

# Operator Quantum Machine Learning: Navigating the Chemical Space of Response Properties

Anders S. Christensen and O. Anatole von Lilienfeld\*

**Abstract:** The identification and use of structure–property relationships lies at the heart of the chemical sciences. Quantum mechanics forms the basis for the unbiased virtual exploration of chemical compound space (CCS), imposing substantial compute needs if chemical accuracy is to be reached. In order to accelerate predictions of quantum properties without compromising accuracy, our lab has been developing quantum machine learning (QML) based models which can be applied throughout CCS. Here, we briefly explain, review, and discuss the recently introduced operator formalism which substantially improves the data efficiency for QML models of common response properties.

**Keywords:** Chemical space · Machine learning · Operators · Quantum chemistry · Response properties



**Anders S. Christensen** is a theoretical chemist who works as postdoctoral researcher at the University of Basel in the lab of Anatole von Lilienfeld. His primary work is developing applicable machine learning models for chemistry. He joined the laboratory in Basel in 2016 after two years of postdoctoral work with Qiang Cui at the University of Wisconsin-Madison. He received a PhD in Theoretical Chemistry from the University of Copenhagen in 2014 researching protein folding and Bayesian inference under supervision of Jan H. Jensen, and obtained an MSC in Nanoscience in 2010 from the University of Copenhagen.



**O. Anatole von Lilienfeld** is Associate Professor of Physical Chemistry at the University of Basel. His laboratory develops methods for first principles-based sampling of chemical compound space using quantum mechanics, super computers, Big Data, and machine learning. Previously, he was an associate professor at the Free University of Brussels. From 2013 to 2015 he was a Swiss National Science Foundation professor at the University of Basel. He worked for Argonne and Sandia National Laboratories, and from 2007 to 2010 he was a distinguished Harry S. Truman Fellow at Sandia National Laboratories, after postdoctoral work with Mark Tuckerman at New York University and at the University of California Los Angeles. In 2005, he was awarded a PhD in computational chemistry, under the supervision of Ursula Rothlisberger at EPF Lausanne. His diploma thesis work was done under the supervision of Martin Quack at ETH Zürich and Nick Handy at the University of Cambridge. He studied chemistry at ETH Zürich, the Ecole de Chimie Polymers et Matériaux in Strasbourg, and the University of Leipzig. He is editor-in-chief of the new IOP journal *Machine Learning: Science and Technology*, has been

awarded the Feynman Prize in Nanotechnology, and is an ERC consolidator grantee.

## 1. Introduction

Chemical compound space (CCS) comprises all the theoretically possible molecules and materials one can conceive of. Just for medium-sized organic drug-like molecules it was estimated to exceed  $10^{60}$ ,<sup>[1]</sup> more than atoms in our solar system.<sup>[2]</sup> Corresponding thermodynamic and kinetic stability, as well as most other observable properties can be calculated, often to a fair degree of accuracy, using well established approximations and frameworks of quantum and statistical mechanics.<sup>[3,4]</sup> Unfortunately, however, the resulting numerical complexity is so dramatic that a first principles based exploration of representative swaths of CCS remains prohibitive for all but its smallest subsets. This computational burden can be reduced through the use of machine learning (ML) which, after regressing on a sufficiently large training data set, affords models with controlled test errors and milli-second prediction speed.

While ML has had a long tradition in the chem- and bioinformatics communities for a long time, its application to potential energy surface fitting for the purpose of vibrational analysis, reaction dynamics, or molecular dynamics already started in the nineties using neural networks (NN),<sup>[5–9]</sup> and kernel methods.<sup>[10–13]</sup> By contrast, inferring solutions of the electronic Schrödinger equation throughout compound space was only made possible using Kernel Ridge Regression (KRR) and neural networks (NN) in 2012<sup>[14]</sup> and 2013,<sup>[15]</sup> respectively. Corresponding demonstrations for crystals followed shortly after<sup>[16–19]</sup> and ever since, a substantial number of papers on this topic has been published in this rapidly growing field. For more details, we refer interested readers to the recent machine learning issues in the *Int. J. Quantum Chem.*<sup>[20]</sup> and *J. Chem. Phys.*,<sup>[21]</sup> as well as the overviews given in refs. [22–27]. Due to the rigorous link to the underlying physical laws of quantum and statistical mechanics, we have dubbed this approach ‘Quantum Machine Learning’ (QML), implying that we aim to model quantum properties using classical ML algorithms. This

