Stacking Gaussian Processes to Improve pK_a Predictions in the SAMPL7 Challenge

Robert M. Raddi and Vincent A. Voelz*

Department of Chemistry, Temple University, Philadelphia, PA 19122, USA

Abstract

Accurate predictions of acid dissociation constants are essential to rational molecular design in the pharmaceutical industry and elsewhere. There has been much interest in developing new machine learning methods that can produce fast and accurate pKa predictions for arbitrary species, as well as estimates of prediction uncertainty. Previously, as part of the SAMPL6 community-wide blind challenge, Bannan et al. approached the problem of predicting pK_as by using a Gaussian process regression to predict microscopic pK_as, from which macroscopic pK_a values can be analytically computed.[1] While this method can make reasonably quick and accurate predictions using a small training set, accuracy was limited by the lack of a sufficiently broad range of chemical space in the training set (e.g., the inclusion of polyprotic acids). Here, to address this issue, we construct a deep Gaussian Process (GP) model that can include more features without invoking the curse of dimensionality. We trained both a standard GP and a deep GP model using a database of approximately 3500 small molecules curated from public sources, filtered by similarity to targets. We tested the model on both the SAMPL6 and more recent SAMPL7 challenge, which introduced a similar lack of ionizable sites and/or environments found between the test set and the previous training set. The results show that while the deep GP model made only minor improvements over the standard GP model for SAMPL6 predictions, it made significant improvements over the standard GP model in SAMPL7 macroscopic predictions, achieving a MAE of 1.5 pK_a.

Introduction

The negative logarithm of the acid dissociation constant, $pK_a = -\log_{10} K_a$, is fundamentally important in drug design. Absorption, metabolism and distribution of a drug are all greatly affected by the protonation state of the compound under various pH conditions.[2,3] pK_a values may be sought for molecules that have yet to be synthesized, or to further understand fundamental reactions. As a consequence, accurate predictions of acid dissociation constants are essential for pharmaceutical companies as well as many other industries.

The problem of pKa prediction continues to be studied due to its significance, and the difficulty of making accurate predictions. The SAMPL challenge provides a unique opportunity to blindly evaluate model performance for pK_a prediction.[4] In the SAMPL6 challenge, predictions for 24 small molecules were made.[5] A survey of the prediction methods employed by SAMPL6 participants shows that most predictions used QM (quantum mechanical)/linear regression methods, while only a handful of participants used QSPR/ML (quantitative structure property relationship/machine learning) methods. In the former, rigorous quantum mechanical calculations are performed to get standard free energies, which are then fitted to linear regression models of experimental data to extract parameters for LFER (linear free energy relationship). QM methods can achieve very good agreement with experiment, but are typically computationally demanding. Alternatively, QSPR/ML methods have less computational expense, and because of this have recently gained much attention, especially in the pharmaceutical industry.[6–8] QSPR/ML methods can make quick and accurate predictions using a curated database of experimental pK_a data combined with physical, chemical and structural descriptors to be used as a training set for a machine learning model.

One particular QSPR/ML approach used in SAMPL6 was a Gaussian process (GP) model from Bannan et al.[1] This model was trained on physical and chemical descriptors that relate to the deprotonation energy. Ten feature calculations were made for each of the 2,700 (2443 monoprotic) small molecules in a private dataset.[9] Bannan showed that GP models have the generality to produce reasonably accurate predictions and uncertainties in those predictions for any type of ionizable group, performing about the median of all SAMPL6 participants.

Encouraged by this finding, we set about attempting to reproduce these results using a training database of small molecules curated from public sources. However, this endeavor revealed two key flaws with the GP approach: (1) the chemical space in the training set was narrowly limited to only monoprotic acids, and (2) the method performs poorly when there is a lack of suitably similar ionizable sites and/or environments between the training set and the test set.[1]

Here, we attempt to remedy these issues through the use of deep Gaussian process (GP) models, which can increase model robustness when there is low structural similarity between the training and test sets, by enabling a larger number of features to be used. Using a deep GP model, the chemical space in the training set can be expanded to include a greater number of polyprotic molecules, although a large number of monoprotic molecules are still required. We train both a standard GP model and a deep GP model on a set of physiochemical descriptors of molecules from a hand-curated database derived from public sources, and then test both models using molecules from the SAMPL6 and SAMPL7 challenges.

Below, we describe our methodology for constructing a training database of ionizable molecules with experimental pKa measurements, and extracting molecular features to train standard and deep GP models for pK_a prediction. We discuss in detail the molecules included in the SAMPL6 and SAMPL7 challenges, and compare the results of standard and deep GP models on these molecular targets.

Computational Methods

Overview of the method. The Gaussian process models used in this work are trained to predict microscopic pK_a values corresponding to the free energy of dissociating a proton from a specific microscopic species AH $\rightarrow$ A^-. The pK_a value measured in experiment corresponds to the macroscopic pK_a, which can be calculated from the complete set of microscopic pK_a (i.e. the network of all possible single-proton dissociations, see SI)[10]

To predict the experimental pK_a of a molecule, the following steps are performed: First, the input molecule is provided as a SMILES string, along with a network of possible microstate transitions. Next, a set of quantitative physical features are calculated to describe each microscopic transition. These features are used by the Gaussian process model to predict microscopic pK_a values and their uncertainties. Finally, the set of microscopic pK_a values are used to calculate a macroscopic pK_a prediction and its uncertainty.

Featurization of input molecules. The flexibility of the standard Gaussian process model comes from the ability to relate molecular features to the variation in deprotonation energy.[1] For each of the molecules in our training database (Figure 2) (and later, for each molecular pK_a prediction), we first perform a 3-D structure minimization using MMFF96s,[11] and then calculate ten different molecular descriptors as training features. Six of the features are AM1-BCC partial charges[12] computed using Open Force Field[13] and RDKit.[14] These are partial charges of the atom of interest (AOI, the atom from which the proton dissociates), atoms 1 bond away from the AOI, and atoms 2 bonds away from the AOI, in both in A^- and AH forms.

The seventh and eighth features are to changes in solvation free energy and the change in enthalpy along AH $\rightarrow$ A^-, both computed using OpenEye-toolkits.[15] The solvation free energy is the free energy change of moving a species from gaseous phase into dilute aqueous solution. It is calculated using the AM1-BCC partial charges as input to a continuum solvent model and Possion-Boltzmann surface area solver. This calculation is performed on the the lowest-energy gas-phase conformation from the ensemble via the MMFF96s forcefield.

The ninth and tenth molecular features are: the solvent-accessible surface area of the deprotonated AOI calculated via the Shrake-Rupley algorithm,[16] and its partial bond order, calculated using the extended Hückel molecular orbital method to obtain the overlap populations, both calculated using RDKit.

In addition to physicochemical features, structural descriptors are used as features for the deep GP model (see below). There are many methods that successfully utilize topological fingerprints as features,[6,8,17] which are useful for selecting training molecules most similar to the test set of input molecules. Here, we use Morgan fingerprints as features for our deep GP model.[18] In short, Morgan fingerprints are topological descriptors compacted into long bit-vectors (black and white boxes) describing fragments (highlighted red) within a given molecule as shown in Figure 1.

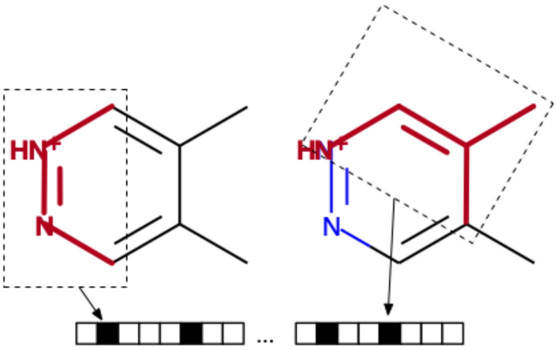

Figure 1: Structural fingerprints are converted into a bit-vector to store information regarding occurrences of specific molecular fragments.

