

# Machine learning of free energies in chemical compound space using ensemble representations: Reaching experimental uncertainty for solvation <sup>EP</sup>

Cite as: J. Chem. Phys. **154**, 134113 (2021); <https://doi.org/10.1063/5.0041548>

Submitted: 24 December 2020 • Accepted: 01 March 2021 • Published Online: 05 April 2021

 Jan Weinreich,  Nicholas J. Browning and  O. Anatole von Lilienfeld

## COLLECTIONS

Paper published as part of the special topic on [Computational Materials Discovery](#)

 This paper was selected as an Editor's Pick



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

### [Machine learning meets chemical physics](#)

The Journal of Chemical Physics **154**, 160401 (2021); <https://doi.org/10.1063/5.0051418>

### [FCHL revisited: Faster and more accurate quantum machine learning](#)

The Journal of Chemical Physics **152**, 044107 (2020); <https://doi.org/10.1063/1.5126701>

### [When do short-range atomistic machine-learning models fall short?](#)

The Journal of Chemical Physics **154**, 034111 (2021); <https://doi.org/10.1063/5.0031215>

The Journal  
of Chemical Physics

**SPECIAL TOPIC:** Low-Dimensional  
Materials for Quantum Information Science

Submit Today!



# Machine learning of free energies in chemical compound space using ensemble representations: Reaching experimental uncertainty for solvation

Cite as: J. Chem. Phys. 154, 134113 (2021); doi: 10.1063/5.0041548

Submitted: 24 December 2020 • Accepted: 1 March 2021 •

Published Online: 5 April 2021



View Online



Export Citation



CrossMark

Jan Weinreich,<sup>1</sup>  Nicholas J. Browning,<sup>2</sup>  and O. Anatole von Lilienfeld<sup>1,2,a)</sup> 

## AFFILIATIONS

<sup>1</sup>University of Vienna, Faculty of Physics, Kolingasse 14-16, AT-1090 Wien, Austria

<sup>2</sup>Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

**Note:** This paper is part of the JCP Special Topic on Computational Materials Discovery.

**a)** Author to whom correspondence should be addressed: [anatole.vonlilienfeld@univie.ac.at](mailto:anatole.vonlilienfeld@univie.ac.at)

## ABSTRACT

Free energies govern the behavior of soft and liquid matter, and improving their predictions could have a large impact on the development of drugs, electrolytes, or homogeneous catalysts. Unfortunately, it is challenging to devise an accurate description of effects governing solvation such as hydrogen-bonding, van der Waals interactions, or conformational sampling. We present a Free energy Machine Learning (FML) model applicable throughout chemical compound space and based on a representation that employs Boltzmann averages to account for an approximated sampling of configurational space. Using the FreeSolv database, FML's out-of-sample prediction errors of experimental hydration free energies decay systematically with training set size, and experimental uncertainty (0.6 kcal/mol) is reached after training on 490 molecules (80% of FreeSolv). Corresponding FML model errors are on par with state-of-the-art physics based approaches. To generate the input representation for a new query compound, FML requires approximate and short molecular dynamics runs. We showcase its usefulness through analysis of solvation free energies for 116k organic molecules (all force-field compatible molecules in the QM9 database), identifying the most and least solvated systems and rediscovering quasi-linear structure–property relationships in terms of simple descriptors such as hydrogen-bond donors, number of NH or OH groups, number of oxygen atoms in hydrocarbons, and number of heavy atoms. FML's accuracy is maximal when the temperature used for the molecular dynamics simulation to generate averaged input representation samples in training is the same as for the query compounds. The sampling time for the representation converges rapidly with respect to the prediction error.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0041548>

## I. INTRODUCTION

An accurate description of solvation free energy is fundamentally important to rationalizing reaction kinetics and product propensities. Therefore, accurate models describing solvation have far reaching utility from drug design to battery development. Computational methods for predicting solvation free energies based on *ab initio* methods,<sup>1–4</sup> while accurate, impose a substantial computational burden and are therefore inherently limited when it comes

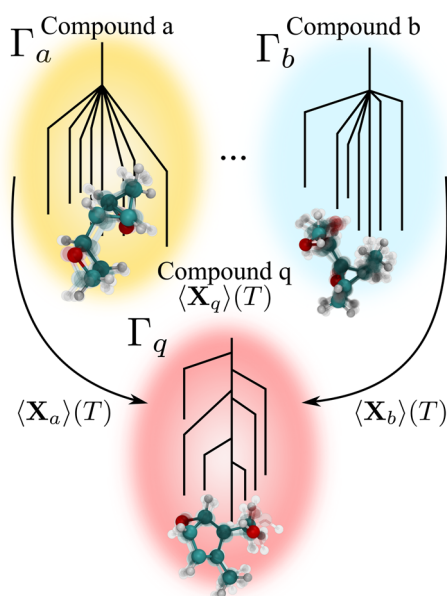
to navigating chemical compound space (CCS). Conversely, more readily available methods based on parameterized force-fields (FFs), implicit solvent models (PCM,<sup>5,6</sup> GBSA,<sup>7</sup> SMD,<sup>8</sup> COSMO,<sup>9,10</sup> and COSMO-RS<sup>11</sup>), or hybrid models (3D-RISM<sup>12–14</sup>) trade reduced computational expense for lower accuracy with respect to experiment. In particular, continuum solvation models exhibit several disadvantages including lack of locality in distinct atomic environments,<sup>15,16</sup> poor modeling of hydrogen-bonding, inaccurate estimates of entropy contributions,<sup>17</sup> as well as poor decoupling

between short-range and long-range effects. Still, in particular, alchemical FF based approaches have become a routine method for free energy calculations.<sup>18</sup>

The recent success of quantum machine learning (QML) in the domain of theoretical and computational chemistry due to unprecedented availability of calculated single-point geometry quantum data has been manifested for challenging molecular problems, such as accurate prediction of molecular electronic properties like atomization energies,<sup>19,20</sup> application to elpasolites,<sup>21</sup> excited states,<sup>23</sup> or fragment based learning with AMONS.<sup>24</sup>

Starting with the work of Behler and Parrinello on high-dimensional neural network potentials,<sup>25,26</sup> there has been an increasing interest in applying ML to molecular dynamics (MD) simulations,<sup>27,28</sup> e.g., using gradient domain ML<sup>29</sup> and molecules in complex environments such as surfaces, water,<sup>30–33</sup> or systems under extreme pressure and temperature.<sup>34</sup> A promising development is coarse-grained ML models.<sup>35–39</sup> Reference 40 provides a general overview about the most recent developments of ML in CCS.

ML based free energy models, on the other hand, are much less established, and potential applications to explore CCS in terms of thermodynamic properties have largely remained unexplored except for some very recent publications.<sup>41–45</sup> Here, we introduce a new ML model capable of predicting ensemble averages, such as free energies of solvation  $\Delta G_{\text{sol}}$ . In particular, our free energy ML (FML) model is designed to deliver both computational efficiency as well as prediction errors, which systematically improve with training set size, thereby being able to reach experimental uncertainty levels. The FML model fills, to the best of our knowledge, an important gap



**FIG. 1.** The FML prediction for a new query compound  $q$  depends on the phase space  $\{\Gamma_i\}$  of all training compounds consisting of conformers, depicted by dis-connectivity graphs, sampled at the same temperature  $T$ , by virtue of ensemble average representations  $\langle \mathbf{X} \rangle(T)$ .

in the field of ML for atomistic simulation by explicitly accounting for an ensemble of molecular conformations through Boltzmann averaged representations,<sup>46–48</sup> rather than through fixed geometry based representations. FML thus avoids the pitfall of neglecting the variance of ensemble properties when basing predictions on fixed geometries only.

