Calculating the Binding Entropy of Host-Guest Systems with Physics-Guided Neural Networks

Alles Rebel
Department of Computer Science
California State University, Los Angeles
Los Angeles, United States
arebel@calstatela.edu

Ali Risheh
Department of Computer Science
California State University, Los Angeles
Los Angeles, United States
arisheh@calstatela.edu

Negin Massoudian
Department of Electrical and Computer Engineering
California State University, Los Angeles
Los Angeles, United States
nmassou2@calstatela.edu

Negin Forouzesh
Department of Computer Science
California State University, Los Angeles
Los Angeles, United States
neginf@calstatela.edu

***Abstract*—In-silico calculation of binding free energy between protein and ligands has vast applications in the early stages of drug discovery. Most of the classical physics-based models, including implicit solvents, ignore entropy contributions from the system. Instead, a simplified solvent entropy is indirectly considered. This simplification is often done because of an under-sampled conformal space due to physics calculation complexity. Machine learning (ML) methods offer a practical venue to incorporate accurate binding entropy predictions from the experiment. While accurate, there are growing concerns about the overfitting of ML models to the training set, lack of interpretation due to its "black box" characteristics, and failure to comply with well-known physical models. Recently emerged, physics-guided models are a class of ML models that combine the robust consistency of physics-based models with the accuracy of modern data-driven algorithms. This work presents a method to design two hybrid models by coupling ML with a physics model. Implementing these hybrid models have been done through careful modification of various model learning parameters or hyperparameters. The proposed hybrid models not only outperform purely data-driven models by at least 10% but also show more consistent performance on both training and test sets. We review the basic theory, investigate binding entropy calculation methods, present hybrid models that take advantage of end-point simulation software, and analyze the performance of these models.**

***Index Terms*—Binding Free Energy, Entropy, Machine Learning, In-silico Drug Discovery**

I. Introduction

Drug discovery and development can span over 12 years and cost over $1 billion [1]. Average cost is reported to be $2.6 billion in 2016 [2]. The drug search finds and evaluates candidate compounds capable of activating or deactivating specific biological targets through conformational changes; the development process involves delivery, toxicity, testing, etc. [3]. High-throughput screening in the early stages of drug discovery uses quick computational methods favoring lower computation time over accuracy [4]. Later stages focus on accuracy at the expense of speed.

Binding entropy plays a major role in determining the change in Gibbs free energy (ΔG) of a reaction; in larger systems, the accuracy of entropy becomes critical for determining ΔG due to increased structure flexibility and solvent interactions [5]. The early stage computer-aided screening is often accomplished with implicit solvent free energy calculation methods such as Molecular Mechanics, Poisson–Boltzmann Solvent Area (MM/PBSA) approaches [6]–[8]. These methods often ignore the contribution of binding entropy due to its computational complexity [5]. Note that these implicit solvent models capture solvent entropy contributions in their polar and non-polar terms and are reliant on parameters in their model to estimate this important contribution [9]–[12].

Quantum methods, such as Ab Initio [13], take significant computational time and are often used for smaller structures; they become impractical when used on larger structures. Hence, finding methods that reduce the computational time and cost of high-level ab initio calculations, especially for large systems, is significant [14]. Calculating the entropy with standard quantum thermochemical methods can be challenging. Appropriate consideration in flexible molecules needs sampling all thermally accessible conformational degrees of freedom and calculating the translational, rotational, and vibrational partition functions. Binding entropy is composed of differences between the unbound and bound states of the systems. Some important components include configurational entropy and solvation entropy. We will focus the discussion on configurational entropy due to being ignored for end-point methods in general. Configurational entropy is associated with the number of ways a molecule can be represented. It has been applied to proteins and other biomolecular systems [15]. To calculate the configurational entropy, the possible conformations of the molecule are separated into a finite number of states. Any of them has been assigned energy

This research has been funded by NSF 2216858 and NIH R16GM146633 awarded to N.F.

calculated with force fields. The calculation of configurational entropy is time-consuming, primarily due to the need for a thorough conformal search. Building models that estimate accurate configurational entropy from the data can offer insight by predicting binding entropy net contribution to free energy.

Physics-guided neural network (PGNN) is a systematic framework for combining the scientific knowledge of physics-based models with neural networks to advance scientific discovery. There are two primary contributions of PGNN: first, it presents an approach to create hybrid combinations of physics-based models and neural network architectures to make full use of both physics and data. Second, it presents a novel framework for training neural network architectures using the knowledge contained in physics-based equations to ensure the learning of physically consistent solutions [16]–[18].

