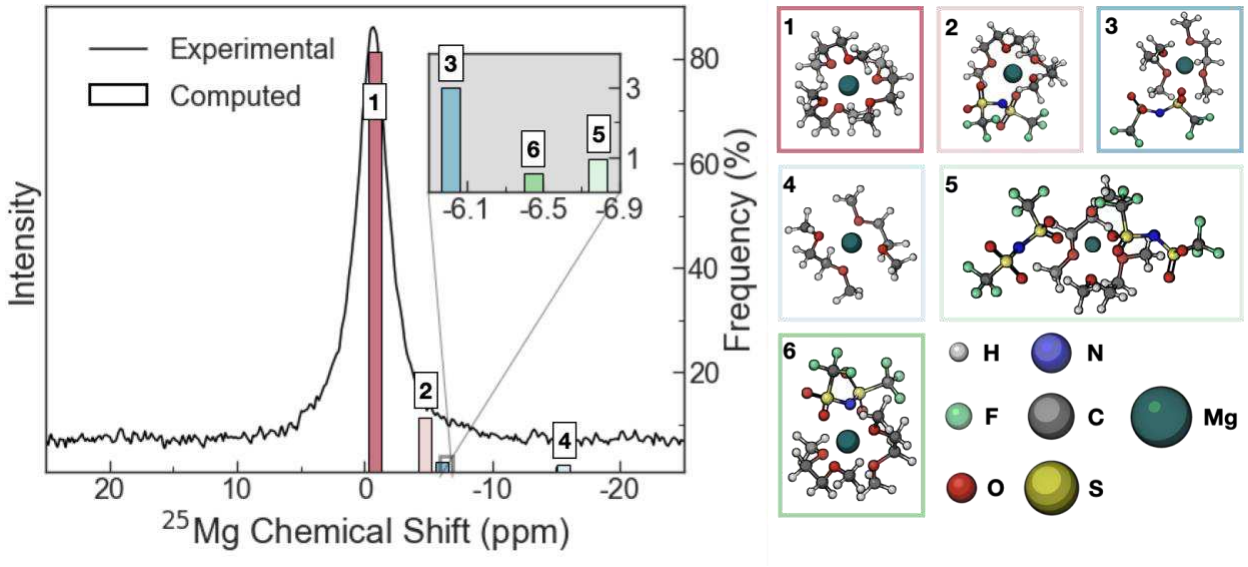

Fig. 3 Predicted ^{25}Mg NMR chemical shifts using the NMR computational protocol and the experimental NMR spectrum along with the corresponding predicted solvation structures of 1:18 $Mg(TFSI)_2$ in DME solution. DFT calculations are performed at the ωB97X/def2-TZVP level of theory using the PCM solvation model

Table 1 DFT predicted chemical shifts for 1:18 $Mg(TFSI)_2$ in DME solution along with deviations from experimental data

Molecule[a]	δ ^{25}Mg (ppm)		δ ^{13}C (ppm)		δ ^{1}H (ppm)	
	PCM[b,c]	SMD[b,c]	Shift[c,d]	Deviation[e]	Shift[d,f]	Deviation[e]
Bulk DME			CH_2: 72.32	CH_2: 0.83	CH_2: 3.74	CH_2: 0.06
			CH_3: 59.16	CH_3: 0.27	CH_3: 3.64	CH_3: -0.01
Configuration 1	-0.809	0.283	CH_2: 72.62	CH_2: -0.26	CH_2: 4.24	CH_2: 0.16
$[Mg(DME)_3]^{2+}$			CH_3: 62.70	CH_3: -0.09	CH_3: 4.07	CH_3: 0.29
Configuration 2	-4.741	-4.649	CH_2: 72.50	CH_2: -0.14	CH_2: 4.20	CH_2: 0.20
$Mg(DME)_2(TFSI)]^{+}$			CH_3: 62.18	CH_3: 0.43	CH_3: 4.11	CH_3: -0.01
Configuration 3	-6.016	2.784	CH_2: 72.16	CH_2: 0.20	CH_2: 4.27	CH_2: 0.13
$Mg(DME)_2(TFSI)]^{+}$			CH_3: 61.64	CH_3: 0.97	CH_3: 4.15	CH_3: -0.05
Configuration 4	-15.559	-3.982	CH_2: 73.37	CH_2: -1.01	CH_2: 4.32	CH_2: 0.08
$[Mg(DME)_2]^{2+}$			CH_3: 62.73	CH_3: -0.12	CH_3: 4.14	CH_3: -0.04
Configuration 5	-6.785	-6.196	CH_2: 71.85	CH_2: 0.51	CH_2: 4.10	CH_2: 0.30
$[Mg(DME)_2(TFSI)_2]$			CH_3: 62.57	CH_3: 0.04	CH_3: 4.14	CH_3: -0.04
Configuration 6	-6.450	-5.815	CH_2: 72.68	CH_2: -0.32	CH_2: 4.29	CH_2: 0.11
$[Mg(DME)_2(TFSI)]^{+}$			CH_3: 62.52	CH_3: 0.09	CH_3: 4.13	CH_3: -0.03

[a] For the difference between configurations 2, 3, and 6, refer to Fig S3,
[b] Compared to experimental peak center at -0.71 ppm,
[c] Using ωB97X/def2-TZVP,
[d] Using PCM model,
[e] Deviation = δ_{exp} - δ_{DFT},
[f] Using M06-2X/def2-TZVP.