As illustrated in Fig. 1, the prediction of an ensemble property for a query compound  $q$  at temperature  $T$  depends on the phase spaces  $\{\Gamma_i\}$  of all training compounds  $i$ . From this point of view, FML infers predictions by combining the information of all the phase spaces at given temperature  $T$  by virtue of the average representation  $\langle \mathbf{X} \rangle$ , i.e., as an integral over the configurationally sampled space.

Our paper is structured as follows: We begin by detailing the FML workflow in Sec. II A with emphasis on the ensemble based representation. Next, we present the results in Sec. IV starting with a numerical demonstration of the necessity of the ensemble representation before assessing the accuracy of FML. Finally, we demonstrate the feasibility of the method for high throughput free energy predictions of 116k organic molecules (a subset of QM9<sup>49</sup>) revealing trends between molecular structure and solubility.

## II. THEORY

We employ a representation based on an ensemble of conformers generated through MD sampling. This gives rise to a unique and temperature dependent representation of the system state. The resulting machine learning framework FML constitutes a physics based approach since this representation is rooted in statistical mechanics.

As shown and discussed below, a comparison with state-of-the-art solvation models reveals that FML retains the promise of being faster, more transferable, and extendable than solvation methods based on conventional fitting of model parameters. While we focus on free energies of solvation, we note that the same methodology might open new pathways for ML applications to other ensemble properties such as protein binding or enthalpy and entropy.

### A. Kernel ridge regression

We use kernel ridge regression<sup>48</sup> (KRR) as a supervised ML method,<sup>50</sup> which can be derived from Gaussian process. KRR non-linearly maps regression and which non-linearly maps input into a high-dimensional feature space, which renders the regression problem linear. The similarity between compounds  $i$  and  $j$  with representations  $\mathbf{X}_i, \mathbf{X}_j$  is measured by applying a Gaussian kernel function  $K$ ,

$$K(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|_2^2}{2\sigma^2}\right), \quad (1)$$

where  $\sigma$  is the kernel-width hyperparameter. The prediction of the property  $p$  of the query compound  $q$  is given by

$$p(\mathbf{X}_q) = \sum_i^{N^{\text{train}}} \alpha_i K(\mathbf{X}_i^{\text{train}}, \mathbf{X}_q), \quad (2)$$

where  $K(\mathbf{X}_i^{\text{train}}, \mathbf{X}_q)$  is the kernel function, evaluated on the query and all training compounds with weight coefficients  $\alpha$ . The unique solution vector for the optimal set of regression coefficients  $\alpha$  is given by

$$\alpha = (\mathbf{K} + \lambda \cdot \mathbb{I})^{-1} \mathbf{p}, \quad (3)$$

with the vector  $\mathbf{p}$  containing all values of the target property in the training set and regularization parameter  $\lambda$ .

## B. Ensemble based representation

While quantum machine learning (QML) is commonly used as a surrogate model for approximate solutions to the electronic Schrödinger equation, i.e., they are associated with exactly one fixed configuration of atoms in a compound (single point). By contrast, the free energy is a property of an ensemble of possible configurational states.

Hence, the intriguing question is how to define a physics based representation of a compound for an ensemble property, such as the free energy of solvation. Here, we have used averages of FCHL19,<sup>46,47</sup> a geometry dependent many-body representation that includes two- and three-body terms where the first term accounts for interatomic distances and the second term accounts for relative orientations of triplets of atoms. In this spirit, representations for thermodynamic properties should be designed similarly, taking into account the ensemble of accessible configurations at a given temperature. We used a thermodynamic ensemble average allowing for a unique definition of a representation, given a set of configurational snapshots obtained by short MD simulations at temperature  $T$  (see Sec. III). First, the FCHL19 representation  $\mathbf{X}(\{\mathbf{r}_i\})$  is computed for all snapshots  $i$  before calculating the ensemble average by numerical integration as follows:

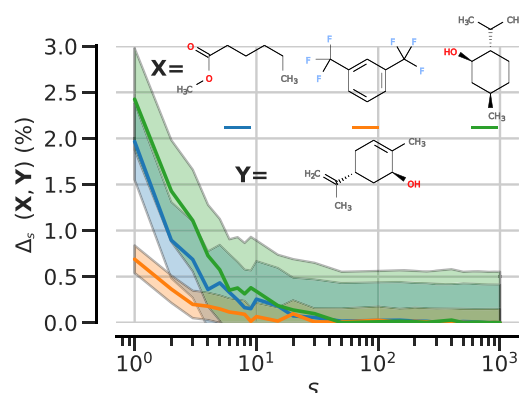
$$\begin{aligned} \langle \mathbf{X} \rangle (T) &= \frac{1}{Z} \int_{\Gamma(T)} \mathbf{X}(\{\mathbf{r}_i\}) e^{-\beta E_i} d\Gamma \\ &\approx \frac{1}{s} \sum_i^s \mathbf{X}_i, \end{aligned} \quad (4)$$

with  $s$  uncorrelated gas-phase solute samples weighted by the respective Boltzmann factor  $e^{-\beta E_i}$  with  $\beta = \frac{1}{k_B T}$  and  $Z$  is the partition function. Note that the average representation  $\langle \mathbf{X} \rangle (T)$  depends on  $T$  since both the integration domain  $\Gamma(T)$  over phase space and the Boltzmann factors depend on  $T$ .

In Fig. 2, we show the convergence of the pairwise distance  $\Delta_s$  between average representations  $\langle \mathbf{X}_s \rangle$  and  $\langle \mathbf{Y}_s \rangle$  for three molecules  $\mathbf{X}$  with respect to a fixed molecule  $\mathbf{Y}$  for a given number of MD samples  $s$ , which we define as

$$\Delta_s(\mathbf{X}, \mathbf{Y}) := \left| \frac{\| \langle \mathbf{X}_s \rangle - \langle \mathbf{Y}_s \rangle \| - \| \langle \mathbf{X}_{s_{\max}} \rangle - \langle \mathbf{Y}_{s_{\max}} \rangle \|}{\| \langle \mathbf{X}_{s_{\max}} \rangle - \langle \mathbf{Y}_{s_{\max}} \rangle \|} \right|, \quad (5)$$

with the maximal number of MD samples  $s_{\max} = 1000$  and the euclidean norm  $\|\cdot\|$ . For the randomly selected examples shown in Fig. 2, we find that for  $s = 10$  uncorrelated molecular geometries, the relative 0.4%. Generally, we find that  $\approx s = 10$



**FIG. 2.** Relative deviation  $\Delta_s(\mathbf{X}, \mathbf{Y})$  with respect to taking all 1000 MD samples in percent [see Eq. (5)] of the distance between the FML representations [see Eq. (4)] of three molecules  $\mathbf{X}$  and fixed  $\mathbf{Y}$  (shown as inset) as a function of the number of MD samples  $s$ .