Standard Gaussian process model. Gaussian process (GP) models treat the prediction of microscopic pK_a values (i.e. single protonation/deprotonation equilibria among a network of many possible microtransitions) as a regression problem. The method seeks to estimate (with uncertainty) an unknown function $f(x)$ that is responsible for mapping input features x to microscopic pK_a values to be converted into relative free energies. To do this, a GP is able to model a family of covariance functions that fit a set of known data points (the training set) and use the average of those curves (the posterior mean) to make predictions given new input data.[19,20]

Gaussian process (GP) regression relies on a mean function μ and a covariance function $k(x, x')$. Here, we use a zero-mean Gaussian process. That is, $f(x) \sim GP(\mu(x) = 0, k(x, x'))$ is defined by an expected spatial location in variable space and a relationship by which different variables are correlated with one another. Two well-known software packages that offer Gaussian process regression are Scikit-learn[21] and GPy.[19] We have used both packages and verified results are the same.

The covariance function is often called a *kernel*, responsible for determining the similarity between two input feature pairs x and x'. A collection of these functions represent the joint variability of the model and can be finely tuned i.e., learned to best represent the data. Here, we use the Matérn32 kernel and RBF kernel for deep GP models (see below). The hyperpaprameters are the length scale $l_{1:d}$, which describes how quickly the unknown function changes as x is varied.[22] The optimal choice of these parameters depend upon the training data.

Deep Gaussian process model. Deep neural networks can be used to significantly increase the number of features in a model to get better predictions, but such methods usually require large training sets. One of the major benefits of Gaussian process models is that small training sets can be used. Here, we propose a deep Gaussian process approach that will allow many more features to be used with a relatively small training set. To do this, we stack GPs such that each layer gets the posterior mean from the previous layer including the original inputs. This can be viewed as a composite multivariate function $g(x) = f_l(f_{l-1}(...f_1(X)))$, where f_i is given by a Gaussian process. For this deep model, we use a Python package called DeepGPy.[20]

A curated database from public sources. As a first step in constructing a ML model for pK_a prediction, we curated a database of small molecules with experimentally measured pK_a. Since the training data used by Bannan et al.[1] was proprietary (OpenEye Scientific Software), we opted to hand-curate a custom database from public sources[23–26] amounting to approximately 3500 small molecules with molecular weights not exceeding 500 Da (Figure 2.a). The histogram of pK_a values for each small molecule in the database reveals a bimodal distribution (Figure 2.b) with the highest frequency around a pK_a of 4. The complete database is freely available at `https://github.com/robraddi/GP-SAMPL7`.

Model fitness and selection. To optimize the model, we measure model fitness using a 3-fold cross-validation technique. By this approach, we are able to determine the inherent performance of the model by splitting the data up into separate training and testing sets, to see if a model parameterized by the training data can predict the testing data. Our kernel hyperparameters $\{\sigma, l\}_{1:d}$ for a given model are optimized by selecting the model from the batch of cross-validation experiments that have the lowest error, highest R^2 value and maximum log likelihood. Measuring model fitness in this way can also help select model hyperparameters such as the type of kernel, combinations of kernels, the number of layers, the number of induc-

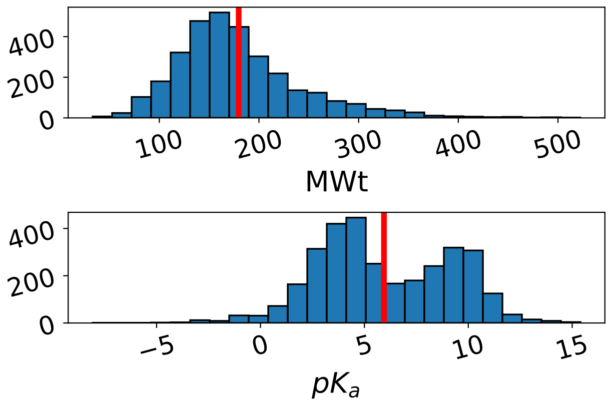

Figure 2: Distribution of (a) molecular weight (avg=180 Da) and (b) pK_a (avg=5.94) for the molecules in the database as described in the main text. Solid red lines denotes the mean.

ing variables, the number of molecules (N) to use for training and even the types of molecules to use—monoprotic or polyprotic.

Damianou et al. incorporate inducing variables inside the deepGPy module to reduce the computational complexity of the model.[20] Inducing variables enable a significant reduction in the number of model parameters for each layer, approximating the true Bayesian posterior by a variational approximation.[20,27]

After many parallel cross-validation experiments, we found that the most robustly predictive models typically have more monoprotic acids as well as having the greatest similarity percentage. Typically, the size (number of molecules) is usually within these limits: $1000 \leq N \leq 1500$, that is, too few or too many compounds in the training set produces weak models (low R^2 and greater error). The similarity percentage metric represents the average structural similarity of each molecule in the training set with the molecule of interest. For example, Table S1 shows average similarity percentage between the training set and the SAMPL7 molecules.

After exploring many different models for the stacked GP, our optimized model uses a RBF kernel, 6 stacked layers with 200 inducing variables per layer and 1474 small molecules in the training set. If we include polyprotic acids inside the training set we limit molecules to have less than four ionizable groups. $-5 \leq pK_a \leq 15$

Similarty filtering to select subsets od relevant training data. Based on our initial cross-validation results, we found that including too many training molecules results in less accurate models. It would therefore be desirable to limit the number of molecules in the dataset, selecting only those with similar structures to a given intended prediction target, i.e. SAMPL molecules, with the idea being that closely-spaced input features should have similar (and more predictable) pK_as. We hypothesized that specifying a similarity threshold to filter the database–leaving only molecules that match the similarity criteria–would result in improved models. The outcomes of these filtering efforts are discussed in Results.

To perform this filtering, Tversky similarity is used with Morgan structural fingerprints.[28] We chose the Tversky similarity metric as it is a general form of Tanimoto similaity metric. We calculated the average similarity between the SAMPL molecules and the molecules inside the training set ($\sim 18\%$ similarity). We then include molecules from the database in our training set if there is a similarity greater or equal to the average with any of the SAMPL molecules.

The SAMPL7 challenge. In the SAMPL7 physical property challenge, pK_a predictions submissions consist of standard state relative free energies of micro-transitions for 22 small molecules, as described in the recent work of Gunner et al.[10] Details for SAMPL7 are available online (http://github.com/samplchallenges/SAMPL7). These details include the experimental measurements for the 22 sulfonamide derivatives,[29] submitted predictions from all participants, and a general analysis of the competition.

Results

Correlation of input features

We first examine the distribution of training molecules in the ten-dimensional feature space used in our GP models. A visualization of the joint probability distribution of any two input features, along with the experimental pK_a, shows some features to be highly correlated, while others are not (Figure 3). The correlation between input features are found by scaling the covariance by the product of the standard deviation of each variable. The greater the linearity, the more closely correlated the two variables are. Correlation contours in Figure 3 emulate the optimized kernel parameters.

Kernel parameters are obtained by using three-fold cross-validation technique and selecting the model with the lowest mean absolute error (MAE), highest R^2 value and largest maximum log likelihood. Highest priority was given to the model with the lowest MAE, then highest R^2. With the optimized kernel parameters for each model, predictions were made for all micro-transitions of SAMPL6 and SAMPL7 molecules. All SAMPL molecules were left outside of the training sets for the results shown below. Six of the SAMPL6 molecules had a few transitions that give rise to known software issues in feature calculations such as free energy of solvation. In this event, the micro-transition is omitted and the micro-pK_a prediction to that micro-transition cannot be made.

Performance of standard and deep GP models

Macroscopic pK_a predictions and their uncertainties are shown in Figure 4 for SAMPL6 (a) and SAMPL7 (b). In both sets of targets, the prediction statistics suggest that the deep GP (blue) has lower mean absolute error and higher R^2 than the standard GP (black).