The benchmarking results for ^{25}Mg chemical shift calculations are displayed in Fig S10. On average, going from left to right, *i.e.,* increasing the number of basis functions, moves most of the predicted chemical shifts corresponding to different electronic environments surrounding the ^{25}Mg nucleus within the bounds of the observed NMR spectrum. From top to bottom, significant variations are observed in the predicted chemical shifts using the four functionals with 6-31+G*, while this difference is less clear with def2-TZVP. In addition, we find that different levels of theory can lead to contradictory conclusions regarding the dominant species in solution. For example, the structure is predicted to be $[Mg(DME)_2(TFSI)]^+$ using PBE1PBE/6-31+G* while the fully solvated $[Mg(DME)_3]^{2+}$ is found with PBE1PBE/def2-TZVP.

Fig 4 shows ^{13}C NMR shifts assigned to CH_3 of DME existing in the bulk solution (labeled 'free CH_3') and DME coordinated to Mg^{2+} (labeled 'bound CH_3') from DFT predictions and experimental measurements. Similar plots for ^{13}C shifts assigned to CH_2 and ^{1}H shifts assigned to CH_3 and CH_2 of both types of DME molecules are shown in Figs S11-S13, respectively. While free and bound DME molecules are distinguishable from experimental and predicted ^{13}C and ^{1}H NMR chemical shifts, it is impossible to differentiate between bound DME at different configurations identified in this work. The spectroscopic differences between the structures may be subtle (see, for example, Table 1 for ^{13}C and ^{1}H chemical shifts in different configurations). On the contrary, ^{25}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}Mg nucleus, thus giving rise to different NMR responses. As displayed in Fig 4 and Fig S11, the highest deviation from experimental ^{13}C shifts are obtained with the 6-31+G* and 6-311++G** basis sets combined with any tested density functional. The basis set from the 'def2' family of Alrichs and coworkers[53], particularly in combination with ωB97X, leads to ^{13}C NMR

414 chemical shift error that approaches the underlying uncertainty in experimental measurements
415 (Table 1). Fig S12 and S13 indicate that for ^{1}H chemical shifts, M06-2X/def2-TZVP outperforms
416 the other tested levels of theory with absolute errors between 0.01 and 0.3 ppm (Table 1).

