QM Assisted ML for ^{19}F NMR Chemical Shift Prediction

Patrick Penner*, Anna Vulpetti*

Biomedical Research, Novartis AG, 4056 Basel, Switzerland

*Corresponding Authors. Emails:
patrick.penner@novartis.com
anna.vulpetti@novartis.com

ORCIDs:
Patrick Penner (0000-0003-4988-6183)
Anna Vulpetti (0000-0002-3114-8679)

Abstract

Ligand-observed ^{19}F NMR detection is an efficient method for screening libraries of fluorinated molecules in fragment-based drug design campaigns. Screening fluorinated molecules in large mixtures makes ^{19}F NMR a high-throughput method. Typically, these mixtures are generated from pools of well-characterized fragments. By predicting ^{19}F NMR chemical shift, mixtures could be generated for arbitrary fluorinated molecules facilitating for example focused screens. In a previous publication, we introduced a method to predict ^{19}F NMR chemical shift using rooted fluorine fingerprints and machine learning (ML) methods. Having observed that the quality of the prediction depends on similarity to the training set, we here propose to assist the prediction with quantum mechanics (QM) based methods in cases where compounds are not well covered by a training set. Beyond similarity, the performance of ML methods could be associated with individual features in compounds. A combination of both could be used as a procedure to split input data sets into those that could be predicted by ML and those that required QM processing. We could show on a proprietary fluorinated fragment library, known as LEF (Local Environment of Fluorine), and a public Enamine data set of ^{19}F NMR chemical shifts that ML and QM methods could synergize to outperform either method individually. Models built on Enamine data, as well as model building and QM workflow tools, can be found at https://github.com/PatrickPenner/lefshift and https://github.com/PatrickPenner/lefqm.

Introduction

Fragment-based screening (FBS) has become a common hit-finding approach for drug discovery. Small and well-characterized fragment libraries typically deliver higher hit rates of ligands with lower molecular weight and higher solubility than those derived from high-throughput screening approaches, making FBS an attractive technology.[1] Since the inception of fragment-based technology in medicinal chemistry, nuclear magnetic resonance (NMR) methods have played, and continue to play, a significant role in the screening of fragments.[2][3]

A particularly attractive ligand-observed NMR technique uses ^{19}F as the NMR active isotope. This method known as FAXS[4] is very sensitive to protein binding, uses low concentrations of fragments, and requires low amounts of unlabeled protein. It can screen and deconvolute large mixtures of compounds due to the wide range of ^{19}F NMR chemical shift, can identify reporter/spy molecules that can then be used to screen molecules not containing fluorine, and measure the binding affinity of the identified hits.[4] The possibility of screening large mixtures makes ^{19}F NMR a high-throughput method.

Establishing and deconvoluting large mixtures of fluorinated compounds typically requires these compounds to have been measured in single experiments, so that their measured ^{19}F NMR chemical shifts can be used as references.[4] This can be done once for a static library of fragments but quickly becomes prohibitive if such a screening library is to be constructed on the fly as a focused set for a particular target or using parallel chemistry. Every compound would have to be measured at least once to generate mixtures of non-overlapping ^{19}F NMR peaks.

In our previous work, we used a machine learning (ML) method to predict the ^{19}F NMR chemical shifts of compounds for focused ^{19}F NMR screenings we termed ^{19}Focused screens.[5] That method was based on a Random Forest architecture but other architectures have been explored for ^{13}C and ^{1}H NMR[6] or even for solid-state NMR[7]. In the analysis of the predictions it became evident that the performance of the ML method was dependent on the similarity of the predicted molecules to the training data set. This was particularly pronounced in the extension of our in-house LEF4000 library by another 1530 molecules, which were chosen to explore new chemical space, and were therefore often dissimilar to the LEF4000 data being used as a training set. The strong outliers this generated would have been very difficult to deconvolute in a mixture of a ^{19}Focused screening.

Quantum mechanics (QM) calculations are a largely training set independent collection of methods that can be used to predict chemical shifts. Their performance in predicting specifically ^{19}F NMR chemical shifts has been increasing steadily.[8][9] Several density functional theory (DFT) methods have shown a favorable balance of speed and accuracy,[8] especially after a correction by linear regression.[9]

In this work, we will further formalize our usage of ML methods to predict ^{19}F NMR chemical shift and then extend it with QM methodology. To that end we will first describe the limits of the ML method's applicability domain. Those compounds that are outside this will then be split off and handled orthogonally by QM. This will involve a more physical description of the molecules that may synergize with the purely empirical ML workflow to increase overall performance, but more importantly to prevent strong outliers.

Methods

Data Sets

Two data sets were used in the course of this work. The first data set was a collection of ^{19}F NMR chemical shifts from our in-house library of fluorinated fragments, the expansion of which from around 4000 (referred to as LEF4000) to around 5500 (referred to as LEF5500) fluorinated molecules was described in the previous publication.[5] The expansion to the LEF5500 represented not only a time-split, but also an expansion of the chemical space contained in the data set, which offered a valuable opportunity for assessing ^{19}F NMR chemical shift prediction.

Molecules with two fluorines at one carbon atom (CF_2), as opposed to molecules with one fluorine at one carbon atom (CF) or three fluorines at one carbon atom (CF_3), had not been considered in the prediction analysis of the last publication. The reason CF_2 had not been a part of model building previously is that, although the CF_3 motif displays a ^{19}F NMR signal which is a singlet, the case of molecules containing CF_2 is more complex.[10] In the prediction of CF_2 containing molecules by ML and QM, we limit the data to those cases in which there is fast exchange with respect to the chemical shift difference between the fluorines in the NMR time scale so that the two fluorines are magnetically equivalent. However, this is not always the case even for the CF_2H motif for which rapid internal rotation is expected. This results in ^{19}F NMR spectra with the two fluorines having different chemical shifts and appearing as doublets due to the splitting arising from the large two-bond $^{2}J_{FF}$ scalar coupling constant. The relative intensities of the two components of each doublet will depend on the ratio of the chemical shift difference of the two doublets, expressed in Hz, divided by J_{FF}. In the approaches described here, we are dealing with CF_2 displaying only one single signal or two peaks very close to each other. This is also a pre-requisite for the averaging of equivalent fluorines in the QM workflow below.

The second data set that we used were the ^{19}F NMR chemical shifts of a combination of diverse fluorinated fragments available from Enamine.[11] Enamine provides two separate plated sets of fluorinated fragments called: “FDS-1000” and “FFL-d6” containing 1000 and 1280 compounds, respectively (Version: 21 May 2021). Both sets have an overlap with the LEF5500, which was used to analyze how consistent the ^{19}F NMR chemical shifts were across data sets measured by two different institutions and two different protocols (See section Experimental Details).

Data Splitting Strategies

The initial ML model selection and validation were performed on the LEF4000. The QM assisted ML workflow was also initially established on the LEF data and then applied to Enamine data. Most results are reported in the main body of the manuscript using the CF containing molecules, the most abundant fluorinated motif, to emphasize the statistical effect. Corresponding results for CF_2 and CF_3 can be found in the supporting information.

The split into tranche 1 and tranche 2 of the LEF expansion was retained from the previous publication and used in our analysis. Typically, models were trained on the LEF4000 and then used to predict tranche 2. Tranche 1 was occasionally used to augment the training set and expose models to a part of the chemical space that the expansion to the LEF5500 had explored. Of the approximately 300 CF_2 containing molecules in the LEF5500 all that had not been part of the LEF4000 before the expansion were added to tranche 2 to perform time-split experiments on these motifs as well. We maintained the split of Enamine data into the FDS-1000 and the FFL-d6 as a train/test split and all overlapping molecules were added to the FDS-1000, the training set.

Experimental Details

Enamine provided ^{19}F NMR chemical shifts measured in 1 mM PBS and reported a KF ^{19}F NMR chemical shift as a standard for every measurement.[11] Occasionally Enamine reports two peaks for a single compound. In these cases, the first peak is taken as a representative. At Novartis the LEF ^{19}F NMR chemical shifts were recorded in 50 mM deuterated Tris buffer at pH 7.6, containing 100 mM NaCl, 100 µM sodium trimethylsilylpropanesulfonate (DSS), and 2% DMSO-d6. Measurements were performed at 296 K (23 °C) with DSS as indirect referencing for ^{19}F NMR chemical shift calibration.