\*Correspondence: Prof. O. A. von Lilienfeld, E-mail: anatole.vonlilienfeld@unibas.ch  
Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, CH-4056 Basel,

usage of terms is in complete analogy to Quantum Monte Carlo (using classical Monte Carlo to sample electron configuration space), or Quantum Molecular Dynamics (integrating Newton's equations of motion with quantum chemically computed forces). Note, however, that this is not to be confused with ML algorithms implemented on quantum computers.

## 2. Learning Curves

For any QML model to be functional, it must *learn*, *i.e.* its predictive power must improve with increasing training set size  $N$ . Vapnik and others showed decades ago that ML models must, if set-up properly, exhibit this behaviour and converge down to arbitrary accuracy in the limit of infinite training set size.<sup>[28]</sup> In practice, however, arbitrary accuracy is neither required, nor do we operate in large data set limits. Therefore the relevant question for any distribution of data points consists of asking how much more predictive power is to be gained for each additional training instance. This dependence is manifested in learning curves which report prediction (or test) error  $E$  as a function of training set size, as discussed for KRR,<sup>[29]</sup> as well as NNs.<sup>[30]</sup>

In accordance with standard ML protocol,<sup>[31]</sup> we stress the importance of (i) prediction errors being obtained exclusively on hold-out data, *i.e.* for input and output samples which have never been seen during training, and (ii) hyper-parameters, noise-levels, and overfitting being accounted for through cross-validation prior to prediction error evaluation. Leading prediction error terms are known to be inversely proportional to training set size,  $E \propto a/N^b$ , implying convenient linear behavior on log-log scales.<sup>[29]</sup> Off-set  $\log(a)$  and slope  $b$  enable the assessment and comparison of different QML models, and imply that prediction errors of newly developed QML models have to be reported for at least three training set sizes in order to be (a) meaningful, and (b) demonstrate linearity. Absence of linearity on log-log curves typically indicates severe problems and can be due to at least one of any of the following short-comings:

- (i) Not all input variables are being accounted for (*e.g.* due to overly coarse representations), resulting in lack of uniqueness.
- (ii) The model is underdetermined due to lack of model complexity (too few parameters, too rigid functional form).
- (iii) Noisy or inconsistent data.

Meeting the uniqueness criterion is a necessary condition for meaningful models in chemistry (proof is given in ref. [32]). This is not always obvious, however, since it also depends on subtleties in the nature of the target property which defines which degrees of freedom are being sampled. For example, molecular graph based representations, such as SMILES strings, can be appropriate for generating QML models of free energies of solvation for those temperature and pressure combinations for which all conformational degrees of freedom are being averaged out while all covalent bonds are being conserved. However, when it comes to properties which are functions of instantaneous geometries (clamped nuclei), *e.g.* HOMO/LUMO eigenvalues or atomic forces, SMILES are too coarse, and representations encoding all degrees of freedom in a unique fashion are necessary.

Low QML model off-sets  $\log(a)$  and steep slopes  $b$  are highly desirable in order to maximize data efficiency and thereby minimize training data needs in order to maximize transferability through CCS. A considerable amount of recent work has been devoted to this endeavour, demonstrating how to lower off-sets through ever-improving representations,<sup>[33–37]</sup> the use of hierarchies among more approximate base line quantum chemistry models,<sup>[38,39]</sup> or optimal training data-selection.<sup>[40]</sup> Steeper slopes have also been obtained through use of AMons (training on local building-blocks selected on the fly),<sup>[41]</sup> as well as through the operator QML approach discussed below.