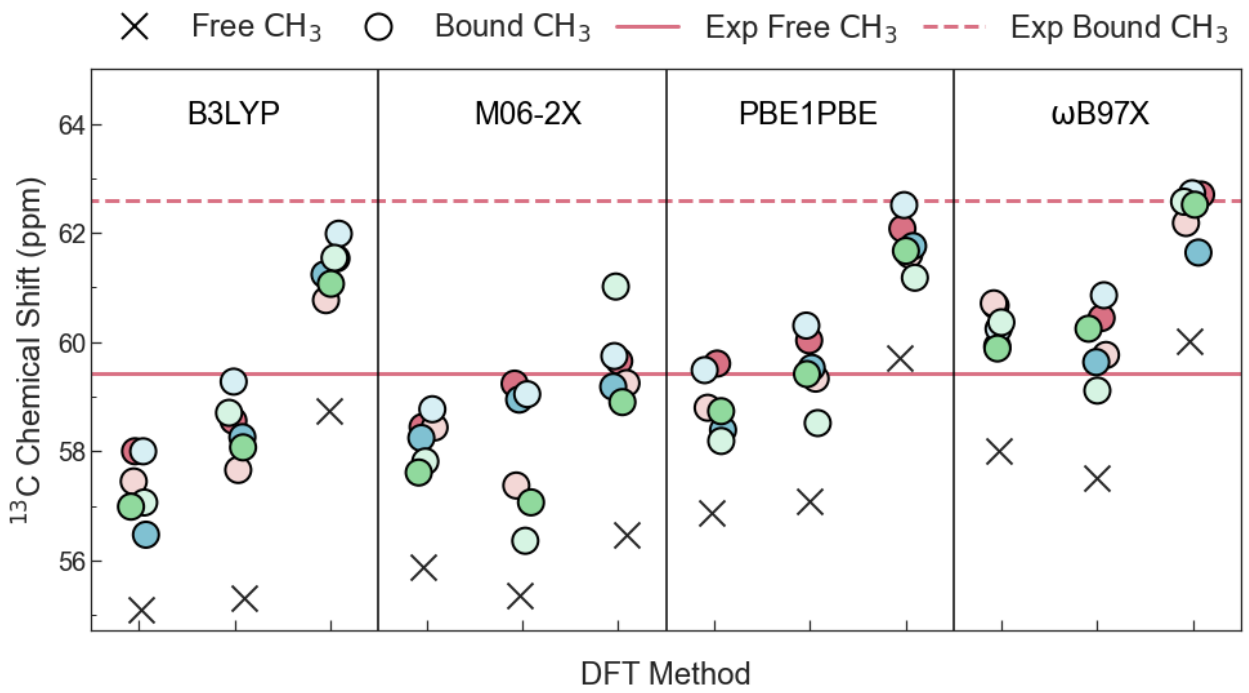

Fig. 4 Strip plot of the computed and experimental ^{13}C NMR chemical shifts assigned to CH_3 of DME coordinated to Mg^{2+} (labeled Bound CH_3) and CH_3 of free DME (labeled Free CH_3). For color code of 'Bound CH_3', please refer to Fig 3. Results from each DFT functional are shown with the basis sets in the following order: 6-31+G*, 6-311++G**, and def2-TZVP

417 We conclude that the choice of the basis set has the highest impact on the accuracy of NMR
418 chemical shift predictions. The 6-31+G* basis set is ruled out as a suitable basis set for NMR
419 calculations of complexes similar to those studied herein due to its degraded accuracy compared
420 to other basis sets, despite its lower computational cost (see Fig S14 for timings). For ^{25}Mg and
421 ^{13}C chemical shifts, the ωB97X/def2-TZVP level of theory is recommended if computational
422 resources are available as its remarkable accuracy and the applicability of def2-TZVP to broader
423 chemical systems make it well worth the additional cost. If computational resources are limited,
424 M06-2X/6-311++G** is recommended for ^{25}Mg shifts as its cost is not prohibitive while still

425 predicting the correct Mg^{2+} solvation structure. Finally, M06-2X with def2-TZVP or 6-311++G**
426 are recommended for ^{1}H chemical shift predictions.

427 Effect of geometry optimization

428 To examine the possibility of making DFT calculations more affordable, we calculated the
429 ^{25}Mg chemical shift of 33 pre-relaxed clusters extracted from CMD simulations. We then compare
430 their deviation from calculations utilizing optimized geometries at the same level of theory (Fig
431 S15). We find a mean absolute deviation of ~ 37.6 ppm between the two types of calculations,
432 with a systematic downfield shift from calculations utilizing fully optimized structures. This result
433 is not surprising due to the sensitivity of the ^{25}Mg nucleus to subtle differences in the local structure
434 and coordination environment. Therefore, relaxing the structures ensures that 'reasonable
435 geometries' are used, and therefore is a prerequisite for obtaining accurate NMR chemical shifts
436 that are comparable to experimental measurements.

437 Choice of the chemical reference

438 Because water is selected as the ^{1}H chemical shift reference, another consideration is the
439 accurate computational representation of the effect of strong hydrogen bonding among water
440 molecules. To this end, calculations on clusters of $(H_2O)_n$ (n = 1 - 4, 6, 8, 10, 12) are performed.
441 At the M06-2X/def2-TZVP level of theory, the isotropic shielding constant of ^{1}H moves upfield
442 when the number of water molecules increases and tends to converge at ~ 27.6 ppm for eight water
443 molecules. In addition, multiple clusters for non-hydrogen-bonding dimethylsulfoxide $(DMSO)_n$
444 (n = 1 - 4, 6), used as a reference in ^{13}C chemical shift calculations, are considered. The use of a
445 DMSO dimer is found to be sufficient, whereby the ^{13}C isotropic shielding constant converges at
446 ~ 150 ppm at the ωB97X/def2-TZVP level of theory. Finally, since intermolecular interactions
447 inevitably exist in DME solution, calculations on $(DME)_n$ (n = 1 - 4) are carried out for predicting

448 the ^{13}C and ^{1}H chemical shifts of DME molecules in the bulk solution, and $(DME)_2$ is found to
449 result in bulk CH_3 and CH_2 chemical shifts that reproduce the experimental data. All calculated
450 isotropic shielding constants for H_2O, DMSO, and DME clusters are included in the dataset
451 associated with this work.

452 Role of the implicit solvation model

453 In addition to the explicit solvent molecules modeled in the Mg^{2+} first solvation shell, an
454 implicit model is used to incorporate long-range electrostatic effects. Implicit solvent models have
455 the advantage of reducing the number of degrees of freedom of the environment (solvent), thereby
456 decreasing the computational cost to describe the dielectric continuum outside the solute cavity.
457 SMD is reliable in many applications[54] and therefore is compared to the PCM results in this work.
458 As evident from the data in Table 1, both methods lead to similar interpretations of experimental
459 results in terms of the most probable solvation structure. However, the PCM method predicts more
460 accurately the chemical shift of the top configuration. The only significant differences between the
461 two methods are for configurations 3 and 4. Similarly, more minor errors associated with ^{13}C and
462 ^{1}H chemical shifts are achieved with PCM than the SMD model using the ωB97X/def2-TZVP and
463 M06-2X/def2-TZVP levels of theory, respectively (Fig S16).