Rooted Fluorine Fingerprint Descriptor

The core of our in-house ^{19}F NMR chemical shift prediction throughout the years have been rooted fluorine fingerprints. They were described in Vulpetti et al. 2009[12] as a variant of the path-based topological torsion count fingerprint using atomic numbers, number of π electrons, and number of heavy atom neighbors to generate bits. All paths up to a certain path length contribute to the set and the counts of bits in the fingerprints. This means a fingerprint with a path length of 7 will count all bits of the paths of length 7, 6, 5, and so on. Because fluorine fingerprints are count fingerprints, they will be compared using Dice similarity, which compares counts as well as whether a bit is present, instead of the more common Tanimoto similarity used for other fingerprints, such as extended connectivity fingerprints. Deviating from the original description, the fluorine fingerprints used in this publication always start their path at a fluorine atom. In the original publication, paths started at a fluorine atom for the CF motif and from the carbon of a trifluoromethyl group for the CF_3 motif. The path length of the fingerprint was optimized as a hyperparameter (See Table S1 of the supporting information).

Machine Learning Methods

Besides the random forest (RF) model shown in the previous publication[5] a number of other architectures were evaluated for the task of ^{19}F NMR chemical shift prediction. Lasso regression was chosen as a reasonable linear baseline. K-nearest neighbor (KNN) regression has been applied to this problem in the past[13] and it was included here as a reference. Random forest and gradient-boosted trees (GBT) serve as similar model architectures that have already been used with some success.[5][14]

The XGBoost GBT implementation was used.[15] Lastly, we also evaluated a model based on Chemprop[16] as a deep-learning representative.

Separate versions of all models were built for the three fluorine motifs: CF, CF_2, and CF_3. Models were selected based on their hyperparameter optimized performance when trained on the LEF4000 data set. Not all models were parametrized using cross-validation, but all parametrized models were eventually run in a cross-validation scenario for selection. The regularization parameter of Lasso regression was optimized using coordinate descent as implemented in the scikit-learn Python package.[17] KNN regression, RF, and GBT model parameters were optimized using a grid search and evaluated in 5-fold cross-validation. Chemprop models were hyperparameter optimized using the provided method based on Bayesian optimization, which internally performs 80:10:10 splits.[16] All hyperparameter configurations can be found in Table S1 of the supporting information.

Quantum Mechanics Setup

The full QM workflow is visualized in Figure 1. Molecules were prepared for QM calculation using several software packages. RDKit[18] was used to strip salts and assign stereochemistry when possible. Unassigned stereocenters were assigned an arbitrary configuration if they were the only stereocenters present in a molecule. Unassigned diastereomers were rejected at this point. The most abundant tautomer and protomer were chosen using MoKa.[19][20][21] Molecules were only used in their charged form for the calculation if they were predicted to be at least 90% charged at physiological pH.

Conformer ensembles were generated using the Conformator program[22] and optimized with GFN2-xTB in water using the ALPB solvent model[23]. These optimized conformers were reduced to a maximum of 15 clusters with the k-means algorithm implemented in scikit-learn[17] and the lowest-energy representative was picked from each cluster. The optimized and clustered conformer ensembles were the input for the DFT calculations.

Turbomole[24] was used to perform NMR shielding constant calculations by first performing a single-point DFT calculation and then deriving the isotropic shielding constants using the Gauge-Including Atomic Orbital (GIAO) method[25] with def2-TZVP[26] as the basis set and the KT3[27] functional. The calculation was performed in water as modeled by the COSMO[28][29] implementation in Turbomole. An example Turbomole control file is given in Section 2 of the supporting information.

After shielding constants had been calculated for all fluorine atoms of all conformers of each molecule their values were combined. At first, the shielding constants for equivalent fluorines (CF_2, CF_3) were averaged by arithmetic mean. Shielding constants of fluorines of different conformers were Boltzmann averaged together using their xTB energies as weights. This left every molecule with one shielding constant for every equivalent set of fluorines. Fluorine shielding constants were converted to chemical shifts with linear regression as in Dumon et al.[8] QM calculations were performed in parallel to ML predictions, so the linear regression was fitted to the same training set as the corresponding ML method.

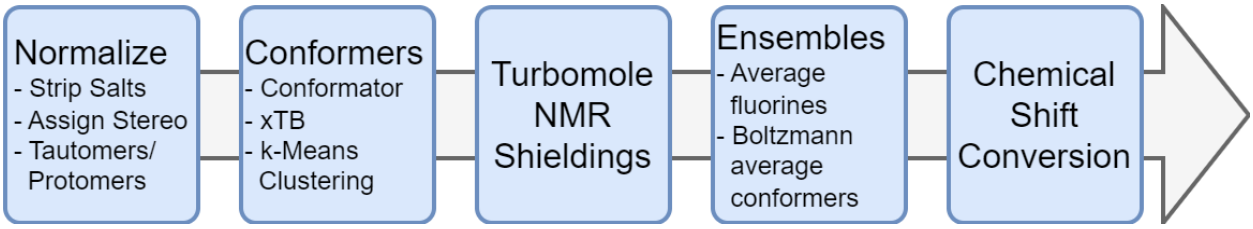

Figure 1 The full QM workflow. Molecules are normalized, followed by conformer generation with xTB optimization. Turbomole is used to calculate NMR shielding constants. The shielding constants for ensembles are combined so that every molecule has one shielding constant for every equivalent set of fluorines. The last step converts shielding constants into chemical shifts using a linear regression correction fitted to a corresponding ML training set.

QM Assisted ML Workflow

Figure 2 outlines the workflow of QM assisted ML. The decision which molecules were out-of-distribution (OOD), i.e. out of the applicability domain, for a ML model was made using two metrics: the Dice similarity of the predicted molecule to the training set of the model using fluorine fingerprints and the number of bits of a molecule's fluorine fingerprint not in the training set of the model. The thresholds of both metrics were optimized for each separate fluorine motif. The LEF4000 was used as the training set and tranche 2 as the test set. Parameters were explored using a grid search. The grid definition can be found in Table S2 of the supporting information. ^{19}F NMR chemical shifts from the QM calculations for molecules outside of the ML model's applicability domain were fed back into the ML model's training set. The ^{19}F chemical shifts were then predicted with a model trained on the augmented training set.

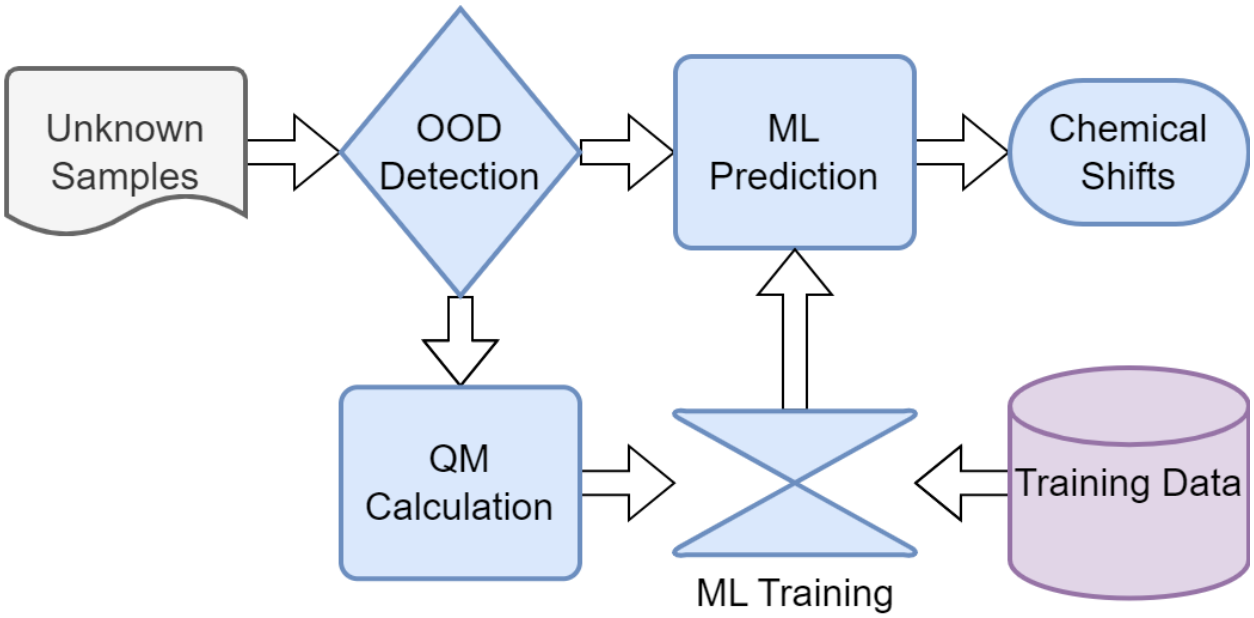

Figure 2 QM assisted ML workflow. Samples with unknown ^{19}F NMR chemical shifts to predict are first fed into out-of-distribution (OOD) detection. OOD samples are processed by QM and the QM ^{19}F NMR chemical shifts are fed into the ML training set. Thus, chemical shift prediction is performed by a ML model trained on experimental ^{19}F NMR chemical shifts as well as QM derived ^{19}F NMR chemical shifts.