A technical issue in comparing our predicted macroscopic pK_a with the experimental values is worth mentioning here. Experiments may report a single pK_a value, while our model sometimes predicts multiple macroscopic pK_a (for different deprotonation events). To decide which predicted macroscopic pK_a to compare to experiment, we use a minimal distance criterion, i.e. the smallest difference between the predicted and observed pK_a values. Other studies have used different criteria, such as the Hungarian algorithm.[1]

Performance on SAMPL6 targets. Our standard GP model performed reasonably well at predicting the macroscopic pK_a of SAMPL6 targets, achieving a coefficient of determination R^2 of 0.59, mean absolute error (MAE) of 1.38 and root mean squared error (RMSE) of 1.61 (Figure 4). These results are similar to those of Bannan et al.:[1] $R^2 = 0.48$, MAE of 1.39 and RMSE of 2.16, despite the fact that we use an entirely different database of molecules to train the model.

The deep GP model performed slightly better at predicting the macroscopic pK_a of SAMPL6 targets, achieving an R^2 of 0.61, MAE of 1.36 and RMSE of 1.62.

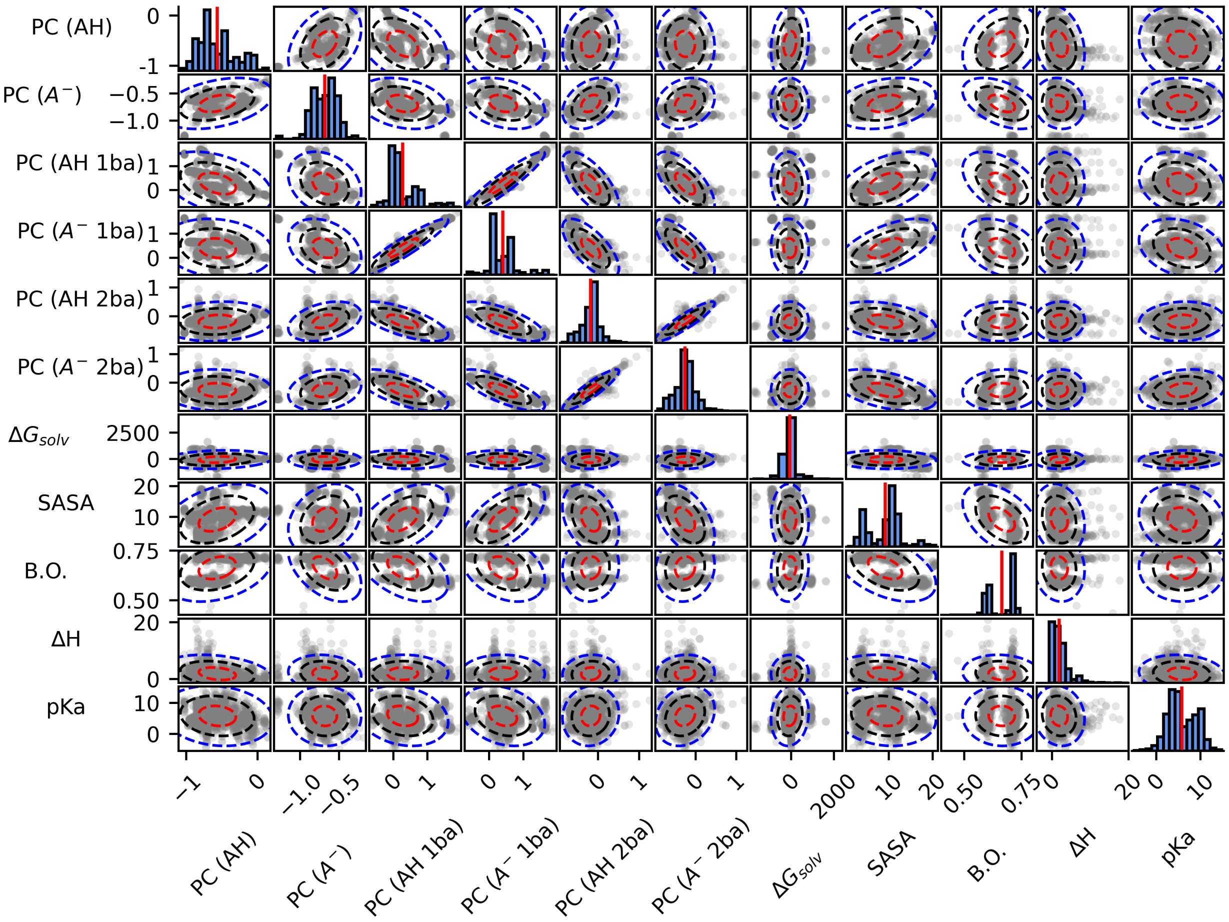

Figure 3: Correlations between input features and pK_a values from the dataset. Input features: AM1BCC partial charge of atom of interest (AOI) protonated (PC (AH)) & deprotonated (PC (A^-)), avg. AM1BCC partial charge of atoms 1 bond away from AOI protonated (PC (AH 1ba)) & deprotonated (PC (A^- 1ba)), avg. AM1BCC partial charge of atoms 2 bond away from AOI protonated (PC (AH 1ba)) & deprotonated (PC (A^- 2ba)), difference in solvation free energy (ΔG_{solv}), SASA (Shrake-Rupley) of AOI (deprotonated), bond order (Mulilken overlap populations) (B.O.), difference in enthalpy (ΔH). Histograms are found along the diagonal with a red line denoting the mean. Correlation curves overlay the raw feature data (gray dots) for 1 σ (red), 2 σ (black) and 3 σ (blue).

Performance on SAMPL7 targets. We anticipated that applying our GP models to SAMPL7 targets would result in poor predictions, due to the lack of similarity between molecules in the training set and the molecules of interest (MOIs) (see Table S1 for similarity percentage). All the compounds in the SAMPL7 challenge contain sulfonamide groups and are difficult to find in open-source databases. Our database contains very few ($\sim$ 40) small molecules with sulfonamide groups.[25] All of these were included in the training set regardless of the number of ionizable sites.

To overcome this issue, we applied a similarity filter to remove from the training set molecules that lacked significantly similarity to our targets, as described in Methods. This filter was used to select a more focused training set for both standard and deep GP models.

Without a similarly filter on the molecules included in the training set, we find the deep GP model predictions produce slightly greater error, with an MAE of 1.7, RMSE of 2.0 and R^2 of 0.64 (Figure S3). These results reflect the benefits of similarity filtering as well as the quality of the deepGP model over the standard GP.

With a similarity filter in place, we expected better results, especially for the deep GP model, which is able to utilize a greater number of effective features. Using the filtered training set, we found that the standard GP model yielded predictions with an R^2 of 0.16, MAE of 3.05 and RMSE of 3.69. In contrast, the deep GP model results gave an R^2 of 0.49, MAE of 1.47 and RMSE of 1.89. These results demonstrate the idea that a deep GP model is more robust than a standard GP model when faced with a paucity of training data that has high correlation with the target molecules.