464 Role of conformer

465 Another consideration in the NMR framework is that it utilizes the lowest-energy conformer
466 from each configuration to initialize the NMR DFT calculations. Previous NMR DFT studies have
467 also reported findings on possible solvation structures based on a single conformer[20,24,55].
468 However, the measured shift is the weighted average of chemical shifts of all possible conformers
469 in solution during the NMR acquisition time. Therefore, we assess the sensitivity of DFT chemical
470 shifts to conformer sampling by starting from the MMFF94 energies of a total of ~ 270,000

conformers of configuration 1 and ~ 4,000 conformers of configuration 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 MM global minimum energy conformer. 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, $\langle\delta\rangle$, and our initial chemical shift estimation, δ_0, as a function of the number of optimizations performed. We note that for each number of optimized structures (N_{opt}) shown on the x-axis of Fig 5, the calculated $\langle\delta\rangle$ is the result of averaging over all possible combinations of N_{opt} from a pool of 15 structures, with a restriction that the MM global minimum energy conformer is included in these combinations. To maintain statistical significance, only N_{opt} resulting in more than 30 possible combinations are used, thus $N_{opt} = 2, 13, 14, 15$ are excluded from the analysis. Variable degrees of errors are obtained with each nucleus type, with the highest difference in the ^{25}Mg chemical shift. In this electrolyte system, a maximum unsigned error of 1.2 ppm in the ^{25}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× the computational 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 MM global minimum is considered for calculations at the higher level of theory. Significantly lower errors are obtained for ^{13}C chemical shifts of CH_2 and CH_3 groups,

while ^{1}H chemical shifts are insensitive to conformer sampling regardless of the type of configuration to which the proton belongs. Therefore, an evaluation of the impact of conformational sampling on DFT predictions for a nucleus of interest should be done whenever possible to boost the confidence in correlations established between experiments and the results of the computational framework described in this work. Such a process would determine whether the conformer issue is critical in the examined case study to possibly avoid instances of multiple conformers that would need to be considered.

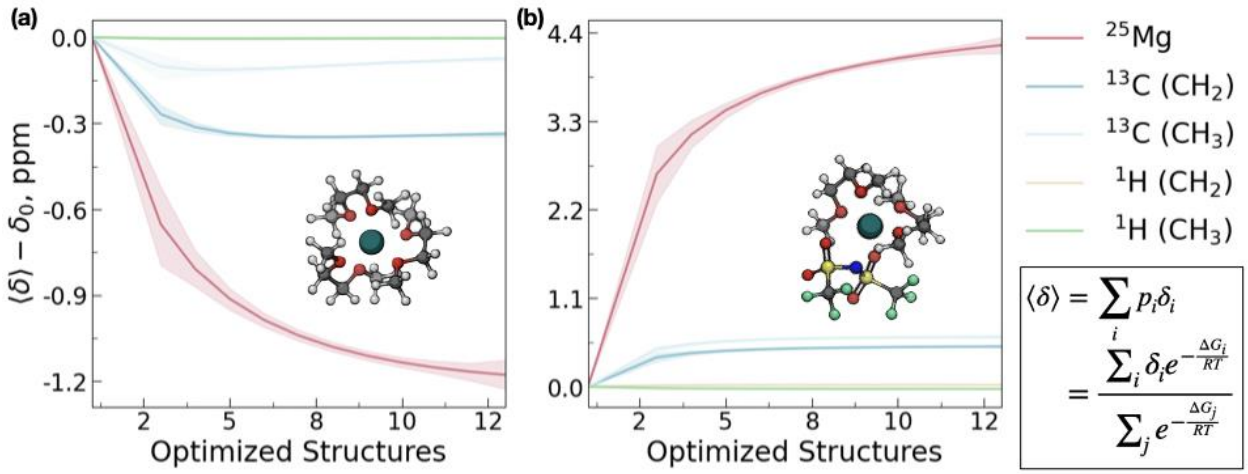

Fig. 5 Effect of multiple conformers for (a) configuration 1 and (b) configuration 2 on ^{25}Mg, ^{13}C, and ^{1}H NMR chemical shifts. In the equation of the Boltzmann weighted average of the chemical shift $\langle\delta\rangle$, p_i and ΔG_i are the Boltzmann weight and the formation energy of structure i relative to the most stable configuration predicted by DFT, respectively