Results

Data Analysis

The LEF5500 library-derived data set and the two Enamine data sets were compared. Figure 3 shows the populations of the different fluorine motifs in the different data sets. The LEF5500 (Figure 3a) was the largest of the data sets, larger than both Enamine data sets put together. The LEF5500 also had the highest proportion of CF_3 containing compounds in comparison to the two Enamine data sets. The LEF5500 also extended most into the -130 to -230 ppm range, which was a direct result of the expansion strategy described in the previous publication.[5] The two Enamine data sets, FDS-1000 (Figure 3b) and the FFL-d6 (Figure 3d), had different distributions of fluorinated motifs. All data sets had a similarly low proportion of CF_2, the least represented fluorine motif, which is likely due to its NMR appearance described above.

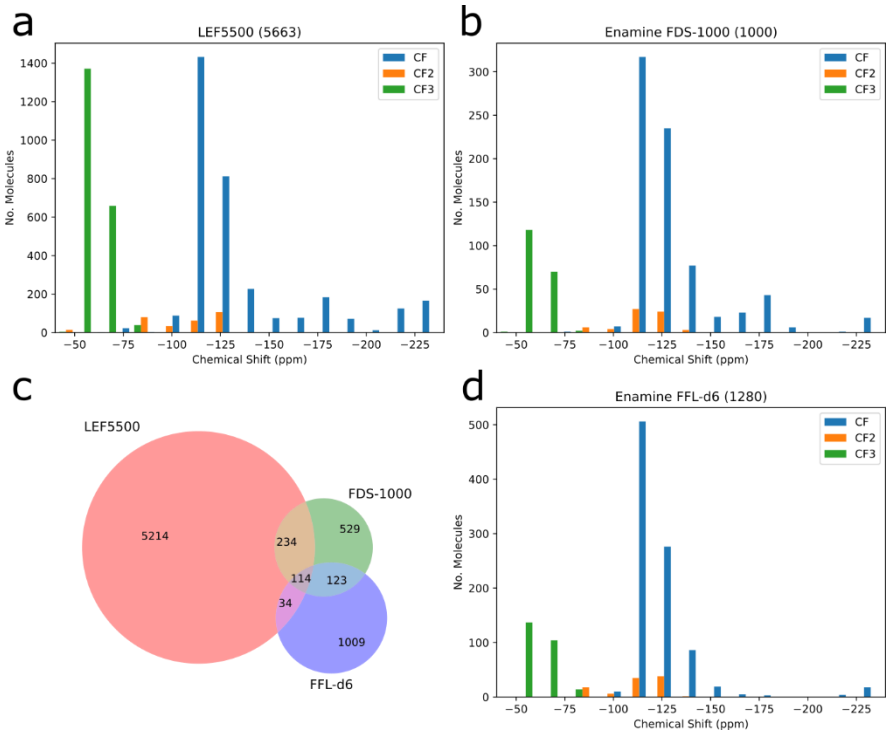

Figure 3 Fluorine motif composition for the three data sets used: (a) LEF5500, (b) FDS-1000, and (d) FFL-d6 and the number of overlapping compounds in (c). The total number of compounds for each data set is given in the diagram title.

Figure 3c shows the number of compounds the three data sets had in common. The largest overlap of 348 molecules was between the FDS-1000 and the LEF5500. There was an overlap of 237 molecules between the FDS-1000 and FFL-d6.

We used the overlap between the LEF5500 and the FDS-1000 to investigate how comparable the measured ^{19}F NMR chemical shifts were across data sets. Both data sets were in very good agreement as can be seen in Figure 4. There were two molecules with very different chemical shifts. Further discussion of these outliers as well as the chemical shift deviations in general can be found in Section 3 of the Supporting Information. The mean absolute deviation for the CF motif was 0.27 ppm. This was substantially worse than the possible deviation of less than 0.03 ppm reported by Rosenau et al. [30], but lower than the deviation expected of two data sets with different experimental procedures and different referencing.

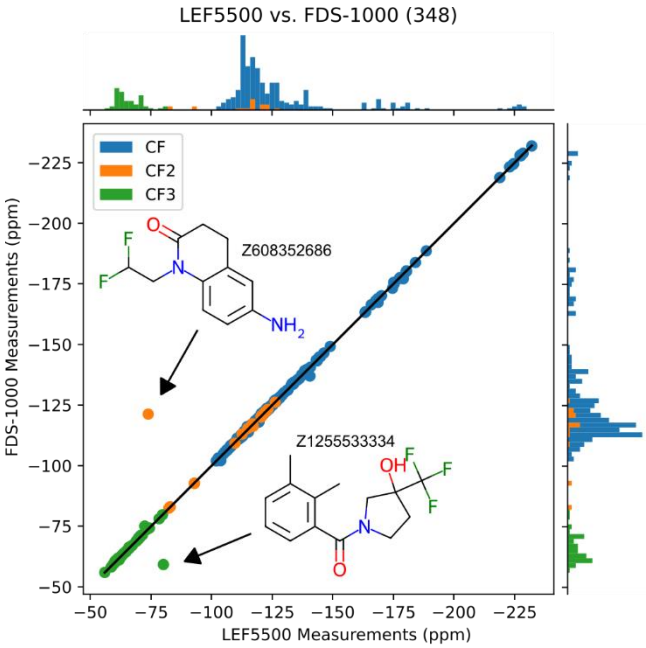

Figure 4 Agreement of ^{19}F NMR chemical shifts between the LEF5500 library and the Enamine FDS-1000. Two outliers, Enamine molecules Z1255533334 and Z608352686, are annotated.

Model Selection on the LEF Data Set

The performance of all models parametrized on the LEF4000 is reported in Table 1. A large part of the problem of predicting ^{19}F NMR chemical shifts appeared to be linear as demonstrated by the high performance of Lasso regression. This is in line with older methods that attempted to model chemical shifts as linear combinations of substructure contributions with impressive success.[31] KNN tended to degenerate into a single nearest neighbor search by cosine similarity. When trained on all data, the best performing KNN models only considered one neighbor (see Table S3 of the supporting information).

Table 1. Cross-validation Model Performance

Architecture	CF (RMSE [a], ppm)	CF_2 (RMSE [a], ppm)	CF_3 (RMSE [a], ppm)
Lasso Regression (LR)	4.868 ± 1.343	6.252 ± 3.112	1.946 ± 0.433
K-nearest Neighbor (KNN)	6.204 ± 0.756	6.305 ± 4.341	2.128 ± 0.447
Random Forest (RF)	4.415 ± 1.103	6.329 ± 2.961	1.926 ± 0.425
Gradient-Boosted Trees (GBT)	4.341 ± 1.281	**5.985 ± 3.220**	1.932 ± 0.435
Chemprop	**4.068 ± 1.561**	12.695 ± 3.367	**1.833 ± 0.864**

[a] Average root-mean-square errors (RMSEs) of 5-fold cross-validation ± their standard deviation on the LEF4000 for all parameterized models.

The top-performing models were RF, GBT, and Chemprop, which performed quite similarly. The most difference between the three models was observed for the CF_2 motif, which had the least training data. The higher standard deviations across the 5-fold cross-validation for CF and CF_3, as well as the erratic learning curve (Figure S4 of the supporting information) shed doubt on the slightly better performance of Chemprop in CF and CF_3. This doubt was substantiated when all parameterized models were used to predict molecules with CF motifs from tranche 2 after being trained on the full LEF4000. Table 2 shows the top-performing models outperforming lasso regression and KNN more clearly, with the two tree-based models taking the lead in mean absolute error (MAE). Chemprop manages to outperform RF in root-mean-square error (RMSE) but is outperformed by GBT in all metrics. The GBT model was determined to be the most effective method at this point.

Table 2. Time-split Model Performance

Architecture	Kendall's Tau	MAE[a] (ppm)	RMSE[b] (ppm)
Lasso Regression (LR)	0.8398	5.662	11.047
K-nearest Neighbor (KNN)	0.7724	6.386	13.150
Random Forest (RF)	0.8475	4.806	9.525
Gradient-Boosted Trees (GBT)	**0.8648**	**4.361**	**8.561**
Chemprop	0.8480	5.539	9.184

Kendall's Tau, [a]Mean absolute error (MAE), and [b]root-mean-square error (RMSE) of the models predicting the ^{19}F NMR chemical shifts on CF motifs of tranche 2 when trained with the LEF4000.

Importance of Features in ML Performance

The previous publication assessed how the performance of a RF model is related to the Dice similarity of fluorine fingerprints in a test set to the training set.[5] A RF model was trained on the LEF4000, with or without tranche 1, and used to predict tranche 2. It yielded good predictions of the ^{19}F NMR chemical shift for those molecules with at least one close analog of a Dice similarity greater than 0.7 in the training set. We here report the performance of predicting tranche 2 after training with the LEF4000 with or without tranche 1 as a function of substructural features present in the fluorine fingerprints.