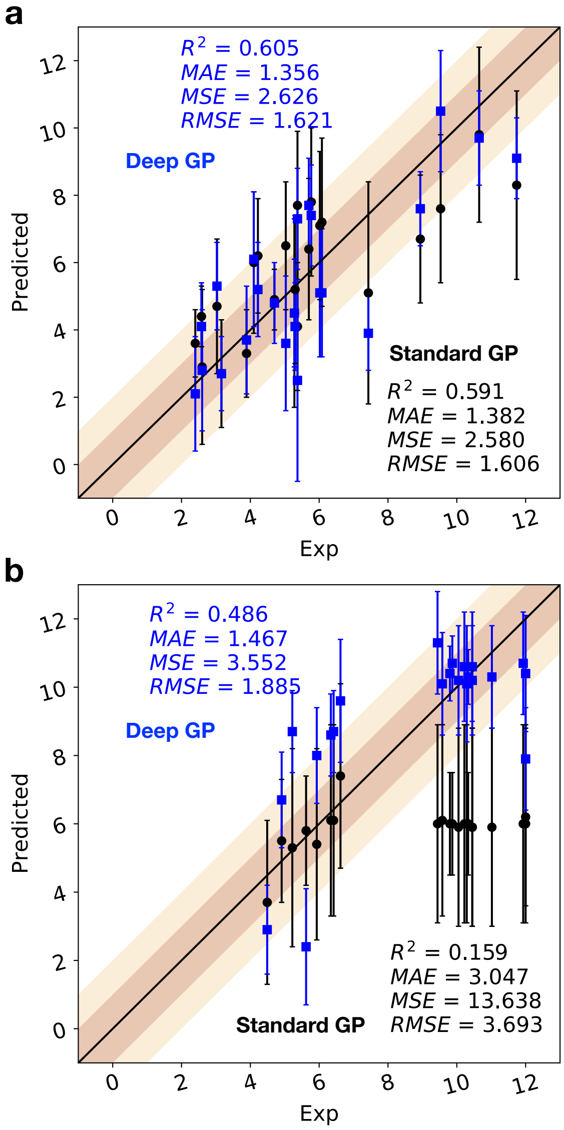

Figure 4: Macroscopic pK_a predictions for (a) SAMPL6 and (b) SAMPL7 compounds. Model statistics for the standard GP are in black (bottom right) and the stacked GP are in blue (top left): determination coefficient (R^2), mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE). Shaded inner region denotes within 1 pK_a and within 2 pK_a for the outer region.

Uncertainty estimates from GP models

Prediction of uncertainties is important to understanding and improving models, regardless of prediction accuracy.[30] In our case, the predicted uncertainties of both standard and deep GP models are helpful in understanding the information each type of uses, and their potential to be successful predictors if presented with more data.

Uncertainties predicted by a standard GP model reflect insufficient training data. One of the goals of this study was to develop a model that was robust when facing uncommon moieties. Arguably, the standard GP model fails this test for the SAMPL7 targets, although its predictions give insight into how GP models function when faced with insufficient training data. Regardless of the target molecule, the standard GP model predicts a pK_a near 6, with large prediction uncertainties. In this case, the GP model deals with unfamiliar input molecules by using the mean pK_a value of the training set ($\sim$ 6) with large error bars as the prediction (Figure 4 (b)). The covariance of the GP random variables are centered about the mean, since the proximity of feature space is far from the molecule of interest and the training set. The lack of correlation between the two sets of input features gives rise to the large uncertainty.

Uncertainties predicted by the deep GP model. Unlike the standard GP model, the deep GP model was able to make relatively accurate predictions with reasonable uncertainties for most molecules. There are only a few instances of poor predictions with misguided uncertainties in the deep GP results. For example, SM22 from SAMPL6 a nitrogen heterocycle with iodide substituents (predicted pK_a of 3.9±1.1, experimental pK_a of 7.43±0.1) and SM28 in SAMPL7 (the only non-sulfonamide) an amide derivative with a nearby sulfone group (predicted pK_a of 7.9 ± 1.5, experimental pK_a of 12).

Upon stacking GPs, predictions are accompanied by relatively smaller uncertainties than those predicted by the standard GP model. For some predictions, however, the deep GP model appears slightly over-confident compared to the standard GP model. The large uncertainties predicted by both models suggest reasonble reporting of uncertainty for SAMPL6 and perhaps even more so for SAMPL7 predictions. Our training set was not strongly conditioned to predict sulfonamide derivatives, and therefore lacked the ability to make accurate predictions.

Why is the deep GP model more predictive than the standard GP model? The deep GP takes into account a larger number of effective features, both by stacking GPs and by filtering the training set by additional structural features.

When trying to predict pKas for molecules that are not found inside the training set, our results show the advantages of reducing the number of molecules inside the training set to be more selective. In doing so, inputs are able to have closer correlations with the training data. Since GP models work well with small training sets without effecting the validity of the model, a similarity filtering approach was implemented to filter out irrelevant compounds.

Discussion

It has been previously shown that Gaussian process regression models are general enough to make predictions with uncertainties for any ionizable group.[1] However, QSAR/ML methods are limited by the prior information. Here, we have shown that GP models can be improved even with limited structural diversity in our training sets. The improvement we observe for deep GP models in this study begs the question of how predictive this approach might be if trained on high-quality commercial-available databases. In this case, we would expect a significant increase in the accuracy the GP model predictions in this case.

Previous work has shown the benefits of curating a diverse training set. For example, Simulation Plus ADMET Predictor from Fraczkiewicz et al.[6] recently partnered with Bayer Pharmaceuticals. Before that, Simulation Plus strictly used public data. With the inclusion of experimental data from Bayer Pharma, however, prediction statistics significantly improved, with R^2 values increasing from 0.87 to 0.93 and MAE decreasing from 0.72 to 0.50.[6]

Here, even in the case of limited publicly available data, we have shown that deep GP models, trained on similar molecules to the target, can lead to significant improvements over standard GP models. Our calculations show that a deep GP model yields more accurate results and increases the robustness of the model without a large/diverse training set. By extending the standard model to a deep GP model we are able to include more features and also allow for a slight increase in the number of polyprotic molecules. While including polyprotic molecules increases the accuracy of macroscopic pK_a predictions, it comes at a cost to predicted relative free energies. In theory the free energy ΔG_{cycle} of any protonation/deprotonation thermodynamic cycle should be zero, but this is not enforced by GP models, leading to increased errors for polyprotic molecules.

A potential future direction for this work would be to study additional features as input to the deep GP model. Since the deep GP model permits higher dimensionality, additional descriptors than the ones describeds above could be explored in future challenges. These include: polar surface area of the protonated atom of interest (AOI), different types of structural fingerprints, HOMO and LUMO energies surrounding the AOI, polarizability, and more. As for the level of theory involved in feature calculations, improvements to increase the level of accuracy would be most beneficial for estimating the free energy of solvation. One concern, however, would be if computational expense would outweigh the performance gain.

Overall, participating in the SAMPL7 challenge was a great way to demonstrate how the standard GP model suffers with a mediocre training set, and how deep GP models that stack GPs and filter out irrelevant molecules can overcome these limitations to an extent. The SAMPL physical property challenges provide excellent target molecules for testing our QSAR/ML models. We look forward to future SAMPL challenges, where we can apply more diverse training sets and incorporate many of the lessons learned here.

The database, code and results are publicly available at https://github.com/robraddi/GP-SAMPL7.

Acknowledgements

RMR and VAV are supported by National Institutes of Health grant R01GM123296. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine)

References

(1) Bannan, C. C.; Mobley, D. L.; Skillman, A. G. SAMPL6 challenge results from pK_a predictions based on a general Gaussian process model. *Journal of computer-aided molecular design* **2018**, *32*, 1165–1177.

(2) Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *Journal of medicinal chemistry* **2008**, *51*, 817–834.

(3) Manallack, D. T.; Prankerd, R. J.; Yuriev, E.; Oprea, T. I.; Chalmers, D. K. The significance of acid/base properties in drug discovery. *Chemical Society Reviews* **2013**, *42*, 485–496.

(4) SAMPL Challenge. https://www.samplchallenges.org.

(5) Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D.; Mobley, D. L. Assessing the accuracy of octanol–water partition coefficient predictions in the SAMPL6 Part II log P Challenge. *Journal of Computer-Aided Molecular Design* **2020**, 1–36.

(6) Fraczkiewicz, R.; Lobell, M.; Goller, A. H.; Krenz, U.; Schoenneis, R.; Clark, R. D.; Hillisch, A. Best of both worlds: Combining pharma data and state of the art modeling technology to improve in silico p K a prediction. *Journal of chemical information and modeling* **2015**, *55*, 389–397.