CONCLUSIONS

In conclusion, we have developed and tested a computational framework that couples first-principle calculations with CMD simulations to robustly and efficiently calculate, analyze, 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 Edisonian approach in selecting possible solvation structures and the significant time required for manual file management, data collection, and error handling. By overcoming these limitations, we were able

to accurately identify multiple stable species present in the solution that contribute to the overall NMR spectral shape. Minimal inputs comprising structures of species in solution and their force field parameters are required to obtain accurate shifts, but the calculation procedure can be tuned by overriding default inputs like the level of theory and solvation model. Factors such as the choice of the force field used to identify the type of speciation in solution, DFT level of theory, implicit solvation model, and conformer sampling are critical in determining the accuracy of predictions made by the framework. We have successfully applied the framework to calculate chemical shifts in a complex multicomponent $Mg(TFSI)_2$/DME solution and resolved the discrepancy in the literature regarding the Mg^{2+} solvation structure in this solution. Our results show formation of solvent separated ion pairs in this electrolyte which is consistent with the experimental NMR results reported in this work and the previously reported SCXRD results[26]. The benchmark test case shows that our procedure can generate reliable results that can facilitate NMR deconvolution assignments to determine ionic association interactions within liquid solutions similar to those reported in this work. An extension of this framework is under development and will be successfully added to the existing one. Features that will be supported include the ability to explore the role of 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 monovalent and multivalent electrolytes whose structure is not intuitive or when the chemical and parameter spaces are too large for human search using conventional non-automated methods. Data collected from the framework is expected to provide fingerprints to guide future experimental investigations of liquid solutions with optimal properties.

531 **METHODS**

532 **Computational**

533 CMD simulations are performed using the LAMMPS simulation package[34] version 3Mar2020
534 (http://lammps.sandia.gov). Initial configurations of ions in the solvent are first obtained by
535 randomly packing the molecules in a cubic box of size $5 \times 5 \times 5$ nm^3 with periodicity in XYZ
536 directions using the PACKMOL package[56]. We consider $MgTFSI_2$ in DME at a salt-to-solvent
537 ratio of 1:18. In FF1, *i.e.,* GAFF[36] parameterization, $TFSI^-$ and DME parameters are obtained by
538 first generating the electrostatic potential of single molecules in Gaussian 16[33] at the B3LYP/6-
539 31+G* level of theory and fitting the electrostatic potential surface of the optimized structures
540 using the RESP method in Antechamber[57]. AMBER force field parameters by Aqvist are used for
541 Mg cations[58]. FF2, corresponding to the OPLS[36] force field, uses TFSI bonded parameters by
542 Lopes/Pádua[59] and nonbonded parameters by Köddermann[60]. DME parameters are taken from the
543 work of Anderson and Wilson[61] except for the parameters of C-C-O-C and O-C-C-O dihedrals,
544 which are based on GAFF parameterization[36]. Lastly, based on FF2, we build FF3
545 parameterization that includes polarization effects via the classical Drude oscillators model[45,62].
546 Drude particles are attached to all atoms, excluding hydrogen and Mg^{2+} due to their relatively small
547 polarizabilities. Atomic polarizabilities and charges for TFSI are based on the APPLE&P force
548 field[63], whereas those for DME are taken from work on poly(ethylene oxide)[64]. Nonbonded
549 parameters for Mg^{2+} cations are adapted from AMOEBA-PRO-13-FF[65]. Force field parameters
550 used in this work are listed in Tables S3-S7.

551 Lennard Jones interactions are truncated at a cutoff distance of 1.2 nm, and the particle-particle
552 particle-mesh (PPPM)[66] method is used to handle long-range electrostatic interactions. With FF3,
553 a Thole damping factor[67] of 1.0 is used to smear the neighboring induced dipoles located on the

same molecule and prevent the 'polarization catastrophe'[68]. Initial structures are subjected to a two-step energy minimization, first using the steepest descent algorithm employing convergence criteria of 1,000 kcal/mol Å and then using a conjugated-gradient minimization scheme with an energy convergence criteria of 10 kcal/mol Å. The two-step minimization allows for the release of strained contacts in the initial configuration. Isothermal-isobaric simulations (NPT) are performed to obtain the correct density on the minimized system using a Nosé/Hoover temperature thermostat and pressure barostat to maintain the temperature at 298.15 K and the pressure at 1 atm for 2 ns. With FF3, Drude particles are thermalized at a lower temperature relative to Drude cores to avoid fast vibrations of the small reduced masses, thus allowing the use of a reasonable time step. The system is then melted to 500.15 K for 2 ns and subsequently quenched to 298.15 K for 3 ns. Following that, canonical ensemble (NVT) simulations are performed for 50 ns using a time step of 0.001 ps at 298.15 K to equilibrate the system. Molecular trajectories are sampled every 5 ps, resulting in 10,000 snapshots, from which properties of interest are calculated.

All DFT calculations are performed using Gaussian 16[33]. Magnetic shieldings are calculated for the extracted Mg^{2+} clusters, ranging in size from 33 to 78 atoms. The benchmark study is performed with twelve combinations of functionals and basis sets (Fig 4) 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)[69] with the B3LYP, M06-2X, and PBE1PBE functionals. Besides the explicit solvent model used in this work, bulk solvent effects are described using a continuum model, particularly PCM[47-49] or SMD[54]. Following the optimization and frequency steps, magnetic response calculations are performed using the gauge-independent atomic orbital (GIAO)[70,71] method at the same level of theory. Chemical shifts are converted to the experimentally observed scale using $\delta_{cluster} = \sigma_{ref} - \sigma_{cluster}$, where $\delta_{cluster}$