Using the feature importance metrics provided by XGBoost[15], specifically the "gain" metric, we could measure the average accuracy gain of splits in the decision trees using a particular feature. The feature with the highest gain importance resembled a substructure of a fluorine bound to an aromatic substructure. The difference between the average chemical shift of all aliphatic CF groups to all aromatic CF groups was around 60 - 70 ppm, demonstrating that the model had extracted a very impactful feature. Typically, fluorines bound to an sp^3 hybridized carbon generate signals in the more shielded (i.e., the more negative) region of the ^{19}F NMR spectrum than those bound to sp^2 hybridized carbon.

An analysis of the changes of feature importance when training with the LEF4000 and tranche 1 as opposed to just the LEF4000 suggested two main reasons for the performance difference. The first reason was the appearance of new features in tranche 1 and tranche 2 that had not been present in the LEF4000. As explained above, the expansion of the LEF4000 was intended to explore new chemical space, which added new fluorine environments encoded as fingerprint bits into the set. Figure 5a visualizes a CF substructure feature present in tranche 1 and tranche 2 but not present in the LEF4000.

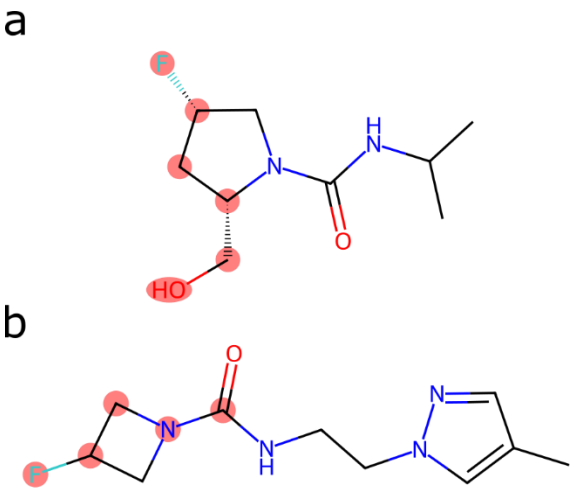

Figure 5 (a) is a fluorine fingerprint path that makes up a bit only present in tranche 1 and tranche 2 highlighted in red in the Enamine molecule Z2051927854 (^{19}F NMR chemical shift: -169.34 ppm). (b) is a fluorine fingerprint path that makes up a feature that becomes more important after the addition of tranche 1 visualized in Enamine molecule Z1658083843 (^{19}F NMR chemical shift: -180.63 ppm)

The closest analogs to the fluoropyrrolidine substructure carrying the hydroxymethyl shown in Z2051927854 in the LEF4000 were four fluoropyrrolidines, which had acids or amides in place of the primary alcohol as well as a different stereochemical configuration. These significant structural differences in the closest analogs of the LEF4000 resulted in poor predictions of the ^{19}F NMR chemical shift. The highlighted fingerprint feature was the one with the highest gain feature importance after the addition of tranche 1 that described this substructure. Other fingerprint features with non-zero feature importance after training with the LEF4000 and tranche 1 also encoded parts of the fluoropyrrolidine. A combination of these bits was likely used by the GBT model to predict the ^{19}F NMR chemical shift. Training with only the LEF4000 resulted in predictions with a MAE of around 6 ppm for these fluoropyrrolidines in tranche 2. Training with the LEF4000 and tranche 1 managed to improve this to a MAE of around 1 ppm by providing more examples of these bits to the GBT.

Another reason for the performance difference was the re-weighting of features present in the LEF4000 after the addition of tranche 1. Figure 5b visualizes a feature that became more important. Although this feature was present in the LEF4000, the ^{19}F NMR chemical shift of molecules that contained it in tranche 2 were predicted with a high MAE of greater than 5 ppm after training with only the LEF4000. Less than 1% of the molecules in the LEF4000 contained this bit and even those could be quite different to the ones in tranche 1 and tranche 2. The percentage of molecules that contained this bit in tranche 1 was higher, at around 5%, which made this bit more important by augmenting its relative frequency. Training with the LEF4000 as well as tranche 1 once again improved the MAE of these molecules to around 1 ppm.

There was substantial variability in the substructure that the bit in Figure 5b crossed. Some of the molecules in the LEF4000 that contained this bit had larger rings, such as the molecule in Figure 5a. Some did not have a ring in that position. The fingerprint does not encode whether an atom is in a ring and therefore the substructure it matched could be a chain. Other bits also re-weighted along with this bit did contain the full azetidine ring. This suggested a more subtle effect of clusters of bits encoding different parts of the same substructure being re-weighted to better fit the data. Because the fluorine fingerprint does not contain stereochemical or ring information, the path from Figure 5b can also be found in Figure 5a. The appearance of the fingerprint feature in Figure 5a and the re-weighting of the feature in Figure 5b will have influenced each other in model training.

Feature-based Out-of-Distribution Detection

The relationship between features and the performance of the GBT model could be used to detect cases where the prediction of the ^{19}F NMR chemical shift would be poor. Figure 6 shows the prediction performance on CF motifs of a GBT model trained on the LEF4000 generating chemical shifts for tranche 2. A grid search determined that CF containing compounds with a Dice similarity of less than 0.8 or at least 3 unknown bits in its fingerprint should be sorted into a subset of samples with unknown features (see Table S4 of the supporting information). The subset of compounds with only known features (Figure 6a) is visibly better predicted than the subset with unknown features (Figure 6b). Roughly a third of the data set made up the subset of known features. A little less than two-thirds of all samples were sorted into the unknown features subset.

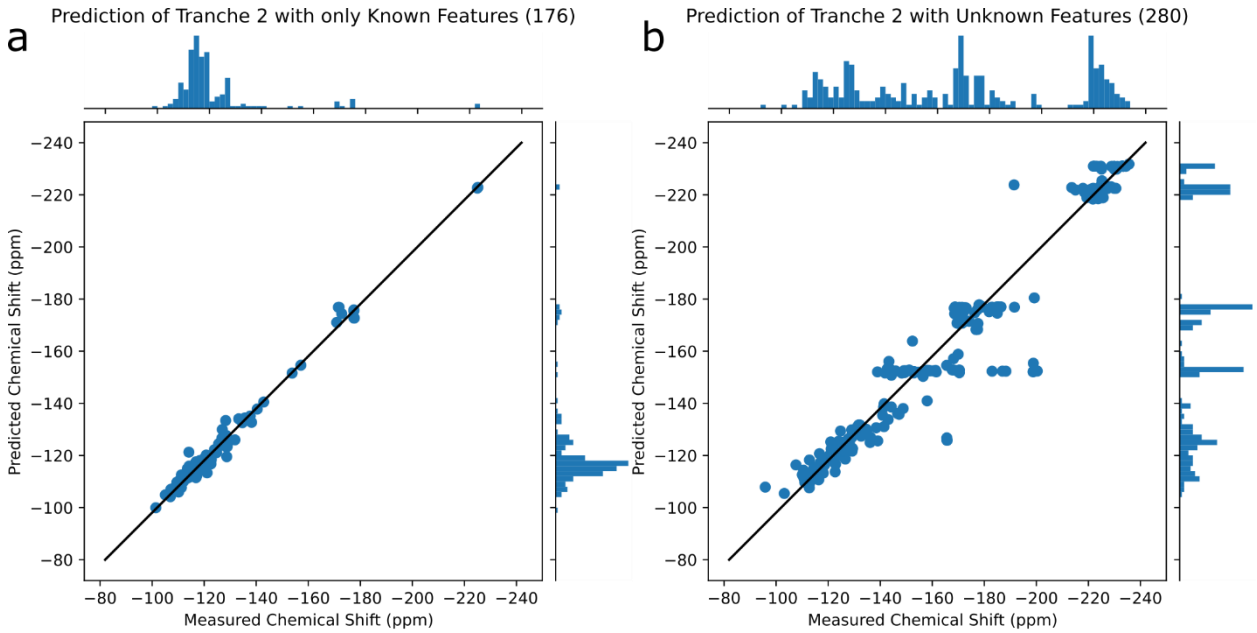

Figure 6 Performance of the GBT model on CF motifs when trained on the LEF4000 and predicting tranche 2. The data was split into subsets of known and unknown features with the size of the subset in the title. The subset with only known features (a) is visibly better predicted than the subset with unknown features (b).