(7) Shields, G. C.; Seybold, P. G. *Computational approaches for the prediction of pKa values*; CRC Press, 2013.

(8) Fraczkiewicz, R. In silico prediction of ionization. **2013**,

(9) pka Prospector, OpenEye Scientific Software.

(10) Gunner, M. R.; Murakami, T.; Rustenburg, A. S.; Işık, M.; Chodera, J. D. Standard state free energies, not pK as, are ideal for describing small molecule protonation and tautomeric states. *Journal of Computer-Aided Molecular Design* **2020**, 1–13.

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

(12) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry* **2002**, *23*, 1623–1641.

(13) Wagner, J. et al. openforcefield/openforcefield: 0.8.0 Virtual Sites and Bond Interpolation. 2020; https://doi.org/10.5281/zenodo.4121930.

(14) Landrum, G., et al. RDKit: Open-source cheminformatics. **2006**,

(15) Software, O. S. Cheminformatics Software: Molecular Modeling Software: OpenEye Scientific. http://www.eyesopen.com./.

(16) Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of molecular biology* **1973**, *79*, 351–371.

(17) Xing, L.; Glen, R. C.; Clark, R. D. Predicting p K a by molecular tree structured fingerprints and PLS. *Journal of chemical information and computer sciences* **2003**, *43*, 870–879.

(18) Gobbi, A.; Poppinger, D. Genetic optimization of combinatorial libraries. *Biotechnology and bioengineering* **1998**, *61*, 47–54.

(19) GPy, GPy: A Gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.

(20) Damianou, A.; Lawrence, N. Deep gaussian processes. Artificial Intelligence and Statistics. 2013; pp 207–215.

(21) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **2011**, *12*, 2825–2830.

(22) Duvenaud, D. The Kernel cookbook: Advice on covariance functions. *URL https://www. cs. toronto. edu/˜ duvenaud/cookbook* **2014**,

(23) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y., et al. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *Journal of computer-aided molecular design* **2011**, *25*, 533–554.

(24) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z., et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* **2018**, *46*, D1074–D1082.

(25) Caine, B. A.; Bronzato, M.; Popelier, P. L. Experiment stands corrected: accurate prediction of the aqueous p K a values of sulfonamide drugs using equilibrium bond lengths. *Chemical science* **2019**, *10*, 6368–6381.

(26) Settimo, L.; Bellman, K.; Knegtel, R. M. Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds. *Pharmaceutical research* **2014**, *31*, 1082–1095.

(27) Titsias, M. Variational learning of inducing variables in sparse Gaussian processes. Artificial intelligence and statistics. 2009; pp 567–574.

(28) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling* **2010**, *50*, 742–754.

(29) Francisco, K. R.; Varricchio, C.; Paniak, T. J.; Kozlowski, M. C.; Brancale, A.; Ballatore, C. Structure Property Relationships of N-Acylsulfonamides and Related Bioisosteres. *European Journal of Medicinal Chemistry* **2021**, 113399.

(30) Nigam, A.; Pollice, R.; Hurley, M. F. D.; Hickman, R. J.; Aldeghi, M.; Yoshikawa, N.; Chithrananda, S.; Voelz, V. A.; Aspuru-Guzik, A. Assigning Confidence to Molecular Property Prediction. 2021.

(31) Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. Multiconformation, density functional theory-based p K a prediction in application to large, flexible organic molecules with diverse functional groups. *Journal of chemical theory and computation* **2016**, *12*, 6001–6019.

Supporting Information

for

Stacking Gaussian Processes to Improve pK_a Predictions in the SAMPL7 Challenge

Robert M. Raddi and Vincent A. Voelz

Analytical computation of macroscopic pK_a using a thermodynamic cycle

Our models predict microscopic pK_a values for each microtransition. We use the following expression from Bochevarov et al[31] to convert microscopic pK_as to macroscopic pK_as, which relates microdissociation constants K_{ij} for each microtransition to total equilibrium concentrations to compute macroscopic equilibrium constants K_a.

$$K_a = \sum_{j=1}^{N_{\text{deprot}}} \frac{1}{\sum_{i=1}^{N_{\text{prot}}} \frac{1}{K_{ij}}} \tag{1}$$

Relative free energies for microtransitions are computed using the microscopic pK_a inside the following relationship as described by Gunner et al.[10] We then converted to kcal mol^{-1}, where $C = 1.36$ kcal mol^{-1}.

$$G_{ij} = nH_{ij}C(-pK_a) \tag{2}$$

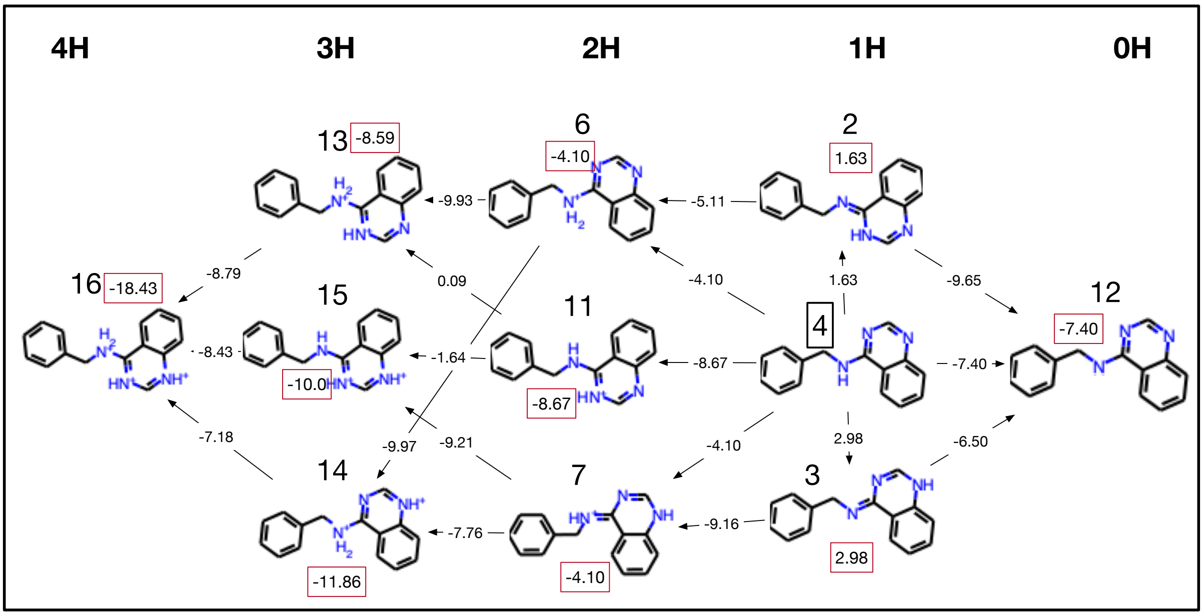

Figure S1: Relative free energies (red boxes) in kcal mol^{-1} of SM07 from SAMPL6, where state 4 is the reference state. Microtransitions are denoted by black arrows with their respective negative micro-pK_a on the inset. Largest difference in energy over all cycles of length 4 correspond to the cycle with states 4,11,13,6 and gives $\Delta G_{cycle} = 14.59$. Note that all other cycles give $\Delta G_{cycle} \leq 6$. Results can be compared with Figure 4B and Table 4 found in Gunner et al.[10]

Model selection & cross-validation

A short example of the selection process is shown in Table S1, where six cross-validation experiments were performed using a Matérn32 kernel: three experiments using only monoprotic acids increasing the number small molecules in the training sets and three additional tests of both monoprotic and polyprotic acids. The best model (shown in bold font) uses a training set of size $N = 1380$, results in $R^2 = 0.85$ and a mean absolute error of 0.91.