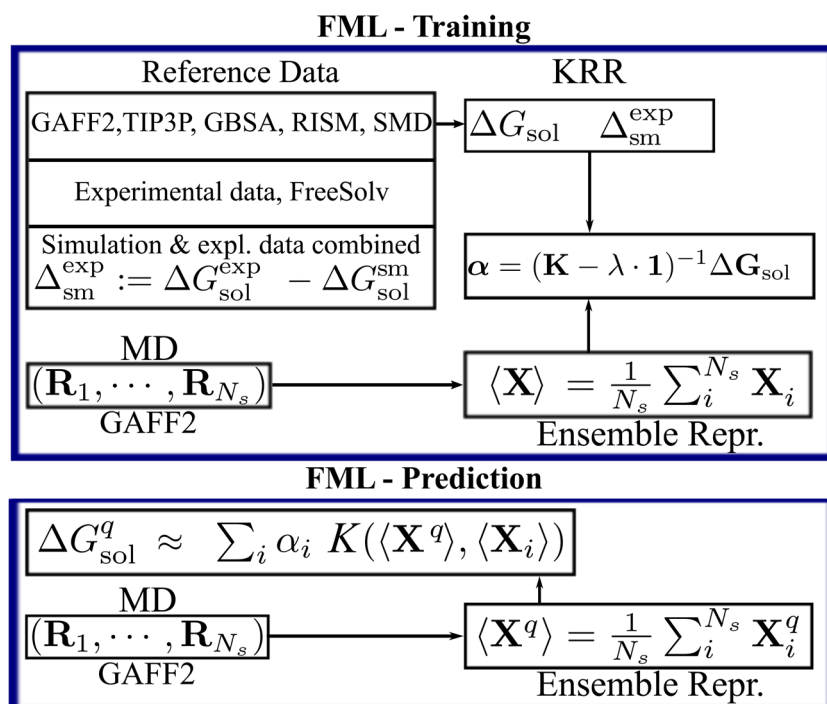
uncorrelated molecular geometries (from MD, see Sec. III) are sufficient to converge the distances below 4% relative deviation (see also Fig. 4 of the [supplementary material](#)).

Our representation, which we also refer to as the FML representation, is similar to what was recently presented in Ref. 51; however, here we rather use experimental free energies for training because of the inherent inaccuracy of conventional of implicit solvation models<sup>17</sup> employed in that work. In addition, there are also a number of QSAR based approaches for representing conformers.<sup>52–54</sup>

In the following, we refer to KRR using a given vacuum geometry as QML and to free energy machine learning using the ensemble averaged FCHL19 representation as FML. Note that unlike most applications of QML, we train on experimental data rather than on computational results from solutions to the Schrödinger Equation. The methodology of FML is fully transferable to free energies from any level of theory, and is both atomistic and *ab initio* in the sense that solely atomic configurations are required as an input without need for molecular connectivities. The generation of the snapshots via molecular dynamics performed using either *ab initio* or force-field based molecular dynamics just as well. In Fig. 3, we show the practical steps for training a FML model for free energies of solvation. First, a set of free energies is calculated, or experimental values are collected from the literature. Subsequently, for all solute configurations, the average over all representations of the trajectories is computed [see Eq. (4)]. Finally, the regression weights  $\alpha$  are determined [see Eq. (3)].

## C. Uniqueness

To illustrate the importance of an ensemble based approach, we use QML to predict a hypothetical value of  $\Delta G_{\text{sol}}$  for all possible fixed conformers of several stereo-isomers. We emphasize that for an ML model to be consistent, the predictions for free energies of solvation of a set of conformers should differ less than the experimental uncertainty. However, in the following, we will show that QML based on a single molecular conformation can indeed lead to inconsistent predictions. For training the QML model, we use



**FIG. 3.** Steps for training an FML model for the free energy of solvation  $\Delta G_{\text{sol}}$  (upper box) and for prediction of  $\Delta G_{\text{sol}}^q$  for a new query compound  $q$  (lower box).

experimental free energies and representation vectors  $\mathbf{X}_{\text{vac}}$  based on vacuum geometries provided by the FreeSolv<sup>55</sup> database.

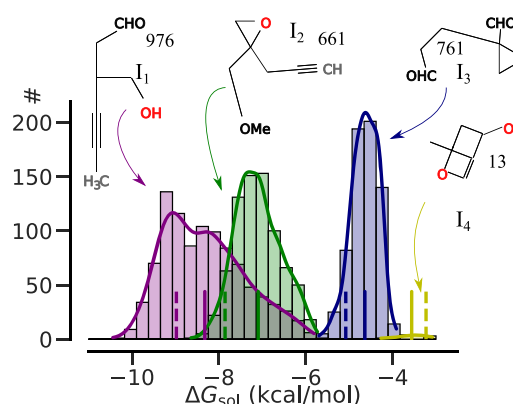
We randomly select four isomers of  $\text{C}_7\text{H}_{10}\text{O}_2$  and use their complete set of conformers resulting from systematic scanning of all dihedral angles of the isomers<sup>56</sup> to compare the FML and QML predictions. The distributions of the corresponding QML predictions of the free energy are shown in Fig. 4. They reveal that for different configurations of the same isomer  $I$ , the predicted free energy  $\Delta G_{\text{sol}}$  can vary by several kcal/mol, well above the experimental uncertainty. Since hundreds of conformational isomers may exist for any given medium sized constitutional isomer (the number of conformational isomers shown in Fig. 4 is  $I_1, \dots, I_4 = 976, 661, 761, 13$ ), it should be obvious that using a single geometry as the representation may lead to considerable prediction errors.

Interestingly, we find (see Sec. IV) that due to an error cancellation QML based on vacuum geometries with weight coefficients,  $\alpha^{\text{vac}}$  can reach MAEs comparable with FML (see Sec. IV A). The reason is that the average of the QML predictions over a set of conformers, denoted by  $\langle \cdot \rangle_c$ , is close to the FML prediction,

$$\underbrace{\left\langle \sum_j^{\text{N}^{\text{train}}} \alpha_j^{\text{QML}} K(\mathbf{X}_j^{\text{QML}}, \mathbf{X}_c) \right\rangle_c}_{\text{Average QML}} \approx \underbrace{\sum_j^{\text{N}^{\text{train}}} \alpha_j^{\text{FML}} K(\langle \mathbf{X}_j \rangle, \langle \mathbf{X}_c \rangle_c)}_{\text{FML}}, \quad (6)$$

using the average representation over all conformers  $\langle \mathbf{X}_c \rangle_c$ . This can be seen by inspecting Fig. 4, where the FML prediction of the average

representation of all conformers is shown as a dashed line and the average of the individual conformer QML free energies is shown as a solid line. The FML weight coefficients  $\alpha^{\text{FML}}$  are obtained using the representation average  $\langle \mathbf{X} \rangle$  [see Eq. (4)] over  $s^{\text{train}} = 300$  MD samples (see Fig. 3).



**FIG. 4.** Distribution of hydration free energies of conformers at 298K estimated by QML for four constitutional isomers ( $\text{C}_7\text{H}_{10}\text{O}_2$ ). Corresponding FML estimates, as well as averages, are denoted by solid and dashed vertical lines, respectively. Insets show molecular graphs of corresponding constitutional isomers and number of conformers.

### D. $\Delta$ -machine learning

Using  $\Delta$ -ML,<sup>57</sup> we promote solvation model (sm) free energies  $\Delta G_{\text{sol}}^{\text{sm}}$  to experimental (exp) uncertainty as follows:

$$\Delta_{\text{sm}}^{\text{exp}}(\langle \mathbf{X} \rangle) \approx \Delta G_{\text{sol}}^{\text{exp}} - \Delta G_{\text{sol}}^{\text{sm}}, \quad (7)$$

using the ensemble representation  $\langle \mathbf{X} \rangle$  as defined in Eq. (4). Note that the FML correction is different for each compound accounting for underestimations as well as overestimations of baseline free energy models. The step for promoting free energies to a predicted experimental value  $\Delta G_{\text{sol}}^{\text{exp}}$  (FML) is given by

$$\Delta G_{\text{sol}}^{\text{exp}}(\langle \mathbf{X} \rangle) \approx \Delta G_{\text{sol}}^{\text{sm}} + \Delta_{\text{sm}}^{\text{exp}}(\langle \mathbf{X} \rangle). \quad (8)$$

Besides these modifications, the workflow remains the same as in Fig. 3.