## 3. Operator Approach

In 2015, we demonstrated that using the same representation and kernel, KRR-based QML models can account for any QM property.<sup>[42]</sup> By analogy with basic quantum mechanics (knowledge about a single wave-function enables evaluation of expectation values for any property's operator) this finding suggests that the kernel assumes the role of the wavefunction, while KRR regression coefficients carry the units of the property in question. These results represent a meaningful baseline with all QML models being functional for all properties (all learning curves indicated constant and negative slopes  $b$ ). We noticed, however, that slope and off-set were more favorable for some properties than for others. In 2017, similar observations were also made for other representations and regressors, including KRR, NNs, and random forests.<sup>[35]</sup>

It is known that learning curves improve upon explicit encoding of invariant relationships between input features and labels. For example, it is beneficial to use rotational invariant features in order to predict the energy of molecules accurately.<sup>[43]</sup> Motivated by the apparent discrepancy of QML models for atomization energies of small organic molecules reaching predictive power after training on only thousands of examples, and the relatively worse performance of the same models for predicting other properties,<sup>[35]</sup> we developed and introduced the Operator QML (OQML) framework.<sup>[44]</sup> Within kernel-based regression,<sup>[45–48]</sup> we decompose the total potential energy  $U^*$  of a query molecule in its electronic ground-state, into a sum of atomic contributions  $E_I$  for each atom  $I$ ,

$$U^* = \sum_{I \in C} E_I(q_I^*) = \sum_{I \in i} \sum_J \alpha_{Jk} k(q_J, q_I^*) \quad (1)$$

where  $q$  is an atomic environment in some chosen representation basis,  $J$  is training atom  $J$ , and  $\alpha_J$  is its regression weight. As a side note, a comparative discussion of energies of atoms in molecules based on QML, basis-set overlap, and alchemical perturbation DFT has just recently been published.<sup>[49]</sup>

The OQML approach extends this KRR model while still ensuring that regression coefficients are also obtained in closed-form. Within this framework, the relationship between a property, typically the energy, and any of its derivatives up to any order (typical response properties) is explicitly enforced through the application of *response operators* to the learning formalism. More specifically and using matrix notation, the response property  $\omega$ , *i.e.* an observable which can be computed by applying a differential operator  $\mathcal{O}$  acting on the energy  $U^*$ , can be approximated by a finite kernel as

$$\omega = \mathcal{O}[U] \approx \mathcal{O}[\mathbf{K}]\alpha. \quad (2)$$

Here, we assume that the regression coefficients do not depend on the perturbation (in complete analogy to force-field parameters, or the electron density for first order perturbation theory within quantum mechanics).

Then, the set of regression coefficients  $\alpha$  minimizes the Lagrangian

$$J(\alpha) = \sum_{\gamma} \beta_{\gamma} \|\mathcal{O}_{\gamma}[\mathbf{U}^{\text{ref}}] - \mathcal{O}_{\gamma}[\mathbf{K}\alpha]\|_{L_2(\Omega_{\gamma})}^2 \equiv \sum_{\gamma} \beta_{\gamma} \int_{\Omega_{\gamma}} [\mathcal{O}_{\gamma}[\mathbf{U}^{\text{ref}}] - \mathcal{O}_{\gamma}[\mathbf{K}\alpha]]^T [\mathcal{O}_{\gamma}[\mathbf{U}^{\text{ref}}] - \mathcal{O}_{\gamma}[\mathbf{K}\alpha]] \quad (3)$$

over some training set of known derivative values of  $\mathcal{O}[\mathbf{U}^{\text{ref}}]$ .  $\Omega_{\gamma}$  is the domain over which the corresponding operator should be minimized, *e.g.* all rotational degrees of freedom if the operator

acts on a  $SO(3)$  group.  $\gamma$  denotes the specific perturbation, so that the model can be trained for multiple properties simultaneously, for example energies, gradients, and dipole moments. Assuming only one perturbation, the resulting expression for the regression coefficients  $\alpha$  which solves the associated normal equations reads,

$$\alpha = \left( \int_{\Omega} \mathcal{O}[\mathbf{K}]^T \mathcal{O}[\mathbf{K}] \right)^{-1} \left( \int_{\Omega} \mathcal{O}[\mathbf{U}^{\text{ref}}]^T \mathcal{O}[\mathbf{K}] \right) \quad (4)$$