Table S1: 3-fold model validation varying the number of compounds in the training set using standard Gaussian process model.

	N	Similarity (%)	MAE	RMSE	R^2
	742	17.68	1.217	1.868	0.678
Mono	**1380**	**18.15**	**0.909**	**1.355**	**0.845**
	2224	17.69	1.149	1.749	0.721
	1038	17.78	1.633	2.195	0.537
Mono & Polyprotic	2105	17.70	0.978	1.465	0.805
	3114	17.59	1.332	1.935	0.639

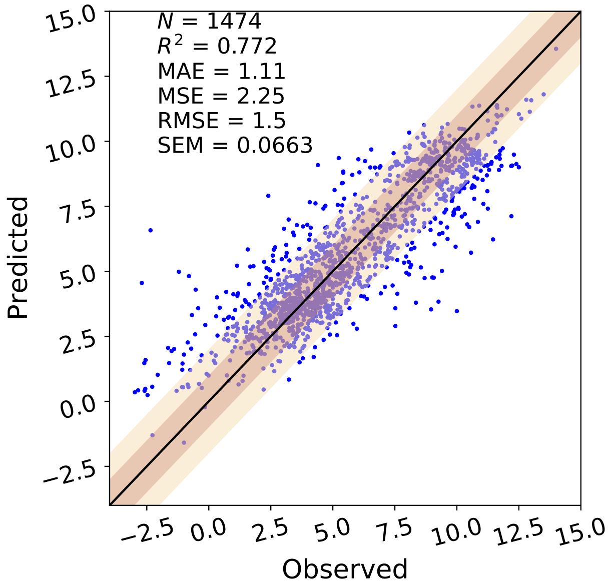

Figure S2: DeepGP model with 6 layers, 200 inducing variables, and 1474 molecules in training. The optimized hyperparameters $\{\sigma, l\}_{1:d}$ from this were used to make the actual predictions.

Table S2: SAMPL7 (SM25-46) relative free energy predictions from standard GP.

Ref	Microstate ID	Charge	Prediction
SM25 m000	SM25 m001	-1	16.48 ± 7.83
SM25 m000	SM25 m003	-1	8.28 ± 3.94
SM25 m000	SM25 m005	1	0.17 ± 5.53
SM25 m000	SM25 m002	0	8.06 ± 3.95
SM26 m000	SM26 m001	-1	7.8 ± 2.69
SM26 m000	SM26 m003	-1	-0.0 ± 0.0

Continued on next page

Table S2 – continued from previous page

Ref	Microstate ID	Charge	Prediction
SM26 m000	SM26 m005	1	-5.63 ± 3.3
SM26 m000	SM26 m002	0	-8.29 ± 3.88
SM27 m000	SM27 m001	-1	8.0 ± 3.95
SM28 m000	SM28 m002	-1	7.43 ± 3.42
SM28 m000	SM28 m004	-1	-0.0 ± 0.0
SM28 m000	SM28 m003	1	-6.09 ± 3.03
SM28 m000	SM28 m001	0	-8.5 ± 3.87
SM29 m000	SM29 m001	-1	8.06 ± 3.95
SM30 m000	SM30 m001	-1	8.11 ± 3.95
SM31 m000	SM31 m001	-1	8.07 ± 3.95
SM31 m000	SM31 m002	1	-8.09 ± 3.87
SM32 m000	SM32 m001	-1	8.08 ± 3.95
SM33 m000	SM33 m001	-1	8.14 ± 3.95
SM34 m000	SM34 m001	-1	8.09 ± 3.95
SM34 m000	SM34 m002	1	-8.0 ± 3.88
SM35 m000	SM35 m001	-1	8.14 ± 3.95
SM35 m000	SM35 m003	-1	8.08 ± 3.95
SM35 m000	SM35 m002	0	0.04 ± 5.59
SM36 m000	SM36 m001	-1	8.12 ± 3.95
SM36 m000	SM36 m003	-1	8.12 ± 3.95
SM36 m000	SM36 m002	0	-0.02 ± 5.59
SM37 m000	SM37 m002	-1	8.18 ± 3.95
SM37 m000	SM37 m004	-1	8.16 ± 3.95
SM37 m000	SM37 m001	1	-8.13 ± 3.88
SM37 m000	SM37 m005	1	-8.02 ± 3.89
SM37 m000	SM37 m003	0	-0.01 ± 5.58
SM38 m000	SM38 m001	-1	8.17 ± 3.95
SM39 m000	SM39 m001	-1	8.2 ± 3.95
SM40 m000	SM40 m001	-1	8.19 ± 3.95
SM40 m000	SM40 m002	1	-8.24 ± 3.85
SM41 m000	SM41 m001	-1	8.23 ± 3.9
SM41 m000	SM41 m002	1	-7.25 ± 3.88
SM42 m000	SM42 m001	-1	10.1 ± 3.75
SM42 m000	SM42 m002	1	-8.02 ± 3.92
SM42 m000	SM42 m003	0	1.91 ± 5.4
SM43 m000	SM43 m001	-1	8.94 ± 3.77
SM43 m000	SM43 m002	1	-7.74 ± 3.9
SM43 m000	SM43 m005	1	-8.1 ± 3.91
SM43 m000	SM43 m003	2	-15.43 ± 7.73
SM44 m000	SM44 m001	-1	8.27 ± 3.87
SM44 m000	SM44 m002	1	-7.39 ± 3.73
SM45 m000	SM45 m001	-1	8.84 ± 3.3
SM45 m000	SM45 m002	1	-7.41 ± 3.79
SM46 m000	SM46 m001	-1	8.27 ± 3.87
SM46 m000	SM46 m002	1	-7.37 ± 3.59

Continued on next page

Table S2 – continued from previous page

Ref	Microstate ID	Charge	Prediction
SM46 m000	SM46 m004	1	−7.81 ± 3.89
SM46 m000	SM46 m003	2	−16.23 ± 7.77

Table S3: SAMPL7 (SM25-46) relative free energy predictions from deep GP.

Ref	Microstate ID	Charge	Prediction
SM25 m000	SM25 m001	-1	24.48 ± 3.47
SM25 m000	SM25 m003	-1	12.28 ± 1.98
SM25 m000	SM25 m005	1	−3.3 ± 5.76
SM25 m000	SM25 m002	0	7.79 ± 4.85
SM26 m000	SM26 m001	-1	6.0 ± 1.79
SM26 m000	SM26 m003	-1	−0.0 ± 0.0
SM26 m000	SM26 m005	1	−4.36 ± 1.75
SM26 m000	SM26 m002	0	−9.16 ± 1.95
SM27 m000	SM27 m001	-1	13.93 ± 2.11
SM28 m000	SM28 m002	-1	7.89 ± 2.61
SM28 m000	SM28 m004	-1	−0.0 ± 0.0
SM28 m000	SM28 m003	1	−6.05 ± 1.74
SM28 m000	SM28 m001	0	−10.79 ± 2.13
SM29 m000	SM29 m001	-1	13.85 ± 2.23
SM30 m000	SM30 m001	-1	13.72 ± 2.32
SM31 m000	SM31 m001	-1	14.02 ± 2.05
SM31 m000	SM31 m002	1	−14.02 ± 2.12
SM32 m000	SM32 m001	-1	14.44 ± 2.2
SM33 m000	SM33 m001	-1	14.17 ± 2.36
SM34 m000	SM34 m001	-1	14.54 ± 2.09
SM34 m000	SM34 m002	1	−14.17 ± 2.14
SM35 m000	SM35 m001	-1	14.53 ± 2.3
SM35 m000	SM35 m003	-1	14.57 ± 2.11
SM35 m000	SM35 m002	0	0.03 ± 3.13
SM36 m000	SM36 m001	-1	14.14 ± 2.17
SM36 m000	SM36 m003	-1	14.16 ± 2.18
SM36 m000	SM36 m002	0	0.13 ± 3.06
SM37 m000	SM37 m002	-1	14.38 ± 2.12
SM37 m000	SM37 m004	-1	14.36 ± 2.08
SM37 m000	SM37 m001	1	−14.28 ± 2.11
SM37 m000	SM37 m005	1	−14.06 ± 2.16
SM37 m000	SM37 m003	0	0.0 ± 3.01
SM38 m000	SM38 m001	-1	15.39 ± 2.08
SM39 m000	SM39 m001	-1	14.46 ± 2.15
SM40 m000	SM40 m001	-1	14.54 ± 2.07
SM40 m000	SM40 m002	1	−13.8 ± 2.02
SM41 m000	SM41 m001	-1	11.78 ± 1.68
SM41 m000	SM41 m002	1	−2.18 ± 2.27
SM42 m000	SM42 m001	-1	13.06 ± 2.62