In this study, we evaluate the accuracy of four popular models for calculating the change in entropy, ΔS. The most accurate method is chosen as the physics-based component of the two proposed hybrid models. These models couple together the accuracy of data-driven machine learning and the consistency of the selected physics-based model. The first hybrid model is based on the Morgan fingerprint [19], also known as extended-connectivity fingerprint [4], which encodes molecular structural characteristics as a vector. The second model is a graph convolution network, which uses a simple encoding of the molecular graph, made of atoms and bonds, to take greater advantage of the information in the molecular structure. This paper is organized as follows:

- Material and Methods: describes the dataset and physics-based models used to calculate the binding entropy and comparison between the models.
- Proposed Hybrid Models: two physics-guided deep learning models are introduced with their featurizers and architecture.
- Result and Discussion: performance of the two proposed hybrid models is assessed and compared with the reference experiments and physics methods.

II. Material and Methods

A. Dataset: Host-Guest Systems

Host–guest systems have been adopted in different applied chemistry fields, including drug development, materials sciences, analytical separation sciences, chemical pollutant cleanup technology, and the agrochemical industry [20]. Due to these systems being smaller than most protein-ligand systems, rapid experimentation and simulation are possible. In addition, these systems are rigid and favorable for software-based conformal searching [21]. In this study, past Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenges and benchmark sets [22] have been utilized. These datasets were put together to evaluate the physical properties of host-guest systems through computational means using experimentally derived references. These systems are provided in repositories containing source files for a host-guest system conformer and with a valid protonation state; however, it is suggested to produce own set of best conformers and protonation as the given one may not be ideal.

SAMPL8 challenge focuses on binding 22 drugs of abuse with the host cucurbit8uril (CB8) [22]. The benchmark dataset used for training and testing contains alpha-cyclodextrin as the host structure and 22 guest structures with a range of 15-25 atoms, including 1-octylammonium and 6-heptenoate. Both datasets include experimental data for binding free energy, binding enthalpy, and binding entropy. The experimental data is validated using Isothermal Titration Calorimetry (ITC) and Nuclear Magnetic Resonance (NMR) spectroscopy.

B. Methodologies

1) Normal Mode Analysis Methodology: Normal mode analysis (NMA) is a post-processing method to compute vibrational modes and, therefore, vibrational entropy from a simulation trajectory. The Amber package [23] was used for implementing NMA [24] on a trajectory generated by the classical molecular dynamics simulator SANDER. The CB8-Meth system conformation was taken as-is from the source repository, parameterized using General Amber Force Field (GAFF), neutralized with sodium or chloride counter-ions using Antechamber, and solvated using the TIP3P water model (1459 water molecules added) through LEaP. The resulting input files were then minimized and heated from 0K to 300K. Equilibrium was verified by observing temperature, total energy, and observing root-mean-square deviation (RMSD) of ring atomic positions in CB8. The time for this process was recorded and included in the final computational time. In addition, starting structures of the host and ligand were taken from this relaxed system.

With the resulting equilibrium structure, a conformation search was performed with a standard classical molecular dynamics simulation. The production runs a combined total of 20 ns with a recording of coordinates every 10 ps to generate a trajectory. Finally, to extract normal modes, the recorded trajectory is used to produce a mass-weighted covariance matrix. Then the covariance matrix is diagonalized to get extract eigenvalues and vectors to produce eigenmodes. The eigenmodes are converted to frequencies and aggregated into the thermodynamic components, including entropy. All these calculations are implemented in CppTraj in Ambertools through $diagmatrix$ function.

2) Verachem Mining Minima (VM2) Methodology: In a very similar process to the NMA methodology, VM2 [25] begins with the raw conformer file. GAFF is used to parameterize the structure. Following the parameterized, charges are neutralized using VM2's vcharge program. Then VM2 attempts to find as many conformers as possible, keeping track of calculated energies. This search is designed to be as exhaustive as possible and efficient (TORK) [26] by attempting to explore the conformal space by rotating or translating non-rigid bonds to find the steepest change in energy. After every search step, a relaxation step is performed to allow conformal changes, moving towards energy minima. Various guest poses

are tried while the host is fixed in place. Once searches yield little to no energy changes, the resulting conformer is recorded and checked if energetically similar to others (and combined if similar). Once the search reveals no new conformers or has exhausted search time, the final binding free energy is calculated using energies associated with each found conformer. The resulting free energy is computationally fast while maintaining accuracy correlated with the experiment. However, for this work, we focus just on the entropy components reported by the software suite.