In comparison to the expression for regression coefficients within conventional KRR, it is clear from these equations that kernel derivatives must play a crucial role in the loss function, constraining the resulting regression solution towards a more meaningful manifold of solutions from which the most probable is being obtained. For further details and discussions, we refer the reader to the original work.<sup>[44,50]</sup>

Maybe the most frequent application concerned deals with derivatives with respect to atomic displacement, *i.e.* forces. Much improved learning curves for forces are obtained when compared to KRR-based QML force models which were trained directly on labels corresponding to force vector elements after rotating each atom into a common reference frame with unique handedness, as done for atoms displaced along normal modes for the first time in 2015.<sup>[51]</sup> Within the most recent revision of FCHL, FCHL19,<sup>[50]</sup> we have provided a detailed analysis of the OQML-based force model which can be applied throughout CCS. Using the correct operators to generate appropriate loss-functions ensures that important physics is automatically being enforced, such as rotational covariance of forces, and that the resulting force models conserve energy. A comparative discussion of other state of the art KRR force models, such as SOAP<sup>[12,52]</sup> and GDML<sup>[53,54]</sup> has also been included in ref. [50].

Learning curves on display in Fig. 1 exemplify the improvements gained when exploiting differential relationships through OQML. Learning curves are shown for QML models of the dipole norm of small organic molecules with up to nine atoms of the elemental types CNOF from the QM9 dataset, saturated with hydrogen atoms.<sup>[55]</sup> The corresponding response operator to the dipole is the negative derivative with respect to a change in an external electric field. Results for two kernel-based models are shown, both using the FCHL representation<sup>[37,44]</sup> for the molecules and a Gaussian kernel function. When training the conventional KRR model, the dipole norm is the only scalar label and the loss function merely regresses its deviation from reference. By contrast in the case of OQML, the operator is applied to the loss function of the energy property before the regression (requiring the extension of the representation to also include a simplified electric field model, denoted as FCHL\*). This results in a loss function with very different terms which trivially reproduces the orientational dependence of the applied field and which yields QML models with much improved off-set and slope of learning curve: The OQML-based machine learns the dipole norm to the same accuracy as KRR with 20 times less data.

Similarly, in ref. [44] we have also demonstrated how OQML can be applied to the second order derivatives defining Hessians which, together with dipole-moments, has enabled the first QML-based direct prediction of an infrared spectrum from vibrational analysis.

#### 4. Conclusions

We have briefly motivated the use of QML models, and highlighted the role of learning curves. Thereafter, the OQML approach was described, and illustrated for the case of dipole moment predictions. We stress that, in analogy to perturbation theory going beyond the Hellmann-Feynman theorem, OQML is sufficiently general to also go beyond first order derivatives. In fact, it is straight-forward to apply the OQML-approach to any