Continued on next page

Table S3 – continued from previous page

Ref	Microstate ID	Charge	Prediction
SM42 m000	SM42 m002	1	-11.27 ± 1.82
SM42 m000	SM42 m003	0	1.67 ± 3.13
SM43 m000	SM43 m001	-1	15.5 ± 2.05
SM43 m000	SM43 m002	1	-12.38 ± 2.05
SM43 m000	SM43 m005	1	-11.7 ± 1.7
SM43 m000	SM43 m003	2	-26.68 ± 4.08
SM44 m000	SM44 m001	-1	11.73 ± 1.68
SM44 m000	SM44 m002	1	-3.69 ± 2.19
SM45 m000	SM45 m001	-1	10.89 ± 1.84
SM45 m000	SM45 m002	1	-3.42 ± 2.24
SM46 m000	SM46 m001	-1	11.82 ± 1.64
SM46 m000	SM46 m002	1	-2.35 ± 2.25
SM46 m000	SM46 m004	1	-13.11 ± 2.14
SM46 m000	SM46 m003	2	-28.38 ± 4.2

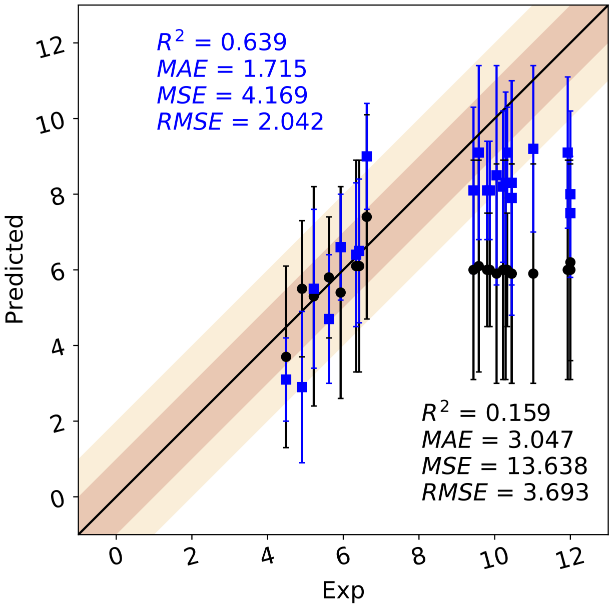

Figure S3: DeepGP model without similarity filtering. Model consists of 6 layers, 200 inducing variables, and 1474 randomly selected database molecules to be used in training. The optimized hyperparameters $\{\sigma, l\}_{1:d}$ from this were used to make the actual predictions.

Table S4: SAMPL6 (SM01-24) and SAMPL7 (SM25-46) macro pKa results for standard GP.

(a) SAMPL6

ID	Prediction	Experiment
SM01	7.6 ± 2.2	9.53 ± 0.01
SM02	6.5 ± 1.9	5.03 ± 0.01
SM03	NaN	7.02 ± 0.01
SM04	7.1 ± 2.2	6.02 ± 0.01
SM05	NaN	4.59 ± 0.01
SM06	4.7 ± 2.0	3.03 ± 0.04
	8.3 ± 2.8	11.74 ± 0.01
SM07	7.2 ± 2.5	6.08 ± 0.01
SM08	6.2 ± 1.7	4.22 ± 0.01
SM09	7.7 ± 2.2	5.37 ± 0.01
SM10	NaN	9.02 ± 0.01
SM11	3.3 ± 1.3	3.89 ± 0.01
SM12	4.5 ± 2.8	5.28 ± 0.01
SM13	7.8 ± 2.2	5.77 ± 0.01
SM14	4.4 ± 0.9	2.58 ± 0.01
	5.2 ± 2.2	5.30 ± 0.01
SM15	4.9 ± 0.9	4.70 ± 0.01
	6.7 ± 1.9	8.94 ± 0.01
SM16	4.1 ± 1.9	5.37 ± 0.01
	9.8 ± 2.6	10.65 ± 0.01
SM17	2.7 ± 1.6	3.16 ± 0.01
SM18	NaN	2.15 ± 0.02
	NaN	9.58 ± 0.03
	NaN	11.02 ± 0.04
SM19	NaN	9.56 ± 0.02
SM20	6.4 ± 2.1	5.70 ± 0.03
SM21	6.0 ± 2.1	4.10 ± 0.01
SM22	3.6 ± 1.0	2.40 ± 0.02
	5.1 ± 3.3	7.43 ± 0.01
SM23	NaN	5.45 ± 0.01
SM24	2.9 ± 2.3	2.60 ± 0.01

(b) SAMPL7

ID	Prediction	Experiment
SM25	3.7 ± 2.4	4.49 ± 0.04
SM26	5.5 ± 1.8	4.91 ± 0.01
SM27	5.9 ± 2.9	10.45 ± 0.01
SM28	6.2 ± 2.6	12.00 ± NaN
SM29	5.9 ± 2.9	10.05 ± 0.01
SM30	6.0 ± 2.9	10.29 ± 0.12
SM31	5.9 ± 2.9	11.02 ± 0.01
SM32	5.9 ± 2.9	10.45 ± 0.02
SM33	6.0 ± 2.9	12.00 ± NaN
SM34	6.0 ± 2.9	11.93 ± 0.05
SM35	6.0 ± 1.5	9.87 ± 0.01
SM36	6.0 ± 1.5	9.80 ± 0.06
SM37	6.0 ± 1.5	10.33 ± 0.02
SM38	6.0 ± 2.9	9.44 ± 0.02
SM39	6.0 ± 2.9	10.22 ± 0.15
SM40	6.1 ± 2.8	9.58 ± 0.01
SM41	5.3 ± 2.9	5.22 ± 0.01
SM42	7.4 ± 2.7	6.62 ± 0.02
SM43	5.8 ± 1.6	5.62 ± 0.02
SM44	6.1 ± 2.8	6.34 ± 0.01
SM45	5.4 ± 2.8	5.93 ± 0.05
SM46	6.1 ± 2.8	6.42 ± 0.01

Table S5: SAMPL6 (SM01-24) and SAMPL7 (SM25-46) macro pKa results for deep GP.