## III. COMPUTATIONAL DETAILS

### A. Molecular dynamics

All MD simulations are performed with OpenMM<sup>58</sup> in vacuum in the NVT ensemble using a Langevin integrator with a friction coefficient of  $1 \text{ ps}^{-1}$  and the small molecule FF GAFF2<sup>59,60</sup> with a time-step of  $\Delta t = 2 \text{ fs}$  and total simulation time of 2 ns using SHAKE.<sup>51</sup> Partial charges are computed with antechamber<sup>59,60</sup> at the AM1-BCC<sup>62</sup> level. An exception is the implicit solvent GBSA<sup>63,64</sup> simulations where we use AMBER<sup>59</sup>/GAFF,<sup>60</sup> while all other simulation parameters are the same. MD samples were selected with 2 ps separation, which is well beyond the maximal correlation time of  $\sim 0.5 \text{ ps}$ .

### B. Machine learning

To optimize the hyper-parameters  $\sigma$  and  $\lambda$  [see Eqs. (1) and (3)], we perform ten-fold cross-validation on the training set only and validate the performance on the test set using the QML package.<sup>65</sup> The training set contains all FreeSolv<sup>55</sup> molecules except for those 146 test set molecules, which are also part of the QM9 database.<sup>49</sup>

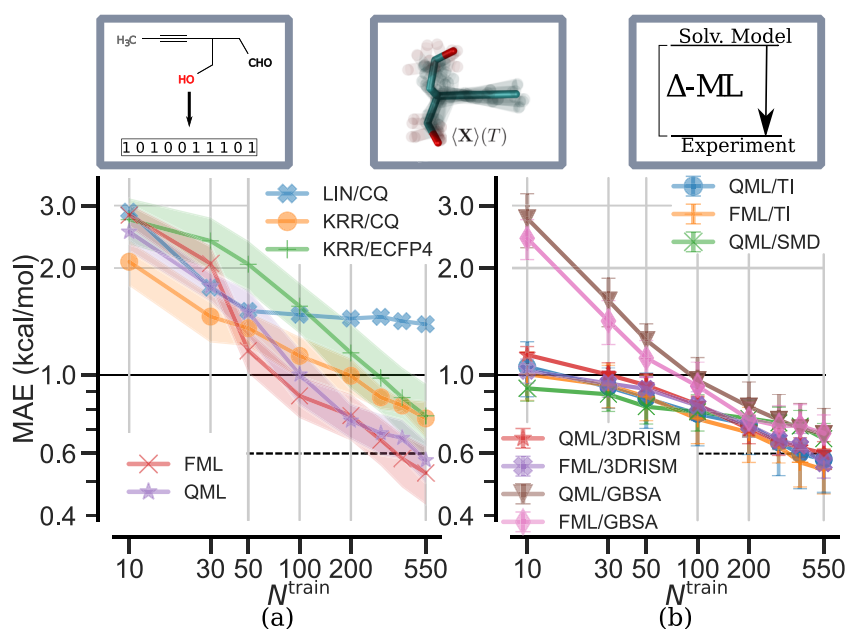
Histograms showing the free energy distribution and the size of all molecules in the FreeSolv database are shown in the [supplementary material](#). For numerous FreeSolv compounds, no value for the experimental uncertainty is reported, but a default value of  $0.6 \text{ kcal/mol}$  corresponding to thermal energy fluctuations at  $298.15 \text{ K}$  is assumed, which we use as our target accuracy, albeit the total average absolute error of all experimental values is slightly smaller ( $0.57 \text{ kcal/mol}$ ).

## IV. RESULTS

### A. Learning curves and free energies of 116k molecules

The performance of several ML models is assessed using learning curves, i.e., the prediction error is reported as a function of training set size  $N$  (see Fig. 5). Most notably, we assess FML, the ensemble average representation [see Eq. (4)], and  $\Delta$ -ML [see Eq. (8)] for predicting experimental free energies of the FreeSolv<sup>55</sup> database.

In addition, we also test the RDKit<sup>67</sup> implementation of the extended connectivity fingerprint<sup>66</sup> (ECFP4) commonly used in cheminformatics as well as a custom QSAR based representation, named CQ (see the [supplementary material](#)). To compare KRR with conventional fitting methods, we also report the MAEs of a multilinear regression model with CQ. The error bars and colored areas



**FIG. 5.** Learning curves for free energies of the FreeSolv database<sup>55</sup> including standard deviation at each number of training molecules  $N$  obtained by ten-fold cross-validation for feature based KRR, QML, and FML in (a). QML/FML based  $\Delta$ -ML with various solvation models as baseline in (b). Solid black line marks thermo-chemical accuracy at  $1 \text{ kcal/mol}$ , and dashed line marks the experimental uncertainty ( $298 \text{ K} \cdot k_B \approx 0.6 \text{ kcal/mol}$ ). The boxes above illustrate the different ML approaches; in the first approach, the molecule is featurized (as in ECFP4<sup>66</sup>), but all 3d information is lost. In the second case (Boltzmann weighted), 3d structures of the molecule are included (FCHL19<sup>46,47</sup>). Third, physics based approaches can be used as a baseline for  $\Delta$ -ML<sup>57</sup> to improve ML predictions, as in (b).

in Fig. 5 show the standard deviation of the MAEs resulting from ten-fold cross-validation over the FreeSolv database. We find that feature based models ECFP4 and CQ [see Fig. 5(a)] perform worst, reaching an MAE of only 0.8 kcal/mol for the maximal training set size of  $N^{\text{train}} = 550$ . This might be due to the fact that feature based representations do not properly weigh the ensemble of molecular conformations, which can have a large influence on predictions of  $\Delta G_{\text{sol}}$  (see Sec. II A).

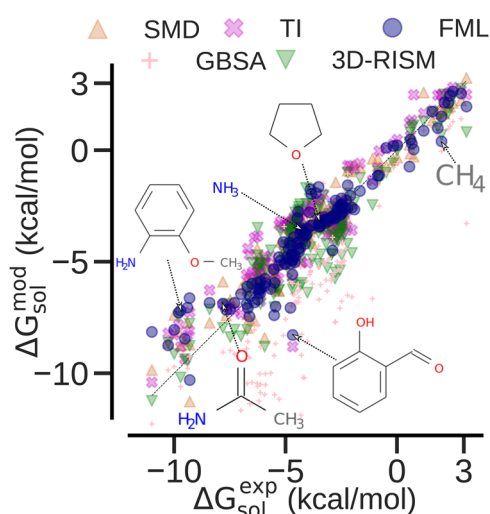
QML, on the other hand, reaches the target accuracy, the experimentally relevant accuracy of 0.6 kcal/mol for about  $N^{\text{train}} = 550$  training molecules. The best FML model trained with ten random MD samples per molecule with  $T = 350$  K essentially results in the same MAE. While both models reach the experimentally relevant accuracy, we find that FML has a slightly smaller offset hitting 1 kcal/mol for  $N^{\text{train}} \approx 50$ , while the QML model trained on the vacuum geometries has an MAE of about 1.5 kcal/mol. The otherwise similar performance of both models may to some extent be attributed to error compensation, as discussed above.

Learning curves for  $\Delta$ -ML<sup>57</sup> are also shown in Fig. 5 using various solvation models as a baseline such as GBSA (GAFF<sup>59,60</sup>), 3D-RISM<sup>12,13</sup> (GAFF<sup>59,60</sup>/TIP3P<sup>68</sup>), SMD<sup>8</sup> (M06-2X<sup>69</sup>/Def2-TZVP<sup>70</sup>), and TI (GAFF<sup>59,60</sup>/TIP3P<sup>68</sup>). The GBSA values have been computed using AMBER<sup>59,60</sup> (see Sec. III), the TI values are from the FreeSolv<sup>55</sup> database, and the values for the two implicit solvation methods 3D-RISM<sup>12,13</sup> and SMD<sup>8</sup> are from the supplementary material of Ref. 14. We find that  $\Delta$ -ML can lower the offset of the learning curves and thus may be useful if experimental data are scarce. FML requires about 100 training molecules to reach 0.8 kcal/mol, whereas  $\Delta$ -ML using TI or 3D-RISM only needs around 50. Note that the SMD baseline model has the lowest offset since SMD has the smallest MAE of 0.93 kcal/mol vs an MAE of 1.12 kcal/mol for TI on the FreeSolv database. However, after inclusion of  $\sim 150$  compounds in training both approaches, direct FML and  $\Delta$ -ML perform similarly and converge to target accuracy.