577 and $\sigma_{cluster}$ are the chemical shift and the isotropic shielding constant of the nucleus of interest in
578 a given cluster, respectively, and σ_{ref} is the calculated isotropic shielding constant of the same
579 nucleus in a suitable reference compound. We use an Mg^{2+} ion coordinated octahedrally by six
580 water molecules, dimethyl sulfoxide, and water, as the chemical references for ^{25}Mg, ^{13}C, and ^{1}H,
581 respectively. To reduce systematic errors, we use secondary references (TMS) by adding 39.5 and
582 4.7 ppm to the calculated chemical shifts of carbon and proton, respectively. These values
583 correspond to the experimental chemical shifts of the secondary references relative to the primary
584 standards. We again stress that all the steps described here are automated within our computational
585 framework except for the polarizable CMD simulations.

586 **Experimental**

587 $Mg(TFSI)_2$ (99.5%, Solvionic) were dried for 48 hours under vacuum at 180 °C, and the DME
588 solvent (Battery-grade, Gotion) was further dried over activated 3 Å molecular sieves in a
589 glovebox until its water content was determined to be below 10 ppm using a Karl-Fisher Titrator
590 (Metrohm). $Mg(TFSI)_2$/DME solutions were prepared inside a glovebox filled with nitrogen right
591 before NMR measurements. ^{1}H and ^{13}C NMR measurements were performed on a Varian DDRS
592 spectrometer with a 17.6 T magnet using a broad-band (BBO) probe with ^{1}H and ^{13}C Larmor
593 frequencies of 748.1 and 188.1 MHz, respectively. The 90˚ pulse widths were 16 µs for ^{1}H and 16
594 µs for ^{13}C. ^{1}H spectra were collected using 30° pulses with a transition number of 16 and a recycle
595 delay of 20 s with a coaxial tube holding $Mg(TFSI)_2$/DME solution and an outer NMR tube holding
596 D_2O (99.9%, from Sigma Aldrich) as an external reference at 4.77 ppm. ^{13}C spectra were collected
597 using 30° pulses with averaging of 1024 transients and a recycle delay of 12 s using a thin-wall 5
598 mm NMR tube. ^{25}Mg NMR spectra were collected at a 14.1 T magnet (Varian DDR spectrometer)
599 with a ^{25}Mg Larmor frequency of 36.7 MHz and a 90˚ pulse width of 20 µs. A small tip angle of

15° with a recycle delay of 0.1 s was used and 128,000 transients were acquired. In order to minimize the spectrometer drift effect on chemical shift, DMSO-d6 and 5 M $MgCl_2$ were used to reference ^{13}C (39.52 ppm) and ^{25}Mg (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 are available in the repository at https://github.com/rashatwi/nmr-dataset.

CODE AVAILABILITY

The open-source LAMMPS-code is used in the CMD simulations while the proprietary Gaussian-code is primarily used in the DFT calculations. The framework shown in Fig 1 is implemented using the MISPR infrastructure, which defines, executes, manages, and stores DFT and CMD workflows. The codes used in this work will be made publicly available with the future release of the MISPR package.

ACKNOWLEDGMENTS

The authors thank Xiaohui Qu (Brookhaven National Laboratory) for the helpful discussions. 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). Work at Pacific Northwest National Laboratory was supported by the Joint Center for Energy Storage Research (JCESR), an Energy Innovation Hub funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences.

AUTHOR CONTRIBUTIONS

RA developed the automated NMR framework and the underlying Python-based codes, performed all the necessary calculations, and had primary writing responsibilities. YC and KSH carried the NMR experiments. VM and KTM guided the experimental aspect of the project. NNR guided and led all 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:

Correspondence:

Correspondence to Nav Nidhi Rajput.

REFERENCES

1 Zhao, E. W. *et al.* In situ NMR metrology reveals reaction mechanisms in redox flow batteries. *Nature* **579**, 224-228 (2020).

2 Kim, H. S. *et al.* Structure and compatibility of a magnesium electrolyte with a sulphur cathode. *Nature communications* **2**, 1-6 (2011).

3 Weber, R. *et al.* Long cycle life and dendrite-free lithium morphology in anode-free lithium pouch cells enabled by a dual-salt liquid electrolyte. *Nature Energy* **4**, 683-689 (2019).

4 Korb, J.-P., Louis-Joseph, A. & Benamsili, L. s. Probing structure and dynamics of bulk and confined crude oils by multiscale NMR spectroscopy, diffusometry, and relaxometry. *The Journal of Physical Chemistry B* **117**, 7002-7014 (2013).

5 Van de Velde, F., Knutsen, S., Usov, A., Rollema, H. & Cerezo, A. 1H and 13C high resolution NMR spectroscopy of carrageenans: application in research and industry. *Trends in Food Science & Technology* **13**, 73-92 (2002).

6 Shimada, I., Ueda, T., Kofuku, Y., Eddy, M. T. & Wüthrich, K. GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures. *Nature Reviews Drug Discovery* **18**, 59-82 (2019).