(a) SAMPL6

ID	Prediction	Experiment
SM01	10.5 ± 1.8	9.53 ± 0.01
SM02	3.6 ± 2.0	5.03 ± 0.01
SM03	NaN	7.02 ± 0.01
SM04	5.1 ± 1.9	6.02 ± 0.01
SM05	NaN	4.59 ± 0.01
SM06	5.3 ± 1.3	3.03 ± 0.04
	9.1 ± 1.2	11.74 ± 0.01
SM07	5.1 ± 1.9	6.08 ± 0.01
SM08	5.2 ± 1.4	4.22 ± 0.01
SM09	7.3 ± 1.5	5.37 ± 0.01
SM10	NaN	9.02 ± 0.01
SM11	3.7 ± 1.6	3.89 ± 0.01
SM12	4.5 ± 1.6	5.28 ± 0.01
SM13	7.4 ± 1.5	5.77 ± 0.01
SM14	4.1 ± 1.3	5.3 ± 0.01
	4.1 ± 1.3	2.58 ± 0.01
SM15	7.6 ± 1.1	8.94 ± 0.01
	4.8 ± 1.2	4.7 ± 0.01
SM16	2.5 ± 3.0	5.37 ± 0.01
	9.7 ± 1.4	10.65 ± 0.01
SM17	2.7 ± 1.1	3.16 ± 0.01
SM18	NaN	2.15 ± 0.02
	NaN	9.58 ± 0.03
	NaN	11.02 ± 0.04
SM19	NaN	9.56 ± 0.02
SM20	7.7 ± 1.4	5.7 ± 0.03
SM21	6.1 ± 2.0	4.1 ± 0.01
SM22	3.9 ± 1.1	7.43 ± 0.01
	2.1 ± 1.7	2.4 ± 0.02
SM23	NaN	5.45 ± 0.01
SM24	2.8 ± 1.8	2.6 ± 0.01

(b) SAMPL7

ID	Prediction	Experiment
SM25	2.9 ± 1.3	4.49 ± 0.04
SM26	6.7 ± 1.4	4.91 ± 0.01
SM27	10.2 ± 1.6	10.45 ± 0.01
SM28	7.9 ± 1.5	$12.0\pm$
SM29	10.2 ± 1.6	10.05 ± 0.01
SM30	10.1 ± 1.7	10.29 ± 0.12
SM31	10.3 ± 1.5	11.02 ± 0.01
SM32	10.6 ± 1.6	10.45 ± 0.02
SM33	10.4 ± 1.7	$12.0\pm$
SM34	10.7 ± 1.5	11.93 ± 0.05
SM35	10.7 ± 0.8	9.87 ± 0.01
SM36	10.4 ± 0.8	9.8 ± 0.06
SM37	10.3 ± 0.8	10.33 ± 0.02
SM38	11.3 ± 1.5	9.44 ± 0.02
SM39	10.6 ± 1.6	10.22 ± 0.15
SM40	10.1 ± 1.5	9.58 ± 0.01
SM41	8.7 ± 1.2	5.22 ± 0.01
SM42	9.6 ± 1.8	6.62 ± 0.02
SM43	2.4 ± 1.7	5.62 ± 0.02
SM44	8.6 ± 1.2	6.34 ± 0.01
SM45	8.0 ± 1.4	5.93 ± 0.05
SM46	8.7 ± 1.2	6.42 ± 0.01

Table S6: SAMPL7 (SM25-46) micro pKa results for deep GP.

Protonated	Deprotonated	Prediction
SM25 m000	SM25 m001	8.99 ± 1.27
SM25 m002	SM25 m001	8.43 ± 1.35
SM25 m004	SM25 m001	5.82 ± 1.33
SM25 m000	SM25 m003	9.02 ± 1.45
SM25 m002	SM25 m003	4.37 ± 1.22
SM25 m004	SM25 m003	3.95 ± 1.43
SM25 m005	SM25 m000	9.16 ± 2.04
SM25 m002	SM25 m000	5.72 ± 3.56
SM25 m005	SM25 m002	8.15 ± 2.27
SM25 m005	SM25 m004	2.86 ± 1.28
SM26 m000	SM26 m001	4.40 ± 1.31
SM26 m002	SM26 m001	8.04 ± 1.51
SM26 m004	SM26 m001	8.34 ± 1.75
SM26 m000	SM26 m003	3.85 ± 1.54
SM26 m002	SM26 m003	8.99 ± 1.24
SM26 m004	SM26 m003	4.67 ± 2.06
SM26 m005	SM26 m000	3.20 ± 1.28
SM26 m002	SM26 m000	6.73 ± 1.43
SM26 m005	SM26 m002	6.29 ± 2.97
SM26 m005	SM26 m004	8.66 ± 1.80
SM27 m000	SM27 m001	10.24 ± 1.55
SM28 m000	SM28 m002	5.80 ± 1.92
SM28 m001	SM28 m002	8.92 ± 1.31
SM28 m000	SM28 m004	7.70 ± 1.44
SM28 m001	SM28 m004	9.53 ± 1.15
SM28 m003	SM28 m000	4.45 ± 1.27
SM28 m001	SM28 m000	7.93 ± 1.56
SM28 m003	SM28 m001	6.85 ± 2.85
SM29 m000	SM29 m001	10.18 ± 1.64
SM30 m000	SM30 m001	10.09 ± 1.70
SM31 m000	SM31 m001	10.31 ± 1.50
SM31 m002	SM31 m000	10.30 ± 1.56
SM32 m000	SM32 m001	10.61 ± 1.62
SM33 m000	SM33 m001	10.41 ± 1.73
SM34 m000	SM34 m001	10.69 ± 1.53
SM34 m002	SM34 m000	10.41 ± 1.57
SM35 m000	SM35 m001	10.68 ± 1.69
SM35 m002	SM35 m001	10.66 ± 1.56
SM35 m000	SM35 m003	10.71 ± 1.55
SM35 m002	SM35 m003	10.68 ± 1.57
SM35 m002	SM35 m000	10.63 ± 1.53
SM36 m000	SM36 m001	10.39 ± 1.59
SM36 m002	SM36 m001	10.30 ± 1.58
SM36 m000	SM36 m003	10.41 ± 1.60
SM36 m002	SM36 m003	10.40 ± 1.58
SM36 m002	SM36 m000	9.77 ± 1.50
SM37 m000	SM37 m002	10.57 ± 1.55
SM37 m003	SM37 m002	10.56 ± 1.57
SM37 m000	SM37 m004	10.55 ± 1.53
SM37 m003	SM37 m004	10.45 ± 1.55
SM37 m001	SM37 m000	10.50 ± 1.54
SM37 m005	SM37 m000	10.34 ± 1.58
SM37 m003	SM37 m000	9.95 ± 1.46
SM37 m001	SM37 m003	10.28 ± 1.56
SM37 m005	SM37 m003	10.17 ± 1.55
SM38 m000	SM38 m001	11.31 ± 1.53
SM39 m000	SM39 m001	10.63 ± 1.57
SM40 m000	SM40 m001	10.69 ± 1.52
SM40 m002	SM40 m000	10.14 ± 1.48
SM41 m000	SM41 m001	8.66 ± 1.23
SM41 m002	SM41 m000	1.59 ± 1.67
SM42 m000	SM42 m001	9.60 ± 1.92
SM42 m003	SM42 m001	8.37 ± 1.26
SM42 m002	SM42 m000	8.28 ± 1.33
SM42 m003	SM42 m000	8.72 ± 1.36
SM42 m002	SM42 m003	1.34 ± 1.66
SM43 m000	SM43 m001	11.39 ± 1.50
SM43 m004	SM43 m001	8.49 ± 1.24
SM43 m002	SM43 m000	9.09 ± 1.51
SM43 m005	SM43 m000	8.59 ± 1.24
SM43 m003	SM43 m000	9.80 ± 1.49
SM43 m002	SM43 m004	9.93 ± 1.59
SM43 m005	SM43 m004	0.35 ± 1.55
SM43 m003	SM43 m002	2.37 ± 1.71
SM43 m003	SM43 m005	10.78 ± 1.55
SM44 m000	SM44 m001	8.62 ± 1.23
SM44 m002	SM44 m000	2.70 ± 1.61
SM45 m000	SM45 m001	8.00 ± 1.35
SM45 m002	SM45 m000	2.51 ± 1.64
SM46 m000	SM46 m001	8.68 ± 1.20
SM46 m002	SM46 m000	1.72 ± 1.65
SM46 m004	SM46 m000	9.64 ± 1.57
SM46 m003	SM46 m000	10.43 ± 1.54
SM46 m003	SM46 m002	10.88 ± 1.53
SM46 m003	SM46 m004	3.46 ± 1.62