Horizontal clusters of compounds in the prediction of the subset with unknown features suggested that the trees in the GBT model did not know how to split some samples and therefore predicted similar ^{19}F NMR chemical shifts for them. The features that would have facilitated this split appeared to have not been in the training set. The higher density of test cases in the unknown features subset in the -130 ppm to -230 ppm range reflected the expansion of the LEF4000 to the LEF5500 and its explicit goal to expand into new chemical space, by embracing alkyl fluorinated motifs.[5] Table S7 in the supporting information shows that the unknown features subset was predicted with almost 5-fold higher MAE and RMSE than the known features subset. It was obvious that the unknown features subset was outside of the applicability domain of the GBT model and required an orthogonal method.

Quantum Mechanics Assisted Machine Learning

LEF Library: LEF4000 and Tranche 2

Not all molecules could be processed by QM mainly due to missing stereochemistry information. Around 10% of the whole LEF5500 data set had ambiguous stereochemistry annotation. Only unassigned enantiomeric compounds could be rescued because their isomers are indistinguishable in NMR with achiral solvent and the stereocenter could therefore be assigned arbitrarily. Those molecules that could be processed were used in the QM assisted ML workflow and the subsequent comparison.

In general, the ^{19}F NMR chemical shifts calculated by the QM method for tranche 2 were better than the chemical shifts predicted by GBT. QM chemical shifts appeared to generalize more effectively. Although the overall trend was towards QM generating less error, comparing QM to GBT on the known and unknown features subsets revealed that GBT was better on the known features subsets and QM was better on the unknown features subset (see Table S7 of the supporting information). This supported our idea that these methods could synergize. Adding QM predictions of the subset with unknown features into the training set for GBT in addition to the LEF4000 outperformed GBT and QM in all metrics as shown in Table 3. The difference in mean-squared-error (MSE), the parameter most sensitive to strong outliers, between GBT and QM assisted GBT on LEF data was statistically significant in a two-sided t-test ($p = 0.008$).

Table 3. Performance of GBT, QM, and QM assisted GBT on CF

Method	Kendall's Tau	MAE[a] (ppm)	RMSE[b] (ppm)
Trained on LEF4000, predicted tranche 2			
GBT	0.8763	3.858	7.464
QM	0.8830	3.623	5.347
QM assisted GBT	**0.9014**	**3.180**	**4.965**
Trained on FDS-1000, predicted FFL-d6			
GBT	**0.7882**	1.818	3.231
QM	0.7607	2.562	3.677
QM assisted GBT	0.7855	**1.814**	**3.094**

Kendall's Tau, [a]Mean absolute error (MAE), and [b]root-mean-square error (RMSE) of GBT, QM, and QM assisted GBT.

Enamine data sets: FDS-1000 and FFL-d6

Having established the QM assisted ML method on the LEF, we applied it to the Enamine data sets by using the FDS-1000 as the training set and the FFL-d6 as the test set. Table 3 also shows the performance for GBT, QM, and QM assisted GBT on Enamine data.

QM by itself performed worse in every metric in an inversion of the trend seen before. Generally, the ML-based method produced lower MAE and RMSE. This suggested that the split between the two Enamine data sets did not represent as much of an extrapolation as the split from the LEF4000 to tranche 2. The split of the test data set, the FFL-d6, into the subset with known features versus the subset with unknown features also supported this. Only a quarter of the FFL-d6 CF samples were sorted into the subset with unknown features as opposed to around two-thirds of samples in tranche 2 of the LEF (see Table S8 of the supporting information).

Figure 7 shows a plot of the similarity between the LEF and Enamine test/train splits. The similarity is calculated as the maximum similarity between all molecules of a test set to any molecule of the training set using the Dice similarity of fluorine fingerprints of length 7 (F-FP-7) and the Tanimoto similarity of extended connectivity fingerprints with a radius of 2 (ECFP4). The FFL-d6 (the Enamine test set) contains only very few molecules with a maximum F-FP-7 Dice similarity lower than 0.6 to molecules of the FDS-1000 (the Enamine training set). Tranche 2 (the LEF test set) on the other hand contains a substantial fraction of molecules with a maximum F-FP-7 Dice similarity lower than 0.6 to molecules of the LEF4000 (the LEF training set). A higher number of dissimilar molecules in tranche 2 suggests that this train/test split required more extrapolation.

The ECFP4 similarity in Figure 7 is given as a reference to a fingerprint describing the whole molecule. It supports that the proportion of dissimilar molecules in tranche 2 is higher than in the FFL-d6. It is interesting to note that all molecules in both sets are quite dissimilar using the ECFP4. The F-FP-7, which is rooted at the fluorine, not only on average finds training molecules with a greater similarity for test molecules but also produces in a more even distribution of similarity.

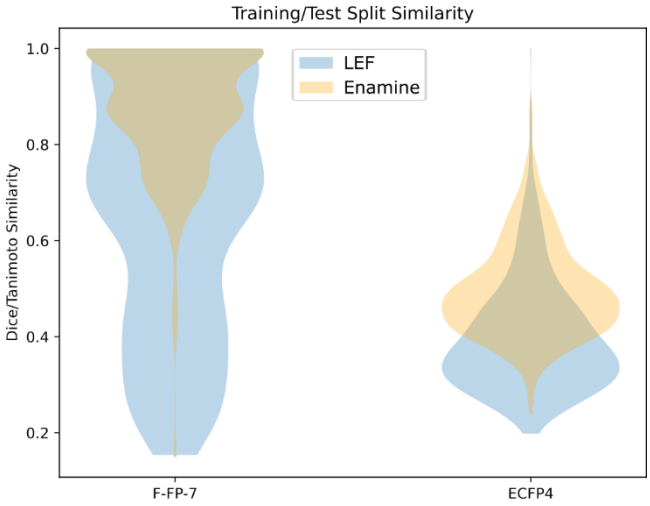

Figure 7 Similarity violin plots for the train/test splits of the LEF4000 to tranche 2 and the FDS-1000 to the FFL-d6 using fluorine fingerprints of length 7 (F-FP-7) and extended connectivity fingerprints with a radius of 2 (ECFP4). The LEF library-derived data is in blue, and the Enamine data is in orange. The distributions are made up of the maximum Dice similarity for the F-FP-7 and the maximum Tanimoto similarity for the ECFP4 of molecules in the test set to molecules in the training set.

Despite the lower overall performance of QM and the lower number of samples processed by QM in the split of the Enamine set, the synergistic QM assisted GBT method still performed better than GBT or QM by themselves in RMSE. The difference in MSE between QM and QM assisted GBT was statistically significant ($p = 0.019$). The MAE did not change much in comparison to GBT but was improved in comparison to QM. The Kendall's Tau dropped slightly in comparison to GBT, which highlights the importance of splitting data sets as precisely as possible. Processing everything by QM would have resulted in lower overall performance in this case. Note that the splitting that was performed on the FFL-d6 used the same Dice similarity and number of unknown bits thresholds as the one performed on tranche 2 of the LEF. The parametrization of the sorting into known and unknown subsets on the time-split in the LEF had been stable enough in this case to extrapolate to a different data set. Results for CF_2 and CF_3 can be found in Section 6 of the Supporting Information.

Runtime Comparison

Training a model on the full Enamine training set took around 3 minutes in total using a single core. The prediction of the full data Enamine test set took around half a minute using a single core. Splitting the Enamine test set, which is effectively a fingerprint comparison of the test set and the training set, took around 5 seconds on a single core. By default, training and prediction are parallelized across 8 cores.

Conformer generation with xTB optimization and clustering took around 4 minutes per molecule, and produced a median of 15 conformers. This meant that over half of the compounds produced the maximum number of conformers. Shielding constant calculations took around 55 minutes per ensemble of conformers. A rough estimate of the effort involved to process the Enamine test set therefore comes to about 1000 CPU hours. The calculations are trivially parallelizable and run overnight on our internal computer cluster. Nonetheless, it is important to note that only processing OOD samples in the Enamine test set reduces the computational load down to one quarter of that.

Characterization of Synergistic Effects

To analyze the synergy between QM and GBT we first generated a learning curve for both using the CF motifs in Enamine data. Figure 8 shows learning curves for GBT and the linear regression correction that converts QM shielding constants into chemical shifts. Both were trained on increasing amounts of randomly sampled FDS-1000 data and used to predict FFL-d6 compounds. The linear regression correction of the QM method showed very little dependence on the training set. After 100 randomly

sampled training data points the RMSE of the linear regression correction changed by less than 1%, with the largest changes ending after 50 points. GBT continually improved with training set size and eventually outperformed QM after around 500 data points. This suggested that in situations with less data, our QM workflow is more robust, but it is eventually outperformed by ML.