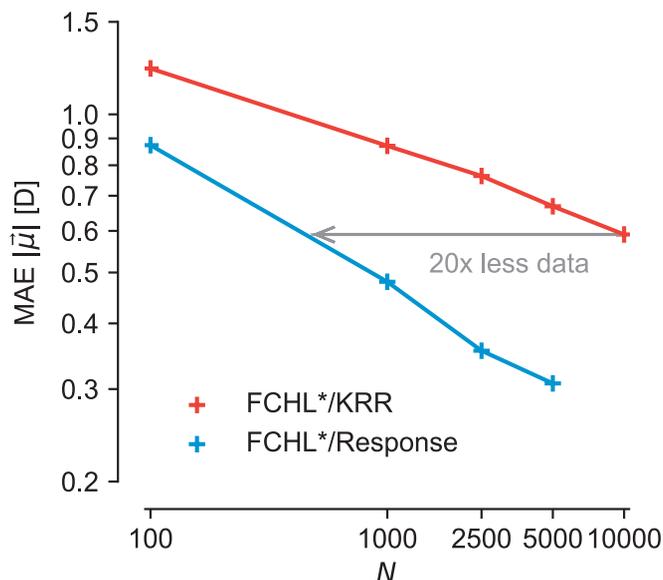


Fig. 1. Improvement of learning curves of dipole moment predictions in QM9 data set<sup>[55]</sup> with (blue) and without (red) response formalism. Both QML models use the FCHL\* representation and a Gaussian kernel function, and differ only by their loss function and the resulting choice of regressor, *i.e.* KRR (red) and OQML (blue). Figure reproduced from ref. [44], licensed under a Creative Commons Attribution (CC BY) license.

property which is a derivative of the energy (or any other reference property), as long as the representation of the molecule can be perturbed accordingly. If and how this framework can also be used for operators other than differential operators remains to be studied.

#### Acknowledgement

We acknowledge support by the Swiss National Science foundation (No. PP00P2 138932, 407540 167186 NFP 75 Big Data, 200021 175747), and from the European Research Council (ERC-CoG grant QML). This work was partly supported by the NCCR MARVEL, funded by the Swiss National Science Foundation.

Received: October 25, 2019

- [1] P. Kirkpatrick, C. Ellis, *Nature* **2004**, 432, 823.
- [2] A. Mullard, *Nature News* **2017**, 549, 445.
- [3] F. Jensen, 'Introduction to Computational Chemistry', West Sussex, England: John Wiley, **2007**.
- [4] M. E. Tuckerman, 'Statistical mechanics: Theory and molecular simulation', Oxford University Press, **2010**.
- [5] B. G. Sumpter, D. W. Noid, *Chem. Phys. Lett.* **1992**, 192, 455.
- [6] S. Lorenz, A. Gross, M. Scheffler, *Chem. Phys. Lett.* **2004**, 395, 210.
- [7] S. Manzhos, T. Carrington, Jr., *J. Chem. Phys.* **2006**, 125, 084109.
- [8] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, 98, 146401.
- [9] J. Behler, *J. Chem. Phys.* **2011**, 134, 074106.
- [10] T. Ho, H. Rabitz, *J. Chem. Phys.* **1996**, 104, 2584.
- [11] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* **2010**, 104, 136403.
- [12] A. P. Bartók, R. Kondor, and G. Csányi, *Phys. Rev. B* **2013**, 87, 184115.
- [13] Z. Li, J. R. Kermode, A. De Vita, *Phys. Rev. Lett.* **2015**, 114, 096405.
- [14] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, *Phys. Rev. Lett.* **2012**, 108, 058301.
- [15] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, *New J. Phys.* **2013**, 15, 095003.
- [16] K. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. Müller, E. Gross, *Phys. Rev. B* **2014**, 89, 205118.
- [17] F. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, *Int. J. Quantum Chem.* **2015**, 115, 1094.
- [18] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, *Phys. Rev. Lett.* **2016**, 117, 135502.
- [19] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, M. A. Marques, *Chem. Mater.* **2017**, 29, 5090.
- [20] M. Rupp, *Int. J. Quant. Chem.* **2015**, 115, 1003.