Similar observations were also reported in Ref. 42, but we find that the target accuracy can be reached using FML without additional solvation calculations for  $\Delta$ -ML. We note that some caution is required for this comparison since the accuracy of  $\Delta$ -ML relies heavily on systematic correlation between the baseline and the target.

## B. Comparison to other models

A comparison of various solvation models with FML is shown in terms of a scatter plot (see Fig. 6) displaying predictions vs experiment. The training set consists of all molecules in FreeSolv but not in QM9, and the overlap (146 molecules) is used as a test set. The MAEs for FML and the tested solvation models are listed in Table I. We find that the best FML model with  $s^{\text{train}} = 10$  and  $T^{\text{train}} = 350$  K results in an MAE of 0.57 kcal/mol, reaching the experimentally relevant accuracy of solvation energy measurements and performing slightly better than SMD with 0.61 kcal/mol. Note that FML has some significant outliers such as methane with an error of about 2 kcal/mol. The largest outlier, however, is 2-hydroxybenzaldehyde with an error of 3.6 kcal/mol. It is worth emphasizing that SMD uses at least 79 of the test set molecules for parameterization, which were not included in training of the FML model. Still, FML has a slightly smaller MAE despite using fewer training points compared to SMD. Furthermore,



**FIG. 6.** Scatter plot comparing free energies of solvation predicted by FML (350 K; 10) (blue) using an ensemble based representation and solvation models: SMD (chocolate), TI (violet), 3D-RISM (green), and GBSA (pink). The FML predictions are also listed in the [supplementary material](#).

we find that FML clearly outperforms TI and 3D-RISM, reaching an MAE of 1 kcal/mol after about  $N^{\text{train}} = 100$  training molecules. We emphasize that a typical SMD (M06-2X<sup>69</sup>/Def2-TZVP<sup>70</sup>) calculation of a molecule in the FreeSolv database with Gaussian09<sup>71</sup> is computationally demanding, while a force-field based MD simulation in vacuum to generate the averaged representation can be run on a single modern central processing unit (CPU) in less than 10 min (see Table I for timing benchmarks).

In addition, FML clearly outperforms COSMO<sup>9,10,72</sup> with B-P86<sup>73,74</sup>/TZVP,<sup>75</sup> resulting in an MAE of 1.94 kcal/mol with higher computational costs. For comparison with COSMO-RS as implemented in COSMOtherm,<sup>76</sup> we refer to a benchmark test<sup>11,77</sup> with a set of 274 molecules resulting in an MAE of 0.52 kcal/mol reaching the accuracy of FML but at higher computational cost (see Table I) and as for all other density functional theory (DFT) methods with a worse scaling with the number of atoms compared to FFs. The predictions of FML for the test set are provided as a table in the [supplementary material](#).

To provide further insight into the prediction errors, 642 FML models were trained on all molecules except for a single test molecule following a “one-versus-all” strategy.

This results in a mean unsigned error of 0.51 kcal/mol, which is slightly smaller than the average reported experimental uncertainty of 0.57 kcal/mol. Note that for 450 of 642 FreeSolv<sup>55</sup> molecules, a default error of  $k_B \cdot T \approx 0.6$  kcal/mol was reported, and our observation indicates that the “true” experimental error may actually be much smaller. Indeed, we find that the average of the reported experimental uncertainties (other than default) is 0.48 kcal/mol. Given the slope and the offset of a continued learning curve (see Fig. 5 in the [supplementary material](#)) as well as 0.48 kcal/mol as an upper boundary of the noise level, we can extrapolate that the learning curve would begin to flatten off for more than  $N^{\text{train}} = 670$  training molecules. We

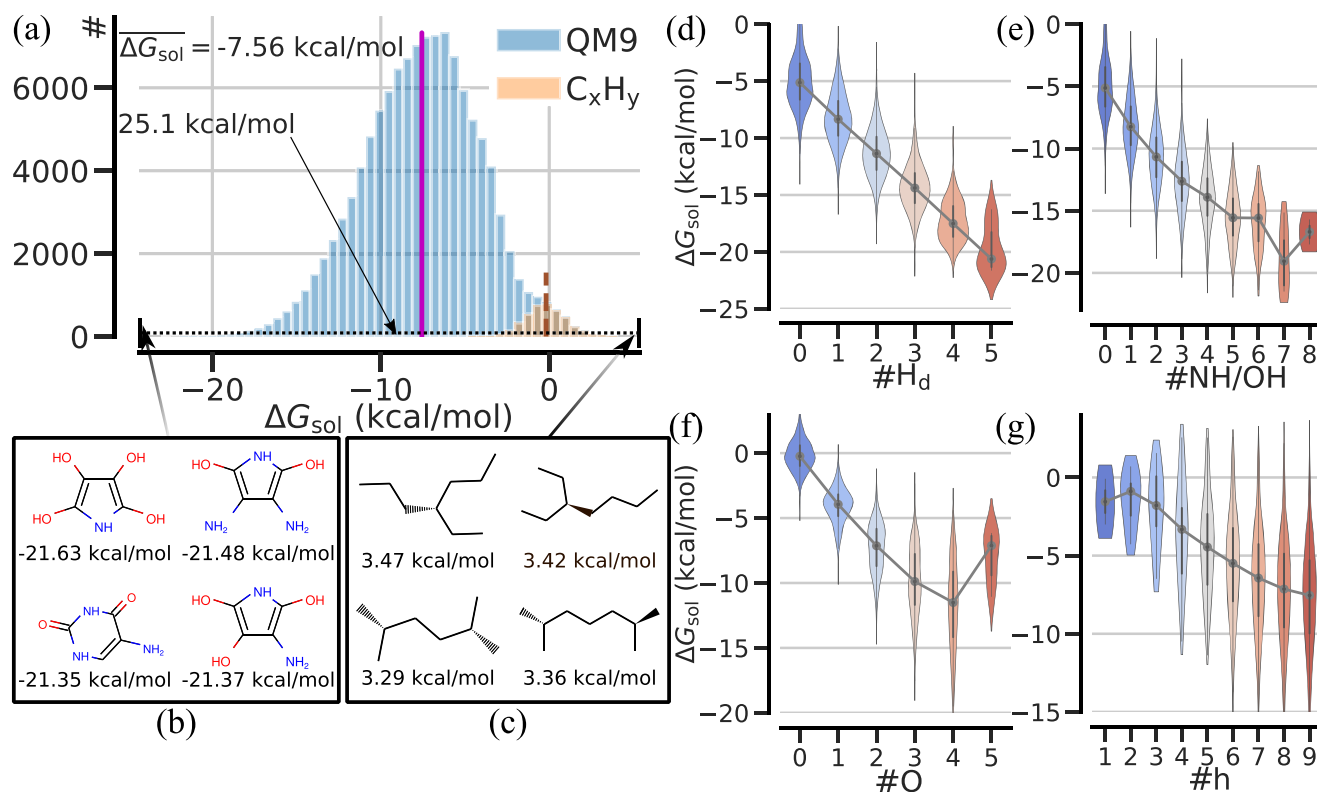
**TABLE I.** Comparison of MAEs, Pearson's  $r$ , and estimated order of CPU time per solute prediction for various solvation models and FML. Approximate number of training molecules  $N^{\text{train}}$  needed for FML to reach the MAE of each respective method. The conversion factor from GPU to CPU  $c = \frac{t_{\text{CPU}}}{t_{\text{GPU}}}$  may vary substantially between  $\sim 10$  and 60 depending on hardware/MD code (here OpenMM<sup>58</sup>).  $g$  is the number of grid points for TI, typically  $\sim 10$ .