3) BEERT Regression Model Methodology: Binding Entropy Estimation for Rotation and Translation (BEERT) is a method to estimate rotational and translational entropy, $\Delta S_{R/T}$, based on the difference between the bound and unbound rotational and translational volumes of a ligand [27]. $\Delta S_{R/T} = \Delta S_R + \Delta S_T = \Delta S_R^C - (\Delta S_R^H + \Delta S_R^G) + \Delta S_T^C - (\Delta S_T^G + \Delta S_T^H)$, represents the sum of changes in rotational and translational binding entropies. These values are the difference between the bound state (complex) subtracted from the sum of ligand and host separately.

To extract the necessary features for this model, a minimization is required. Once the system has been minimized, a short molecular dynamics simulation is executed with the same constraints as those used during the minimization process to produce a trajectory sampling the conformations near the low energy conformation. Using CPPTraj, the center of mass coordinate is calculated for the starting structure and for each coordinate in the trajectory - this difference is then summed together and averaged based on the number of snapshots. Finally, this number is multiplied by the Boltzmann constant, resulting in $\Delta S_{R/T}$ estimation. This method covers translational and rotational contributions; however, due to the computational complexity of vibrational entropy calculations, it is suggested to use alternative methods accounting for this term [27].

4) OLE Regression Model Methodology: Ordinary least squares (OLS) is a linear least squares method to approximate the unknown parameters in a linear regression model. It minimizes the sum of the squares of the differences between those predicted by the linear function of the independent variable and the observed dependent variable. In this approach [15], conformer ensembles and the corresponding entropies of over 120,000 small molecules were evaluated with up to 20 rotatable bonds and comprising over 12 million conformers to develop models to predict conformational entropy across a wide range of molecules. To extract the necessary features for the model, RDKit was used.

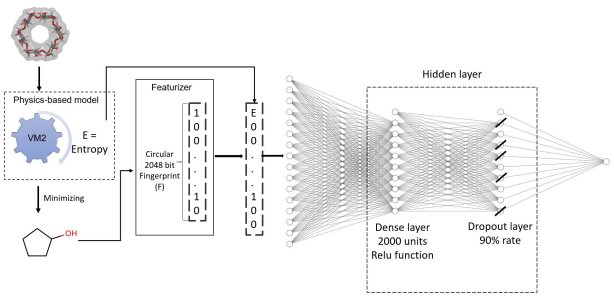
III. Proposed Hybrid Models

A graph convolution model and a traditional deep neural network were developed using PyTorch and the Deepchem open-source framework [28]. Both models were evaluated using Root Mean Squared Error (RMSE) to evaluate prediction error for binding entropy. 10-fold cross-validation was used due to lack of data, where each fold contained an 80% training : 20% test data split, with no early stopping. Both models had their hyper-parameters optimized using a grid search against RMSE. Both models take advantage of the physics calculations embedded into both the minimized structure data and the calculations to produce the estimated binding entropy from VM2. ADAM optimizer was utilized along with dropout regularization to reduce dependence on the input data. Each model employs a specific featurizer to generate the fingerprints; the explanation of each model featurizer is discussed after the model definition. In addition, the number of layers, number of hidden units, and dropout percentage are tuned using a simple grid search for the parameters [29], resulting in the lowest RMSE across 100 epochs.

A. Hybrid Model I: The Physics Guided Linear Network

The input data is the host-guest system that will have its binding entropy estimated. The data is fed into two primary pathways: first, VM2 to perform a conformer search and produce a binding entropy prediction, then the resulting data is concatenated to VM2 predicted entropy and, next, fed into the corresponding linear model. The neural network architecture is a stacking of one dense and dropout layer (hidden layer). The number of layers, neurons, learning rate, and rate of dropout layer is estimated by DeepChem Grid Hyper-parameter Optimization. According to Figure 1, each hidden layer contains a densely connected layer with 2000 units and $ReLU$ activation function followed by a dropout function to prevent overfitting. The learning rate is equal to 0.001.