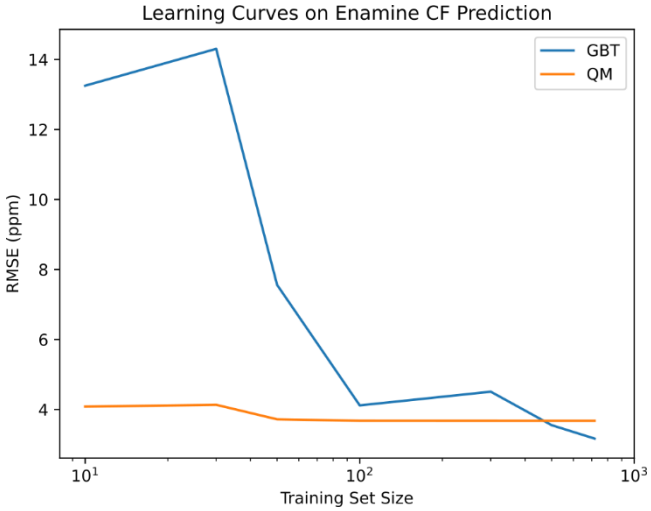

Figure 8 Learning curves for GBT and the QM linear regression correction trained on randomly sampled FDS-1000 data and predicting the full FFL-d6. Note the logarithmic scale on the training set size. Only the CF containing samples are plotted. The GBT model was reparametrized using 5-fold cross-validation at each point.

QM differs from the ML methods in several fundamental ways. One is that QM by its use of molecular geometries can discriminate between stereoisomers. Figure 9a shows all four possible stereoisomers for a molecule with two stereocenters. According to the QM calculation, the CF_3 motif can have ^{19}F NMR chemical shifts ranging from -62.98 ppm to -73.83 ppm depending on which of the four stereoisomers is used as an input. The cis-(2S, 3R) stereoisomer, shown in Figure 9b, was measured to have a shift of -70.34 ppm and QM predicted it as -72.75 ppm. The rooted fluorine fingerprint does not include stereochemistry and would therefore not be able to distinguish between stereoisomers.

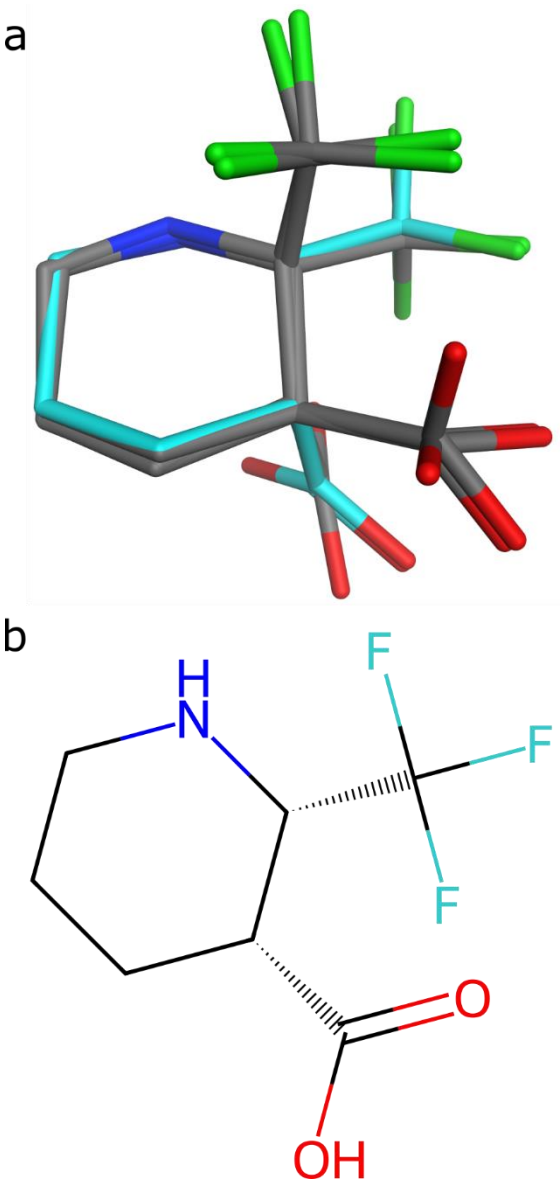

Figure 9 (a) Enumerated stereoisomers and (b) the reported stereoisomer of Enamine molecule Z2718608919. The most likely conformer by xTB energy is visualized. Hydrogens have been omitted for clarity. The cis/trans positioning of the acid and the CF_3 group to each other are likely the cause of the large ppm range seen in the QM calculation.

Another aspect that QM handles is conformer effects. Figure 10 shows a population plot of conformers as well as the corresponding conformers for a CF motif. There was a significant difference in chemical shift between the most likely and the least likely conformers in the ensemble. None of the chemical shifts of the conformers coincided with the measured chemical shift. The most likely (i.e., the lowest energy) conformer was associated with a chemical shift of -131.44 ppm compared to the measured chemical shift of -134.66 ppm, an error greater than 3 ppm. A Boltzmann average over the conformer ensemble produced a better chemical shift prediction of -135.60 ppm, which is an error of around 1 ppm. This highlights the importance of sampling more than just the lowest energy conformation for precise ^{19}F NMR chemical shift predictions.

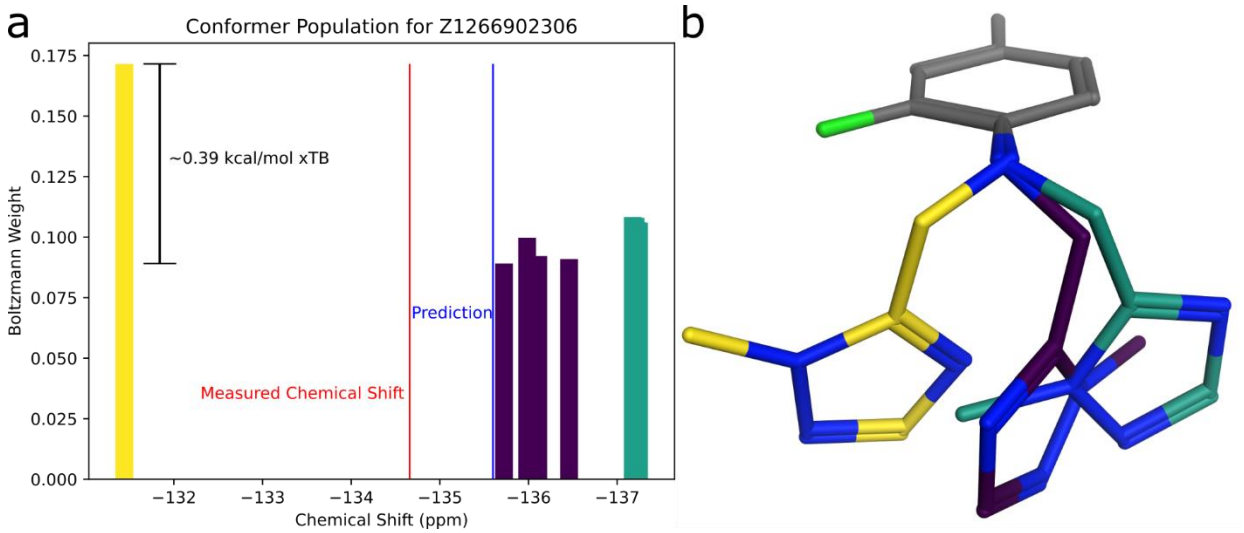

Figure 10 Conformer population effects on the calculation of ^{19}F NMR chemical shift for Enamine molecule Z1266902306. (a) shows a plot of conformer clusters and the chemical shift that would be calculated for them, as well as their Boltzmann weights (i.e., their likelihood) in water. A red line shows the measured chemical shift and a blue line the prediction. A range in black between the highest and lowest Boltzmann weight denotes what the difference in likelihood equates to in kcal/mol at 298.15 K (25 °C) calculated by xTB. Three non-symmetric conformers that represent the conformer clusters are shown in (b). Hydrogens have been omitted for clarity. The carbon atoms of the different side-chain conformers are colored by their xTB energy in water, where darker colors signify higher and brighter colors signify lower energies. The colors of the conformers correspond to the distributions in chemical shift and Boltzmann weights in the plot that they represent.

As an example of the QM workflow's robustness, Figure 11 shows a collection of quinolines from the FFL-d6 (Figure 11 top, test set) and their close analogs in the FDS-1000 (Figure 11 bottom, training set). The two indoles in the training set have a greater similarity by fluorine fingerprint Dice similarity because the larger substituent of the quinoline from the training set is attached at a different position than in the test set. For reference, the F-FP-7 Dice similarity values can be found in Table S11 of the supporting information. This changes multiple bits in the fingerprint and makes it harder for the GBT model to accurately predict these examples. The FFL-d6 quinolines have a ^{19}F NMR chemical shift in the range of -107.73 ppm to -109.93 ppm and are all predicted at around -119.47 ppm to -120.87 ppm, a consistent 10 ppm error. The QM workflow can more accurately assess the nature of the ring systems involved and consider the effect that changes in substitution make, which leads to better predictions of -108.37 ppm to -111.76 ppm for the FFL-d6 quinolines. The QM workflow is not as dependent on approximations as the ML method and can therefore generalize better.