Class	Model	MAE (kcal/mol)	$r$	$N^{\text{train}}$	$\sim$ CPU h/solute	References	Year of publication
DFT	SMD	0.61 <sup>14</sup>	0.96	400	High	8 and 71	2009
	COSMO	1.94	0.90	20	$10^{-1}$	9, 10, and 72	1993
	COSMO-RS	0.52 <sup>78</sup>	0.91	...	$10^{-1}$	11, 72, 76, and 78	2000
	DCOSMO-RS	0.94 <sup>78</sup>	0.87	100	$10^{-1}$	79	2006
FF	3D-RISM	0.99 <sup>14</sup>	0.90	100	$10^{-1}$	12 and 13	1998
	TI	0.93 <sup>55</sup>	0.94	100	$g \times c \times 10^{-1}$	55	2017
	GBSA	2.41	0.84	20	$10^{-2}$	63 and 64	2004
ML	FML	0.57	0.95	490	$10^{-2}$	This work	2020

find that the two molecules with identifier mobley\_2523689 and mobley\_3201701 with the largest prediction errors of 5.3 kcal/mol and mobley\_3201701 4.4 kcal/mol also have the largest reported experimental uncertainty of 1.93 kcal/mol. This may indicate that the

reported experimental uncertainty may cause the large deviation for these outliers (see Fig. 6 in the [supplementary material](#)).

Such a one-versus-all scheme may be useful to identify possible candidates with high experimental error, but we note that a large



**FIG. 7.** Free energy distribution of 116k molecules predicted by FML in (a); small histogram corresponds to a subset of 4907 hydrocarbons with averages shown as solid and dashed lines, respectively. Four molecules with most negative (b) and positive (c)  $\Delta G_{\text{sol}}$  in clockwise order. The mean of  $\Delta G_{\text{sol}}$  for simple descriptors such as number of H-bond donors  $H_d$  (d), NH/OH groups (e), oxygen atoms O in stoichiometry  $C_{n_c}H_{n_h}O_{n_o}$  (f), and heavy atoms h (g).



prediction error is not a sufficient condition but rather an indication of inconsistency with the rest of the dataset. A large prediction error may also mean that the molecule in question has a very different chemical structure such that the structure–property relationship cannot be sufficiently explained by the training set.

### C. Predicted solvation for 116k organic molecules

We use FML to calculate free energies of a large dataset of organic molecules, a subset of CCS, in order to identify trends between solubility and structure. More specifically, 116k molecules of the QM9 dataset have been considered in order to predict the free energy distribution in Fig. 7(a). For comparison, we also show the corresponding distribution of  $\Delta G_{\text{sol}}$  for the FreeSolv database in the supplementary material. The FML model was trained on the complete FreeSolv database with coincident QM9 molecules removed from training. We find that the free energy distribution is approximately Gaussian with a mean value of  $\Delta G_{\text{sol}} = -7.56$  kcal/mol spanning a range of 25.1 kcal/mol. The molecules with the most negative and positive  $\Delta G_{\text{sol}}$  are QM9 compounds indexed 26 712 and 118 570 with  $-21.63$  kcal/mol and  $3.47$  kcal/mol, respectively [see Figs. 7(b) and 7(c)]. The top 50 most soluble and least soluble molecules, according to FML estimates, are also shown in the supplementary material. We find that the most soluble molecules have a planar ring structure. On the other hand, the 4907 hydrocarbons of QM9 occupy the right tail of the distribution. Molecules with many hetero atoms bonded in NH/OH groups in a planar ring structure tend to have a very negative  $\Delta G_{\text{sol}}$ , while aliphatic linear molecules tend to be less soluble. This can be explained by the large difference in electronegativity between the H and the N, O atoms leading to polar bonds eventually resulting in H-bonds, lowering the enthalpy of solvation. Alkanes, on the other hand, have a negligible polarity and thus only interact with water through much weaker van der Waals interaction resulting in low solubility.

These simple rules for the solubility lead to trends for the average free energy  $\Delta G_{\text{sol}}$  such as approximately linear relations with the number of H-bond donors  $H_d$  or the number of OH and NH groups NH/OH, as shown in Figs. 7(d) and 7(e). Furthermore,  $\Delta G_{\text{sol}}$  decreases with the number of oxygen atoms for molecules with stoichiometry  $C_{n_c}H_{n_h}O_{n_o}$  and with the number of heavy atoms  $h$  [see Figs. 7(f) and 7(g)]. The former effect is due to enthalpic H-bond contributions from hydroxyl groups, and the latter is due to weak van der Waals interactions, which roughly scale with the molecule size. Such linear free-energy relationships (LFER) are well known and used for predictive models, e.g., for solvation<sup>80,81</sup> or organic chemistry.<sup>82</sup>

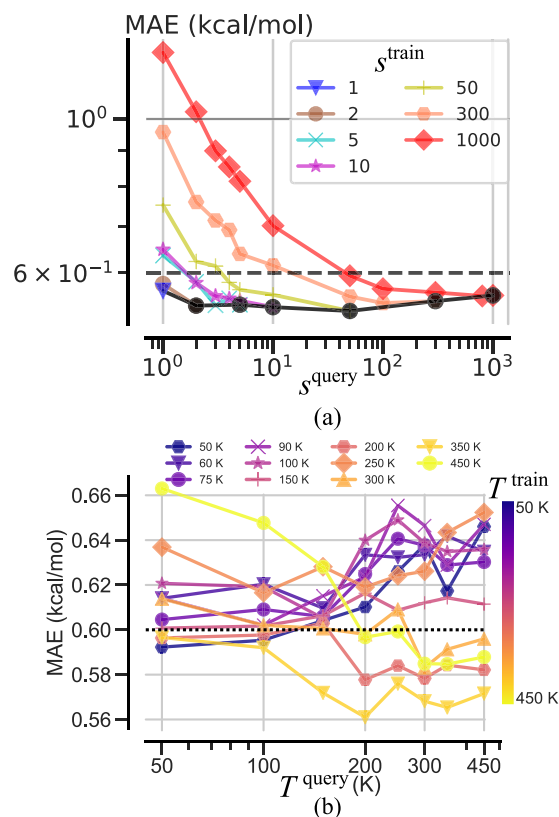
The 116k predictions of FML for the subset of QM9<sup>49</sup> are available in the supplementary material.

### D. Analysis of the model

We study convergence of the MAE as a function of the number of MD samples for the training  $s^{\text{train}}$  and test molecules  $s^{\text{query}}$ , given an MD sampling temperature of  $T^{\text{query}} = T^{\text{train}} = 350$  K. Note that both  $s^{\text{train}}$  and  $s^{\text{query}}$  refer to the representation [sum in Eq. (4)] and not to the number of training or test molecules (here always 620 and 22 respectively, corresponding to 29 fold cross-validation over FreeSolv<sup>35</sup>).