Fig. 1. Physics guided linear network model



B. Featurizer I: Extended-Connectivity Fingerprints

Extended-connectivity fingerprints (ECFPs) [4] are derived using a variant of the Morgan algorithm [19]. The ECFP generation process has three sequential stages:

- An initial assignment stage in which each atom has an integer identifier assigned to it.
- An iterative updating stage in which each atom identifier is updated to reflect the identifiers of each atom's neighbors
- A duplicate identifier removal stage in which multiple occurrences of the same feature are reduced to a single representative in the final feature list

First, atoms are assigned integer identifiers, *e.g.*, atomic numbers. These initial atom identifiers are collected into an initial fingerprint set. Next, each atom is assigned a unique identifier, and the identifiers of its immediately neighboring atoms are

into an array. To avoid order dependence, the neighbors are ordered using their identifiers and the order of the attaching bonds. A hash function is applied to reduce the array of identifiers back into a new, single-integer identifier. Once all atoms have generated their new identifiers, they replace their old identifiers with their new identifiers. The new atom identifiers are added to the fingerprint set. This iteration is repeated a specified number of times. Once completed, duplicate identifiers in the set are removed, and the remaining integer identifiers in the fingerprint set define the ECFP fingerprint. The process is shown in Figure 2.

The output of the featurizer is a 2048 binary array which is a unique fingerprint for each molecule and provides a compact version of the feature matrix. This fingerprint uniquely captures the topology of the input and uniquely characterizes information such as neighboring atoms types and number of bonds and substructures numerically [4]. In this study, in addition to the output of the featurizer, an additional term is appended which is a float representing the estimated binding entropy generated by VM2.

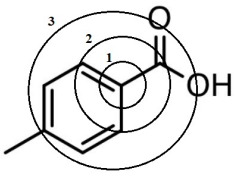


Fig. 2. ECFP process overview over a guest molecule in SAMPL8 dataset known as 4-Hydroxybenzoic. Each circle and number indicate the molecules involved in the hash and the step number, respectively. In each round, the atoms get a new hash number until all atoms are considered, and the final fingerprint is generated.

C. Hybrid Model II: The Physics Guided Graph Convolution Network

The main difference between Hybrid Model I and Hybrid Model II is their featurizers. The linear model used hashed molecular fingerprints capturing topography directly while the graph convolution network uses graph convolution featurizer to generate molecular graph input fingerprint. First, a minimized system is produced by our physics-based model, along with a predicted binding entropy. The predicted binding entropy is concatenated to each atom feature vector in the input. Next, the minimized system is passed into a graph convolution featurizer to generate a representation of the minimized structure in which features of each atom are summed with nearby atoms. Finally, the modified molecular feature matrix is fed to two graph convolution layers, followed by a pooling layer, a normalization layer, and a graph gather layer to generate the internal fingerprint vector for final prediction through dense layers. The process is shown in Figure 3.

D. Featurizer II: Molecular Graph Convolutions

In 2015 a new approach was proposed for generating molecular fingerprint based on a convolution deep learning

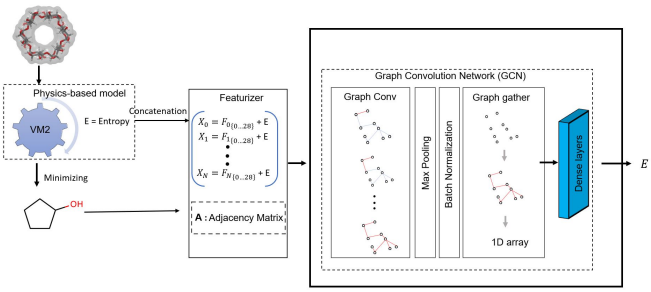


Fig. 3. Physics guided graph convolution network model

algorithm [30]. The first generation of this featurizer used a linear combination of atoms and neighbors to produce a new feature for each atom. By using a Softmax function, all new atomic features are transformed into a fixed-size of vector to be fed to a neural network for extracting molecular features. Molecular Graph Convolutions (MolGraphConv) [31] featurizer used in this work, introduces two new concepts for generating fingerprints. The first concept is introducing an atom layer that contains an n-dimensional feature vector associated with each atom. Therefore the atom layer is a two-dimensional matrix indexed first by the atom and then by specific features of that atom. The next concept is a pair layer which contains an n-dimensional vector associated with each pair of atoms modeling a relationship between specific atoms. Note that the pair input can contain information not just about the direct edges of that atom but about any arbitrary atom in the structure. The end result is a featurizer that represents features contributed by a specific atom and relationships between the atoms through a graph representation.