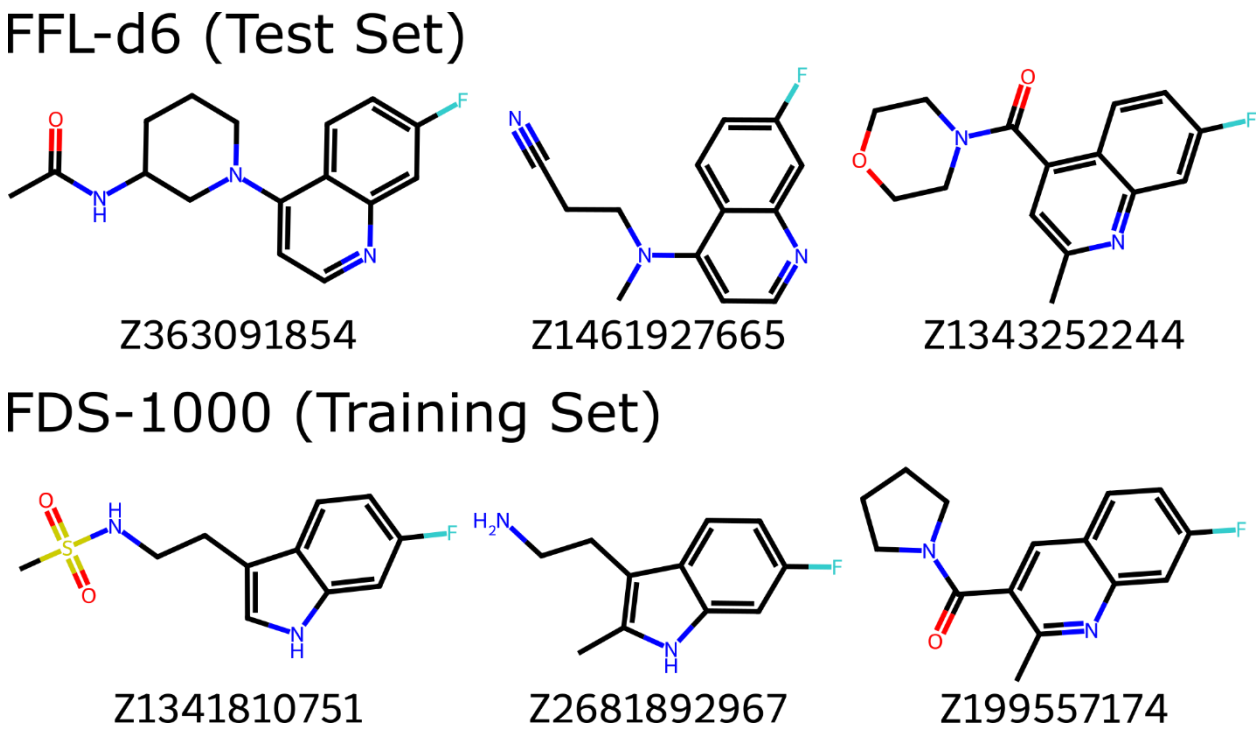

Figure 11 At the top poorly predicted quinolines from the FFL-d6 present in the test set and at the bottom their close analogs from the FDS-1000 present in the training set. The indoles from the FDS-1000 have a higher fluorine fingerprint Dice similarity to the test set molecules than the quinoline Z199557174 in the bottom because the larger substituent at the quinoline is attached at a different position than in the test set molecules. This means that the indoles from the FDS-1000 also have more bits in common with the quinolines from the FFL-d6, which probably confounds the ML prediction.

Discussion

ML-based workflows using the rooted fluorine fingerprint achieve high correlations in the prediction task. Even though the MAEs are still orders of magnitude higher than achievable experimental errors[30], the predictions are already useful in application.[5] ML workflows are however expected to fail in situations when the predicted compounds contain features that are not covered by the training set. Some ML architectures managed to extrapolate better than others. Previously used models such as KNN were outperformed by more complex models. On the other hand, Chemprop was also outperformed, probably because the size of the training sets could not yet justify a deep-learning model. All models incurred a significant number of strong outliers. In these cases, a model derived from physics in the form of a QM-based workflow is more successful. The given QM workflow was able to generalize better in predicting ^{19}F NMR chemical shifts for fluorines in new local environments and was largely independent of a training set. It was on the other hand less precise on in-distribution samples compared to the ML workflow. Handling out-of-distribution samples by QM and in-distribution samples by ML led to better prediction performance and significantly fewer strong outliers. A few limitations and possible extensions of both the ML and QM methodology are discussed in the following paragraphs in more detail.

In this work, we focused our efforts on the in-house LEF library-derived data set and the publicly accessible Enamine data sets. We kept the two data sources and models trained on these sets separated due to the difference in their experimental conditions, despite observing very good fluorine chemical shift correlation in the overlap between both data sets. Evaluating whether and under what conditions data sets could be combined is key for extending available data sets. One essential part of this is consistent referencing of ^{19}F NMR chemical shift between institutions, which has been shown to lower deviations by

orders of magnitude.[30] In the absence of a consensus, reporting referencing methods may already facilitate corrections to correlate chemical shift measurements. ML model building could then leverage the availability of more ^{19}F NMR chemical shift data sets of fluorinated molecules in aqueous solution, for example, from other vendors or academic groups, to expand the applicability domain.

The efforts to understand the model's performance could also be extended to try and reverse engineer the most impactful features and build a fully human-interpretable model. Building a single decision tree, ideally at minimal loss of performance due to the reduction of model complexity, with the most impactful features could start to approximate the more complex GBT model. This would make patterns detected by the model more accessible, as well as generally improve the trust in the model.

The out-of-distribution detection presented here can be seen to lack some beneficial complexity. Splitting samples by number of unknown bits and Dice similarity cannot take into account the re-weighting of known features that we observed in the LEF expansion upon the addition of tranche 1. Furthermore, the OOD detection could be extended to specifically focus on phenomena that QM can handle and the ML method based on the fluorine fingerprint cannot, such as stereochemical effects.

Building up a ^{19}F NMR chemical shift prediction model from the ground up is a difficult task. Public data sets are a good starting point. Alternatively, a different starting point could use the presented QM workflow or a similar one to bootstrap such a model from minimal or no experimental data. While a model trained on QM results would not have the same precision as a model trained on experimental data, it could be a way to start predicting and screening mixtures. The ^{19}F NMR chemical shifts measured as a consequence could then be fed back into the model starting a continuous improvement cycle. Another option would be to train a model on QM generated data and then fine-tune on available experimental data in a transfer learning workflow[32].

The problem with generating QM data is the large amount of CPU time necessary. QM assisted ML as we have presented it, minimizes the number of QM calculations necessary for good-quality predictions. Integrating QM parameters directly into the model requires the users to perform QM calculations for all predictions they want to make. Decoupling these systems gives the user more freedom. The splitting and re-training time necessary for the presented workflow have negligible runtimes in comparison to the QM workflow and the user has full discretion about which samples to process by QM. The system is intentionally designed in such a way that even if samples cannot be processed by QM for whatever reason the ML system is still able to provide a prediction at baseline performance.

The QM workflow could also be extended methodologically by, for example, enumerating all diastereomers of a stereochemically unannotated molecule and predicting them all. This would initially approximate the expected error of a predicted ^{19}F NMR chemical shift due to stereochemistry. These calculation results could also be used to hypothesize what the actual stereochemistry may be. The presented QM workflow represents a cost-efficient approximation that is adequate for our application but with many avenues for further improvement. For example, one could use the Δ-learning paradigm recently applied to NMR chemical shift prediction[33] or replace the linear regression correction with a non-linear method such as has been done for ^{1}H and ^{13}C NMR chemical shift prediction[34].

In general, the QM assisted ML workflow demonstrates on one hand the difference between purely empirical and more physical models, as well as their synergistic potential. The purely empirical model was more successful at precise predictions in the chemical space that was covered in the training set. The QM workflow was more robust in extrapolation and less dependent on a training set. Knowing how and when to combine the two methodologies can lead to a workflow that benefits from the performance of both.

Acknowledgments
The authors would like to acknowledge Rainer Wilcken, Jimmy Kromann, and Hagen Münkler for their guidance in setting up QM calculations, Claudio Dalvit and Jochen Sieg for their comments on the manuscript, and the Novartis Innovation Postdoctoral Fellowship program for their support of P.P.

Author Contributions
Both authors contributed substantially to the methodology, analysis, and writing of the manuscript.