7 Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663-668 (2020).

8 Deng, X. *et al.* Nuclear magnetic resonance studies of the solvation structures of a high-performance nonaqueous redox flow electrolyte. *Journal of Power Sources* **308**, 172-179 (2016).

9 Cresce, A. V. *et al.* Solvation behavior of carbonate-based electrolytes in sodium ion batteries. *Physical Chemistry Chemical Physics* **19**, 574-586 (2017).

10 Wang, F. *et al.* Highly reversible zinc metal anode for aqueous batteries. *Nature materials* **17**, 543-549 (2018).

11 Harks, P., Mulder, F. & Notten, P. In situ methods for Li-ion battery research: A review of recent developments. *Journal of power sources* **288**, 92-105 (2015).

12 Plewa-Marczewska, A., Kalita, M., Marczewski, M. & Siekierski, M. NMR studies of equilibriums in electrolytes: Ionic pairing in glymes. *Electrochimica Acta* **55**, 1389-1395 (2010).

13 Barone, V., Improta, R. & Rega, N. Quantum mechanical computations and spectroscopy: From small rigid molecules in the gas phase to large flexible molecules in solution. *Accounts of chemical research* **41**, 605-616 (2008).

14 Rajput, N. N., Seguin, T. J., Wood, B. M., Qu, X. & Persson, K. A. in *Modeling Electrochemical Energy Storage at the Atomic Scale* 79-124 (Springer, 2018).

15 Hu, J. Z., Jaegers, N. R., Hu, M. Y. & Mueller, K. T. In situ and ex situ NMR for battery research. *Journal of Physics: Condensed Matter* **30**, 463001 (2018).

16 Dracinsky, M., Möller, H. M. & Exner, T. E. Conformational sampling by ab initio molecular dynamics simulations improves NMR chemical shift predictions. *Journal of chemical theory and computation* **9**, 3806-3815 (2013).

17 Abella, L., Philips, A. & Autschbach, J. Ab initio molecular dynamics study of sodium NMR chemical shifts in the methylamine solution of [Na+[2.2. 2] cryptand Na−]. *Physical Chemistry Chemical Physics* **23**, 339-346 (2021).

18 Casabianca, L. B. & De Dios, A. C. Ab initio calculations of NMR chemical shifts. *The Journal of chemical physics* **128**, 052201 (2008).

19 Hu, M. Y. *et al.* In situ natural abundance 17O and 25Mg NMR investigation of aqueous Mg (OH) 2 dissolution in the presence of supercritical CO2. *Environmental science & technology* **50**, 12373-12384 (2016).

20 Hu, J. Z. *et al.* 25Mg NMR and computational modeling studies of the solvation structures and molecular dynamics in magnesium based liquid electrolytes. *Nano energy* **46**, 436-446 (2018).

21 Yesiltepe, Y. *et al.* An automated framework for NMR chemical shift calculations of small organic molecules. *Journal of cheminformatics* **10**, 1-16 (2018).

22 Gao, P., Zhang, J., Peng, Q., Zhang, J. & Glezakou, V.-A. General protocol for the accurate prediction of molecular 13C/1H NMR chemical shifts via machine learning augmented DFT. *Journal of Chemical Information and Modeling* **60**, 3746-3754 (2020).

23 Chen, Y. *et al.* Role of Solvent Rearrangement on Mg2+ Solvation Structures in Dimethoxyethane Solutions using Multimodal NMR Analysis. *The Journal of Physical Chemistry Letters* **11**, 6443-6449 (2020).

24 Wan, C. *et al.* Natural abundance 17O, 6Li NMR and molecular modeling studies of the solvation structures of lithium bis (fluorosulfonyl) imide/1, 2-dimethoxyethane liquid electrolytes. *Journal of Power Sources* **307**, 231-243 (2016).

25 Rajput, N. N., Qu, X., Sa, N., Burrell, A. K. & Persson, K. A. The coupling between stability and ion pair formation in magnesium electrolytes from first-principles quantum mechanics and classical molecular dynamics. *Journal of the American Chemical Society* **137**, 3411-3420 (2015).

26 Salama, M. *et al.* Unique behavior of dimethoxyethane (DME)/Mg (N (SO2CF3) 2) 2 solutions. *The Journal of Physical Chemistry C* **120**, 19586-19594 (2016).

27 *SDBSWeb*, <https://sdbs.db.aist.go.jp>.

28 Fulmer, G. R. *et al.* NMR chemical shifts of trace impurities: common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist. *Organometallics* **29**, 2176-2179 (2010).

29 MongoDB Inc., M., 2014.

30 Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314-319 (2013).

31 Jain, A. *et al.* FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience* **27**, 5037-5059 (2015).

32 *Custodian*, <https://github.com/materialsproject/custodian>.

33 Gaussian 16 Rev. C.01 (Wallingford, CT, 2016).

34 Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics* **117**, 1-19 (1995).

35 O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33 (2011).

36 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **25**, 1157-1174 (2004).