By inspection of Fig. 8(a), we find that FML models trained with large  $s^{\text{train}}$  tend to have a larger offset in the MAEs and a larger number of test samples  $s^{\text{query}}$  is needed to reach the same accuracy as a model with small  $s^{\text{train}}$ . Note that we only show up to  $s^{\text{query}} \leq s^{\text{train}}$  for the query molecule representation on the x axis since a model evaluated on query FML representations with more MD samples than used for training would certainly not be desirable for ML. All ML models improve with  $s^{\text{query}}$  until the MAEs saturate at  $s^{\text{train}} = s^{\text{query}}$  where FML generally achieves the highest accuracy. Interestingly, we find that the accuracy of the FML models increases only up to  $s^{\text{train}} \approx 10$  and saturates beyond. This is because we find the FML representation to be sufficiently converged for approximately 10 random MD samples per molecule (see Fig. 2).

Next, we investigate the temperature dependence of the FML representation. To this end, the MAEs of several ML models  $\text{FML}(T^{\text{train}}; s^{\text{train}})$  with training molecule sampling temperatures  $T^{\text{train}}$  between 50 K and 450 K are compared for various query molecule sampling temperatures  $T^{\text{query}}$  with  $s^{\text{train}} = s^{\text{query}} = 10$  [see Fig. 8(b)]. We find that the FML models usually perform best if



**FIG. 8.** Comparison of MAEs of FML models using different number of samples in the training and query molecule representation ( $T^{\text{query}} = T^{\text{train}} = 350$  K) with  $N^{\text{train}} = 620$  distinct training molecules (a). Various temperatures for sampling the test  $T^{\text{query}}$  and training molecules  $T^{\text{train}}$  in (b) and resulting MAEs. Solid line marks thermo-chemical accuracy at 1 kcal/mol, and dashed line marks the experimental uncertainty ( $298 \text{ K} \cdot k_B \approx 0.6 \text{ kcal/mol}$ ) with  $N^{\text{train}} = 496$  training molecules evaluated on the test set with 146 molecules.

$T^{\text{train}} \approx T^{\text{query}}$ , e.g., model FML (50 K) performs best for  $T^{\text{train}} = T^{\text{query}} = 50$  K and the MAE increases slightly for higher temperatures  $T^{\text{query}}$ . On the other hand, FML (450 K) performs worse at lower than at higher temperature. This indicates, as suggested by Eq. (4), that the average representation indeed results in models specific for the temperature and phase space  $\Gamma(T)$  spanned by the training MD samples.

## V. CONCLUSION

We have studied the role of the representation for ML models of ensemble properties. As one would expect, numerical results confirm that the representation should be rooted in statistical mechanics since ML based on a single molecular geometry can lead to large and spurious prediction errors. The definition of a representation based on Boltzmann weighted averages does not only resolve this issue but also naturally introduces temperature dependence.

In a similar manner, FML models could be constructed that depend on pressure, or chemical potential, to account for increasingly more realistic canonical and grand-canonical ensembles. The numerical performance of FML is encouraging: FML reaches experimental uncertainty for relatively small training sets and at low computational cost for new query estimates. As such, it is better or on par with state-of-the-art models in the field, and it emerges as a viable alternative whenever sufficient training data are available. Furthermore, we find that  $\Delta$ -ML can improve the predictions for small training set sizes; however, FML can reach experimental relevant accuracy without requiring additional solvation model calculations. Furthermore, we stress that it is straightforward to improve the accuracy and transferability of FML by adding more experimental data points. This is not necessarily true for other more conventional solvation models. To demonstrate the usefulness of a transferable FML model, we have predicted solvation energies for 116k organic molecules of the QM9 database (see Fig. 7). The results confirm known trends, namely, that molecules with a high solubility tend to have many hetero atoms and are arranged in planar ring structures, while linear aliphatic molecules tend to have a lower solubility. We have therefore demonstrated that FML can be used to study solvation throughout CCS and that it might prove useful to identify additional non-trivial structure–property relationships.

## SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for a table of predicted free energies of solvation on the test set and information about the CQ-representation. We also show the predicted most and least soluble molecules of QM9 as well as the distribution of heavy atoms in the FreeSolv<sup>55</sup> database. In addition, we show a histogram illustrating the convergence of the FML representation and the individual prediction errors and outliers of the model as well as an extended learning curve.

## ACKNOWLEDGMENTS

O.A.v.L. acknowledges support from the Swiss National Science Foundation (Grant No. 407540\_167186 NFP 75 Big Data) and the European Research Council (Grant No. ERC-CoG grant QML and

H2020 projects BIG-MAP and TREX). This project received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement Nos. 952165 and 957189. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains. This work was partly supported by the NCCR MARVEL, funded by the Swiss National Science Foundation. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 772834).

## DATA AVAILABILITY

The 116k predictions of FML for the subset of QM9<sup>49</sup> are available as an ASCII text file (QM9\_116k.txt) for download in the [supplementary material](#). The predictions of FML for the validation set are provided as a table in a PDF file in the [supplementary material](#) (SI.pdf). The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

- R. Jinnouchi, F. Karsai, and G. Kresse, *Phys. Rev. B* **101**, 060201 (2020).
- Y. Basdogan, M. C. Groenenboom, E. Henderson, S. De, S. B. Rempe, and J. A. Keith, *J. Chem. Theory Comput.* **16**, 633 (2020).
- L. Mones, A. Jones, A. W. Götz, T. Laino, R. C. Walker, B. Leimkuhler, G. Csányi, and N. Bernstein, *J. Comput. Chem.* **36**, 633 (2015).
- H. Takahashi, H. Kambe, and A. Morita, *J. Chem. Phys.* **150**, 114109 (2019).
- F. Fogolari, A. Brigo, and H. Molinari, *J. Mol. Recognit.* **15**, 377 (2002).
- B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel, and P. J. Stephens, *J. Phys. Chem. A* **106**, 6102 (2002).
- A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *J. Chem. Theory Comput.* **9**, 609 (2013).
- A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **113**, 6378 (2009).
- A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 799 (1993).
- A. Klamt, *J. Phys. Chem.* **99**, 2224 (1995).
- A. Klamt and F. Eckert, *Fluid Phase Equilib.* **172**, 43 (2000).
- D. Beglov and B. Roux, *J. Phys. Chem. B* **101**, 7821 (1997).
- A. Kovalenko and F. Hirata, *Chem. Phys. Lett.* **290**, 237 (1998).
- D. Roy and A. Kovalenko, *J. Phys. Chem. A* **123**, 4087 (2019).
- Y. Basdogan, A. M. Maldonado, and J. A. Keith, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **10**, e1446 (2020).
- J. Guerard and J. S. Arey, *J. Chem. Theory Comput.* **9**, 5046 (2013).
- J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. van der Spoel, *J. Chem. Theory Comput.* **13**, 1034 (2017).
- A. S. J. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, A. Rizzi, J. Scheen, M. R. Shirts, G. Tresadern, and H. Xu, *Living J. Comput. Mol. Sci.* **2**, 18378 (2020).
- M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Phys. Rev. Lett.* **108**, 058301 (2012).
- F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, *J. Chem. Theory Comput.* **13**, 5255 (2017).
- F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, *Phys. Rev. Lett.* **117**, 135502 (2016).
- M. Schwilk, D. N. Tahchieva, and O. A. von Lilienfeld, "Large yet bounded: Spin gap ranges in carbenes," [arXiv:2004.10600 \[physics.chem-ph\]](#) (2020).
- J. Westermayr and P. Marquetand, "Machine learning for electronically excited states of molecules," *Chem. Rev.* (published online 2020).