- [21] M. Rupp, O. A. von Lilienfeld, K. Burke, *J. Chem. Phys.* **2018**, *148*, 241401.
- [22] O. A. von Lilienfeld, *Int. J. Quant. Chem.* **2013**, *113*, 1676.
- [23] O. A. von Lilienfeld, 'Towards the Computational Design of Compounds from First Principles' ,vol. IX, Mathematical Physics Studies, Springer, **2014**.
- [24] R. Ramakrishnan, O. A. von Lilienfeld, 'Machine Learning, Quantum Chemistry, and Chemical Space', vol. 30, pp. 225–256. John Wiley & Sons, Inc., **2017**.
- [25] B. Huang, N. O. Symonds, O. A. v. Lilienfeld, 'Quantum machine learning in chemistry and materials', Handbook of Materials Modeling: Methods: Theory and Modeling, **2018**, pp. 1–27.
- [26] O. A. von Lilienfeld, *Angew. Chem. Int. Ed.* **2018**, *57*, 4164.
- [27] F. A. Faber, O. Anatole von Lilienfeld, *Mater. Informat. Methods, Tools and Applications* **2019**, 171.
- [28] V. Vapnik, 'The nature of statistical learning theory', Springer science & business media, **2013**.
- [29] C. Cortes, L. D. Jackel, S. A. Solla, V. Vapnik, J. S. Denker, *Adv. Neural Inform. Processing Syst.* **1994**, 327.
- [30] K. R. Müller, M. Finke, N. Murata, K. Schulten, S. Amari, *Neural Comp.* **1996**, *8*, 1085.
- [31] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, K. R. Müller, *J. Chem. Theory Comput.* **2013**, *9*, 3404.
- [32] O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp, A. Knoll, *Int. J. Quantum Chem.* **2015**, *115*, 1084.
- [33] K. Hansen, F. Biegler, O. A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, *J. Phys. Chem. Lett.* **2015**, *6*, 2326.
- [34] B. Huang, O. A. von Lilienfeld, *J. Chem. Phys.* **2016**, *145*, 161102.
- [35] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2017**, *13*, 5255.
- [36] W. Pronobis, A. Tkatchenko, K.-R. Müller, *J. Chem. Theory Comput.* **2018**, *14*, 2991, 2018.
- [37] F. A. Faber, A. S. Christensen, B. Huang, O. A. von Lilienfeld, *J. Chem. Phys.* **2018**, *148*, 241717.
- [38] R. Ramakrishnan, P. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2015**, *11*, 2087.
- [39] P. Zaspel, B. Huang, H. Harbrecht, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2018**, *15*, 1546.
- [40] N. J. Browning, R. Ramakrishnan, O. A. von Lilienfeld, U. Roethlisberger, *J. Phys. Chem. Lett.* **2017**, *8*, 1351.
- [41] B. Huang, O. A. von Lilienfeld, *arXiv preprint arXiv:1707.04146*, **2017**, submitted.
- [42] R. Ramakrishnan, O. A. von Lilienfeld, *Chimia* **2015**, *69*, 182.
- [43] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, O. A. von Lilienfeld, K.-R. Müller, *NIPS Proc.* **2012**, *25*, 449.
- [44] A. S. Christensen, F. A. Faber, O. A. von Lilienfeld, *J. Chem. Phys.* **2019**, *150*, 064105.
- [45] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, *IEEE Trans. Neural Networks* **2001**, *12*, 181.
- [46] B. Schölkopf, A. J. Smola, 'Learning with kernels: support vector machines, regularization, optimization, and beyond', MIT press, **2002**.
- [47] V. Vovk, 'Kernel Ridge Regression', Berlin, Heidelberg: Springer Berlin Heidelberg, **2013**, pp. 105–116.
- [48] T. Hastie, R. Tibshirani, J. Friedman, 'The Elements of Statistical Learning: data mining, inference and prediction', Springer series in statistics, New York, N.Y.: Springer, **2001**.
- [49] G. F. von Rudor, O. A. von Lilienfeld, *J. Phys. Chem. B* **2019**, doi: 10.1021/acs.jpccb.9b07799.
- [50] A. S. Christensen, F. A. Faber, L. Bjornholm, D. Glowacki, O. A. von Lilienfeld, *J. Chem. Phys.* **2018**, *148*, 241717.
- [51] M. Rupp, R. Ramakrishnan, O. A. von Lilienfeld, *J. Phys. Chem. Lett.* **2015**, *6*, 3309.
- [52] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816.
- [53] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller, *Sci. Adv.* **2017**, *3*, e1603015.
- [54] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, A. Tkatchenko, *Computer Phys. Commun.* **2019**, *240*, 38.
- [55] R. Ramakrishnan, P. Dral, M. Rupp, O. A. von Lilienfeld, *Sci. Data* **2014**, *1*, 140022.