37 Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of computational chemistry* **17**, 490-519 (1996).

38 Landrum, G. RDKit: Open-Source Cheminformatics Software. (2016).

39 Chai, J.-D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of chemical physics* **128**, 084106 (2008).

40 Rajput, N. N. *et al.* Elucidating the solvation structure and dynamics of lithium polysulfides resulting from competitive salt and solvent interactions. *Chemistry of Materials* **29**, 3375-3379 (2017).

41 Blau, S., Spotte-Smith, E., Wood, B., Dwaraknath, S. & Persson, K. Accurate, Automated Density Functional Theory for Complex Molecules Using On-the-fly Error Correction. (2020).

42 Fadel, E. R. *et al.* Role of solvent-anion charge transfer in oxidative degradation of battery electrolytes. *Nature communications* **10**, 1-10 (2019).

43 Borodin, O., Behl, W. & Jow, T. R. Oxidative stability and initial decomposition reactions of carbonate, sulfone, and alkyl phosphate-based electrolytes. *The Journal of Physical Chemistry C* **117**, 8661-8682 (2013).

44 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. *Journal of the American Chemical Society* **118**, 11225-11236 (1996).

45 Lamoureux, G. & Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *The Journal of chemical physics* **119**, 3025-3039 (2003).

46 Kubisiak, P. & Eilmes, A. Solvation of Mg2+ Ions in Mg (TFSI) 2–Dimethoxyethane Electrolytes—A View from Molecular Dynamics Simulations. *The Journal of Physical Chemistry C* **122**, 12615-12622 (2018).

47 Miertuš, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chemical Physics* **55**, 117-129 (1981).

48 Miertus, S. & Tomasi, J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chemical physics* **65**, 239-245 (1982).

49 Pascual-ahuir, J.-L., Silla, E. & Tunon, I. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *Journal of Computational Chemistry* **15**, 1127-1138 (1994).

50 Beck, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys* **98**, 5648-5646 (1993).

51 Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical chemistry accounts* **120**, 215-241 (2008).

52 Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of chemical physics* **110**, 6158-6170 (1999).

53 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. *Physical Chemistry Chemical Physics* **7**, 3297-3305 (2005).

54 Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B* **113**, 6378-6396 (2009).

55 Andersen, A. *et al.* Structure and dynamics of polysulfide clusters in a nonaqueous solvent mixture of 1, 3-dioxolane and 1, 2-dimethoxyethane. *Chemistry of Materials* **31**, 2308-2319 (2019).

56 Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* **30**, 2157-2164 (2009).

57 Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* **25**, 247-260 (2006).

58 Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* **94**, 8021-8024 (1990).

59 Canongia Lopes, J. N. & Pádua, A. A. Molecular force field for ionic liquids composed of triflate or bistriflylimide anions. *The Journal of Physical Chemistry B* **108**, 16893-16898 (2004).

60 Köddermann, T., Paschek, D. & Ludwig, R. Molecular dynamic simulations of ionic liquids: A reliable description of structure, thermodynamics and dynamics. *ChemPhysChem* **8**, 2464-2470 (2007).

61 Anderson, P. M. & Wilson*, M. R. Developing a force field for simulation of poly (ethylene oxide) based upon ab initio calculations of 1, 2-dimethoxyethane. *Molecular Physics* **103**, 89-97 (2005).

62 Dequidt, A., Devemy, J. & Padua, A. A. Thermalized Drude oscillators with the LAMMPS molecular dynamics simulator. *Journal of chemical information and modeling* **56**, 260-268 (2016).

63 Borodin, O. Polarizable force field development and molecular dynamics simulations of ionic liquids. *The Journal of Physical Chemistry B* **113**, 11463-11478 (2009).

64 Borodin, O. & Smith, G. D. Development of quantum chemistry-based force fields for poly (ethylene oxide) with many-body polarization interactions. *The Journal of Physical Chemistry B* **107**, 6801-6812 (2003).

65 Shi, Y. *et al.* Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of chemical theory and computation* **9**, 4046-4063 (2013).

66 Hockney, R. W. & Eastwood, J. W. *Computer simulation using particles*. (crc Press, 1988).

67 Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **59**, 341-350 (1981).

68 Borodin, O. *et al.* Insights into the structure and transport of the lithium, sodium, magnesium, and zinc bis (trifluoromethansulfonyl) imide salts in ionic liquids. *The Journal of Physical Chemistry C* **122**, 20108-20121 (2018).

69 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. *The Journal of chemical physics* **132**, 154104 (2010).

70 Ditchfield, R. Self-consistent perturbation theory of diamagnetism: I. A gauge-invariant LCAO method for NMR chemical shifts. *Molecular Physics* **27**, 789-807 (1974).

71 Wolinski, K., Hinton, J. F. & Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *Journal of the American Chemical Society* **112**, 8251-8260 (1990).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIAnautomtedframeworkforhighthroughputpredictionsofNMRchemicalshiftswithinliquidsolutions.pdf
- flatRajputepc.pdf