Competing Interests
The authors declare no competing interests with regard to this work.

Data and Software Availability
Results and models not built on the proprietary LEF data are available publicly. ^{19}F chemical shift prediction models built on Enamine data as well as tools to build these can be found at https://github.com/PatrickPenner/lefshift. This repository also contains a short tutorial on integrating QM derived ^{19}F chemical shifts into model building. A second repository that contains QM workflow tools to calculate ^{19}F chemical shifts can be found at https://github.com/PatrickPenner/lefqm. Both repositories contain the Enamine data used in this study as example data.

Supporting Information
- Supporting Information QM Assisted ML for 19F NMR Chemical Shift Prediction: Supporting information on the data sets, machine learning parameters and performance, the QM setup, and the results.
- Data.zip: Collection of public Enamine training data, test data, QM, and ML results as CSV files.

References
[1] Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H (2016) Twenty years on: the impact of fragments on drug discovery. Nat. Rev. Drug Discov. 15(9):605–619. https://doi.org/10.1038/nrd.2016.109

[2] Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534. https://doi.org/10.1126/science.274.5292.1531

[3] Walsh L, Erlanson DA, de Esch IJP, Jahnke W, Woodhead A, Wren E (2023) Fragment-to-lead medicinal chemistry publications in 2021. J. Med. Chem. 66(2):1137–1156. https://doi.org/10.1021/acs.jmedchem.2c01827

[4] Dalvit C, Vulpetti A (2019) Ligand-based fluorine NMR screening: principles and applications in drug discovery projects. J. Med. Chem. 62(5):2218–2244. https://doi.org/10.1021/acs.jmedchem.8b01210

[5] Vulpetti A, Lingel A, Dalvit C, Schiering N, Oberer L, Henry C, Lu Y (2022) Efficient screening of target-specific selected compounds in mixtures by 19F NMR binding assay with predicted 19F NMR chemical shifts. ChemMedChem 17(13):e202200163. https://doi.org/10.1002/cmdc.202200163

[6] Guan Y, Shree Sowndarya SV, Gallegos LC, St. John PC, Paton RS (2021) Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network. Chem. Sci. 12(36):12012–12026. https://doi.org/10.1039/d1sc03343c

[7] Paruzzo FM, Hofstetter A, Musil F, De S, Ceriotti M, Emsley L (2018) Chemical shifts in molecular solids by machine learning. Nature Communications 9(1):4501. https://doi.org/10.1038/s41467-018-06972-x

[8] Dumon AS, Rzepa HS, Alamillo-Ferrer C, Bures J, Procter R, Sheppard TD, Whiting A (2022) A computational tool to accurately and quickly predict 19F NMR chemical shifts of molecules with fluorine–carbon and fluorine–boron bonds. Phys. Chem. Chem. Phys. 24(34):20409–20425. https://doi.org/10.1039/D2CP02317B

[9] Saielli G, Bini R, Bagno A (2014) Computational 19F NMR. 2. organic compounds. RSC Adv. 4(78):41605–41611. https://doi.org/10.1039/c4ra08290g

[10] Vulpetti A, Dalvit C (2013) Design and generation of highly diverse fluorinated fragment libraries and their efficient screening with improved 19F NMR methodology. ChemMedChem 8(12):2057–2069. https://doi.org/10.1002/cmdc.201300351

[11] Enamine. *Fluorinated Fragment Library - Enamine*. https://enamine.net/compound-libraries/fragment-libraries/fluorinated-fragment-library (accessed 2023-05-04)

[12] Vulpetti A, Hommel U, Landrum G, Lewis R, Dalvit C (2009) Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. J. Am. Chem. Soc. 131(36):12949–12959. https://doi.org/10.1021/ja905207t

[13] Vulpetti A, Landrum G, Rüdisser S, Erbel P, Dalvit C (2010) 19F NMR chemical shift prediction with fluorine fingerprint descriptor. J. Fluorine Chem. 131(5):570–577. https://doi.org/10.1016/j.jfluchem.2009.12.024

[14] Lu Y, Anand S, Shirley W, Gedeck P, Kelley BP, Skolnik S, Rodde S, Nguyen M, Lindvall M, Jia W (2019) Prediction of pKa using machine learning methods with rooted topological torsion fingerprints: application to aliphatic amines. J. Chem. Inf. Model. 59(11):4706–4719. https://doi.org/10.1021/acs.jcim.9b00498

[15] Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 785–794

[16] Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, Guzman-Perez A, Hopper T, Kelley B, Mathea M, Palmer A, Settels V, Jaakkola T, Jensen K, Barzilay R (2019) Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model. 59(8):3370–3388. https://doi.org/10.1021/acs.jcim.9b00237

[17] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12(85):2825–2830

[18] Landrum G, Tosco P, Kelley B, Ric, Sriniker, Gedeck, Vianello R, Cosgrove D, NadineSchneider, Kawashima E, N D, Dalke A, Jones G, Cole B, Swain M, Turk S, AlexanderSavelyev, Vaucher A, Wójcikowski M, Take I, Probst D, Ujihara K, Scalfani VF, Godin G, Pahl A, Berenger F, JLVarjo, Strets, JP, DoliathGavid (2022) rdkit/rdkit: 2022_09_3 (Q3 2022) release

[19] Milletti F, Storchi L, Sforna G, Cruciani G (2007) New and original pKa prediction method using grid molecular interaction fields. J. Chem. Inf. Model. 47(6):2172–2181. https://doi.org/10.1021/ci700018y

[20] Milletti F, Storchi L, Sforna G, Cross S, Cruciani G (2009) Tautomer enumeration and stability prediction for virtual screening on large chemical databases. J. Chem. Inf. Model. 49(1):68–75. https://doi.org/10.1021/ci800340j

[21] Milletti F, Vulpetti A (2010) Tautomer preference in PDB complexes and its impact on structure-based drug discovery. J. Chem. Inf. Model. 50(6):1062–1074. https://doi.org/10.1021/ci900501c

[22] Friedrich N-O, Flachsenberg F, Meyder A, Sommer K, Kirchmair J, Rarey M (2019) Conformator: A novel method for the generation of conformer ensembles. J. Chem. Inf. Model. 59(2):731–742. https://doi.org/10.1021/acs.jcim.8b00704

[23] Bannwarth C, Ehlert S, Grimme S (2019) GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. J. Chem. Theory Comput. 15(3):1652–1671. https://doi.org/10.1021/acs.jctc.8b01176

[24] Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Electronic structure calculations on workstation computers: The program system turbomole. Chem. Phys. Lett. 162(3):165–169. https://doi.org/10.1016/0009-2614(89)85118-8

[25] Ditchfield R (1974) Self-consistent perturbation theory of diamagnetism. Mol. Phys. 27(4):789–807. https://doi.org/10.1080/00268977400100711

[26] Weigend F, Ahlrichs R (2005) 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(18):3297. https://doi.org/10.1039/b508541a

[27] Keal TW, Tozer DJ (2004) A semiempirical generalized gradient approximation exchange-correlation functional. J. Chem. Phys. 121(12):5654–5660. https://doi.org/10.1063/1.1784777

[28] Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc., Perkin Trans. 2 (5):799–805. https://doi.org/10.1039/p29930000799

[29] Schäfer A, Klamt A, Sattel D, Lohrenz JCW, Eckert F (2000) COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. Phys. Chem. Chem. Phys. 2(10):2187–2193. https://doi.org/10.1039/b000184h

[30] Rosenau CP, Jelier BJ, Gossert AD, Togni A (2018) Exposing the origins of irreproducibility in fluorine NMR spectroscopy. Angew. Chem. Int. Ed. 57(30):9528–9533. https://doi.org/10.1002/anie.201802620

[31] Fifolt MJ, Sojka SA, Wolfe RA, Hojnicki DS, Bieron JF, Dinan FJ (1989) A chemical shift additivity method for the prediction of fluorine-19 chemical shifts in fluoroaromatic compounds. J. Org. Chem. 54(13):3019–3023. https://doi.org/10.1021/jo00274a011

[32] Han H, Choi S (2021) Transfer Learning from Simulation to Experimental Data: NMR Chemical Shift Predictions. J. Phys. Chem. Lett. 12(14):3662–3668. https://doi.org/10.1021/acs.jpclett.1c00578

[33] Büning JBK, Grimme S (2023) Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT. J. Chem. Theory Comput. https://doi.org/10.1021/acs.jctc.3c00165

[34] Gao P, Zhang J, Peng Q, Zhang J, Glezakou V-A (2020) General protocol for the accurate prediction of molecular 13C/1H NMR chemical shifts via machine learning augmented DFT. J. Chem. Inf. Model. 60(8):3746–3754. https://doi.org/10.1021/acs.jcim.0c00388

Table of Contents Artwork