- <sup>24</sup>B. Huang and A. O. von Lilienfeld, *Nat. Chem.* **12**, 945 (2020).
- <sup>25</sup>J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- <sup>26</sup>J. Behler, *J. Chem. Phys.* **145**, 170901 (2016).
- <sup>27</sup>O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, “Machine learning force fields,” [arXiv:2010.07067](https://arxiv.org/abs/2010.07067) [physics.chem-ph] (2020).
- <sup>28</sup>S. Manzhos and T. Carrington, *Chem. Rev.* (published online 2020).
- <sup>29</sup>H. E. Sauceda, M. Gastegger, S. Chmiela, K.-R. Müller, and A. Tkatchenko, “Molecular force fields with gradient-domain machine learning (GDML): Comparison and synergies with classical force fields,” [arXiv:2008.04198](https://arxiv.org/abs/2008.04198) [physics.chem-ph] (2020).
- <sup>30</sup>M. L. Paleico and J. Behler, *J. Chem. Phys.* **153**, 054704 (2020).
- <sup>31</sup>V. Quaranta, M. Hellström, and J. Behler, *J. Phys. Chem. Lett.* **8**, 1476 (2017).
- <sup>32</sup>M. Hellström and J. Behler, *J. Phys. Chem. B* **121**, 4184 (2017).
- <sup>33</sup>B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1110 (2019).
- <sup>34</sup>B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, *Nature* **585**, 217 (2020).
- <sup>35</sup>J. Wang, S. Chmiela, K.-R. Müller, F. Noé, and C. Clementi, *J. Chem. Phys.* **152**, 194106 (2020).
- <sup>36</sup>F. Noé, G. De Fabritiis, and C. Clementi, *Curr. Opin. Struct. Biol.* **60**, 77 (2020).
- <sup>37</sup>C. Scherer, R. Scheid, D. Andrienko, and T. Berau, *J. Chem. Theory Comput.* **16**, 3194 (2020).
- <sup>38</sup>K. H. Kanekal and T. Berau, *J. Chem. Phys.* **151**, 164106 (2019).
- <sup>39</sup>C. Hoffmann, R. Menichetti, K. H. Kanekal, and T. Berau, *Phys. Rev. E* **100**, 033302 (2019).
- <sup>40</sup>B. Huang and O. A. von Lilienfeld, “*Ab initio* machine learning in chemical compound space,” [arXiv:2012.07502](https://arxiv.org/abs/2012.07502) [physics.chem-ph] (2020).
- <sup>41</sup>J. Gebhardt, M. Kiesel, S. Riniker, and N. Hansen, *J. Chem. Inf. Model.* **60**, 5319 (2020).
- <sup>42</sup>J. Scheen, W. Wu, A. S. J. S. Mey, P. Tosco, M. Mackey, and J. Michel, *J. Chem. Inf. Model.* **60**, 5331 (2020).
- <sup>43</sup>H. Lim and Y. Jung, “MLSolv-A: A novel machine learning-based prediction of solvation free energies from pairwise atomistic interactions,” [arXiv:2005.06182](https://arxiv.org/abs/2005.06182) [stat.ML] (2020).
- <sup>44</sup>S. Axelrod and R. Gomez-Bombarelli, “Molecular machine learning with conformer ensembles,” [arXiv:2012.08452](https://arxiv.org/abs/2012.08452) [cs.LG] (2020).
- <sup>45</sup>F. H. Vermeire and W. H. Green, “Transfer learning for solvation free energies: From quantum chemistry to experiments,” [arXiv:2012.11730](https://arxiv.org/abs/2012.11730) [physics.chem-ph] (2020).
- <sup>46</sup>A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. A. von Lilienfeld, *J. Chem. Phys.* **152**, 044107 (2020).
- <sup>47</sup>F. A. Faber, A. S. Christensen, B. Huang, and O. Anatole von Lilienfeld, *J. Chem. Phys.* **148**, 241717 (2018).
- <sup>48</sup>B. Huang, N. O. Symonds, and O. A. von Lilienfeld, [arXiv:1807.04259](https://arxiv.org/abs/1807.04259) (2018).
- <sup>49</sup>R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Sci. Data* **1**, 140022 (2014).
- <sup>50</sup>V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, 1998).
- <sup>51</sup>C. Rauer and T. Berau, *J. Chem. Phys.* **153**, 014101 (2020).
- <sup>52</sup>A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, and C. Duraiswami, *J. Am. Chem. Soc.* **119**, 10509 (1997).
- <sup>53</sup>V. E. Kuz'min, A. G. Artemenko, P. G. Polishchuk, E. N. Muratov, A. I. Hromov, A. V. Liahovskiy, S. A. Andronati, and S. Yu. Makan, *J. Mol. Model.* **11**, 457 (2005).
- <sup>54</sup>D. V. Zankov, M. Matveieva, A. Nikonenko, R. Nugmanov, A. Varnek, P. Polishchuk, and T. Madzhidov, “QSAR modeling based on conformation ensembles using a multi-instance learning approach,” [chemrxiv:13456277](https://arxiv.org/abs/13456277) (2020).
- <sup>55</sup>G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley, *J. Chem. Eng. Data* **62**, 1559 (2017).
- <sup>56</sup>D. Lemm, J. C. Kromann, G. F. von Rudorff, and O. A. von Lilienfeld, “Clockwork,” (unpublished).
- <sup>57</sup>R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- <sup>58</sup>P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, *PLoS Comput. Biol.* **13**, e1005659 (2017).
- <sup>59</sup>J. Wang, W. Wang, P. A. Kollman, and D. A. Case, *J. Mol. Graphics Modell.* **25**, 247 (2006).
- <sup>60</sup>J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- <sup>61</sup>J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- <sup>62</sup>A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, *J. Comput. Chem.* **21**, 132 (2000).
- <sup>63</sup>A. Onufriev, D. Bashford, and D. A. Case, *Proteins* **55**, 383 (2004).
- <sup>64</sup>J. r. Weiser, P. S. Shenkin, and W. C. Still, *J. Comput. Chem.* **20**, 217 (1999).
- <sup>65</sup>A. Christensen, F. Faber, B. Huang, L. Bratholm, A. Tkatchenko, K. Müller, and O. v. Lilienfeld (unpublished) (2017), <https://github.com/qmlcode/qml>.
- <sup>66</sup>D. Rogers and M. Hahn, *J. Chem. Inf. Modell.* **50**, 742 (2010).
- <sup>67</sup>RdKit: Open-source cheminformatics, Python package.
- <sup>68</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>69</sup>Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.* **119**, 525 (2008).
- <sup>70</sup>F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- <sup>71</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian16 Revision C.01 (Gaussian Inc., Wallingford, CT, 2016).
- <sup>72</sup>TURBOMOLE V7.2 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- <sup>73</sup>A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- <sup>74</sup>R. Ahlrichs, F. Furche, and S. Grimme, *Chem. Phys. Lett.* **325**, 317 (2000).
- <sup>75</sup>A. Schäfer, C. Huber, and R. Ahlrichs, *J. Chem. Phys.* **100**, 5829 (1994).
- <sup>76</sup>F. Eckert and A. Klamt, CosmoTherm, Cosmologic GMBH & CoKG, Leverkusen, Germany, version c2.1, revision 01.07, 2007.
- <sup>77</sup>A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco, and F. J. Luque, *Acc. Chem. Res.* **42**, 489 (2009).
- <sup>78</sup>A. Klamt and M. Diedenhofen, *J. Phys. Chem. A* **119**, 5439 (2015).
- <sup>79</sup>S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, and F. Neese, *J. Phys. Chem. A* **110**, 2235 (2006).
- <sup>80</sup>R. W. Taft, J.-L. M. Abboud, M. J. Kamlet, and M. H. Abraham, *J. Solution Chem.* **14**, 153 (1985).
- <sup>81</sup>T. N. Borhani, S. García-Muñoz, C. Vanesa Luciani, A. Galindo, and C. S. Adjiman, *Phys. Chem. Chem. Phys.* **21**, 13706 (2019).
- <sup>82</sup>M. Bragato, G. F. von Rudorff, and O. A. von Lilienfeld, *Chem. Sci.* **11**, 11859 (2020).