This differs from the prior featurizer in structure and to how information is modeled. The prior featurizer focused on topographic or similar substructures throughout the system; MolGraphConv featurizer works by representing individual atoms and their relation to other atoms and allows a network to find the hierarchical structure through the convolution of the extracted features. In this study, the estimated binding entropy is inserted as a single float to each atom-specific feature vector. Inserting the same value into all atoms is based on the idea that binding entropy applies to every atom of the entire complex regardless of relationship between atoms.

E. Baseline Model: Purely Data-driven Model

To provide a baseline for comparison between the proposed hybrid models, a purely data-driven model is generated with the non-minimized SMILES fed into the non-modified featurizer. All other aspects of the model match exactly to the proposed hybrid model, including learning rate, layers, dropout, and other hyperparameters. These networks were hyperparameter optimized separately; however, both hybrid and purely data-driven models resulted in the same optimal hyperparameters. The same training procedure and testing procedure were applied to the purely data-driven models as the proposed hybrid models.

IV. RESULTS AND DISCUSSION

A. Selection of the Physics-based Model

The accuracy of the four physics-based methodologies (*i.e.*, NMA, VM2, BEERTm, and OLE) in calculating binding entropy is tested on a host-guest system, see Figure 4. Included is also the time taken to produce the result. The goal of this preliminary survey is to assess the accuracy of the aforementioned methods. The final result of this survey is the selection of the method that our proposed hybrid models will utilize to generate the physics component for the input structures.

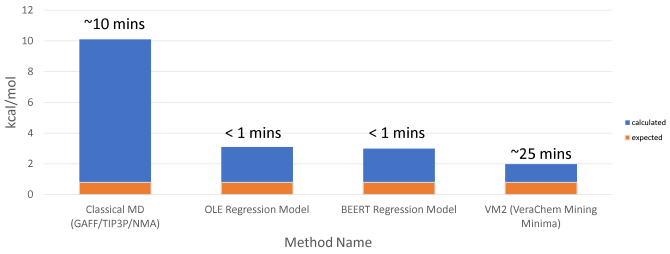


Fig. 4. All methods run upon CB8-Methamphetamine system. Orange indicates the value derived from the experiment. Blue is the value generated by applying the method. Time in minutes is transposed above each corresponding method, indicating the total computation time.

The results shown in Figure 4 indicates VM2 as the most accurate entropy result of the methods surveyed. The output of VM2, in particular, includes a minimized structure, conformer appearance frequencies, and total binding free energy (and therefore binding entropy) estimation. All these items made VM2 a good candidate for developing a hybrid model. Although estimates are provided for binding free energy, binding enthalpy, and binding entropy - due to the use of implicit solvation - binding entropy does not include all solvent contributions. In addition, preliminary results indicate that both OLE and BEERT regression models, models that are statistical or data-driven, have accurate results. This further motivated the development of the hybrid model, coupling together the statistical nature of machine learning with the physics-based information generated from simulation tools.

There are different versions of VM2 available for binding free energy calculations. Table I shows the average performance of five versions of VM2 across all host-guest systems with the standard deviation. This work chose to utilize VM2's Random search simulation mode due to the lowest error across all structures.

TABLE I
ACCURACY OF DIFFERENT VERSIONS OF VM2 IN TERMS OF CALCULATING BINDING ENTROPY.

Simulation Type	RMSE (kcal/mol)
Fast VM2 Random	15.20 ± 2.06
Fast VM2 Single	15.95 ± 1.84
VM2 Random	14.77 ± 2.02
VM2 Single	14.90 ± 2.06

B. Training Phase: Cross-Validation

The convergence over epoch and the associated error can be seen in Figure 5 and Figure 6. It is demonstrated that both hybrid models converged for every fold; the converged error on the training set of data is less than 0.01 on Model I on 0.1 on Model II across all folds. This level of error indicates that Model I's internal representation aligns closely with the output of the features. On the other hand, Model II's featurizer appears to be significantly more sensitive to topological changes and better at normalizing the dependence on input features. During training, no batch regularization was used due to the size of the data set. Dropout regularization was used on each epoch. Both Model I and Model II took about 60 epochs to have all folds converge to a similar error on the training set.

Fig. 5. Model I, Physics Guided Deep Neural Model, convergence across folds.

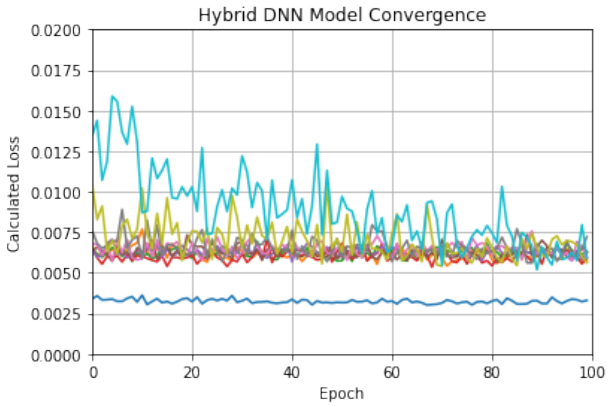
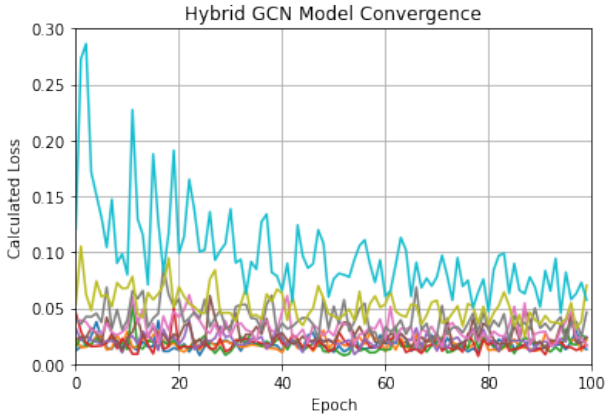


Fig. 6. Model II, Physics Guided Graph Convolution Model, convergence across folds.



C. Test Phase: Model Accuracy

Figure 7 shows error averaged across all folds between a purely data-driven model and our proposed hybrid models on test data. The figure indicates the impact to accuracy and precision due to incorporation of physics information. Comparing the purely data-driven results (*i.e.*, DNN and GCN)

against the hybrid models (*i.e.*, PGDNN and PGGCN), a significant jump in accuracy is observed.

Fig. 7. Model Performance (RSME) on Test Sets averaged across all folds.

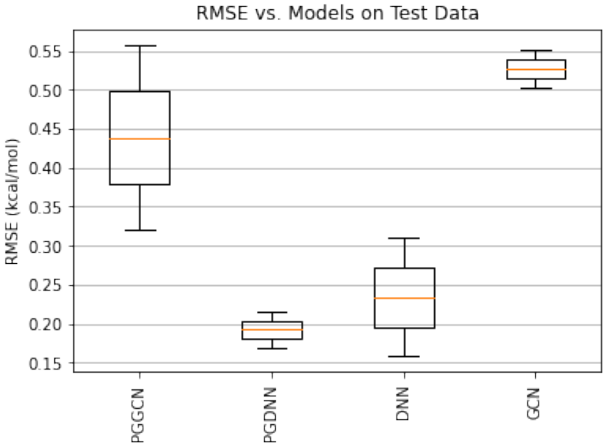


Table. II shows the RSME on the training dataset as well as test dataset. While the predictive performance is better overall on the test dataset, the performance on the training set averaged across all folds indicate there is additional headroom for model improvement. It should be noted that the hybrid models show more consistent results when executed on the training and test sets. This consistency is the result of embedding physics-based features in the model which makes it more "transferrable" across different sets.

TABLE II
AVERAGE RMSE OF MODELS IN KCAL/MOL ACROSS ALL FOLDS.

Model	Training Set	Test Set
Pure data-driven DNN	0.19 ± 0.03	0.23 ± 0.10
Pure data-driven GCN	0.75 ± 0.08	0.52 ± 0.03
Model I (PGDNN)	0.24 ± 0.11	0.19 ± 0.03
Model II (PGGCN)	0.47 ± 0.12	0.43 ± 0.16

D. Correlation Analysis

Both Figure 8 and Figure 9 show model error across training and test sets between a purely data-driven model and our proposed hybrid model. This analysis aims to understand what adding physics-based features to the original data-driven model changes. While Figure 8 does not indicate a large difference between data-driven and hybrid model performance on the same system - Figure 9 shows good correlation between data-driven and hybrid for larger errors; however, for smaller errors, the hybrid model outperformed the data-driven model.

V. CONCLUSION

Purely data-driven models can outperform physics-based models but are always at the risk of overfitting. Another critique is that the results of data-driven models often lack meaningful interpretation due to the "black box" characteristics of machine learning. When data-driven models are coupled with physical models, the resulting hybrid model inherits high accuracy from the former and interpretability from the latter.

Fig. 8. Model I Correlation Plot across both the train and test sets of data.

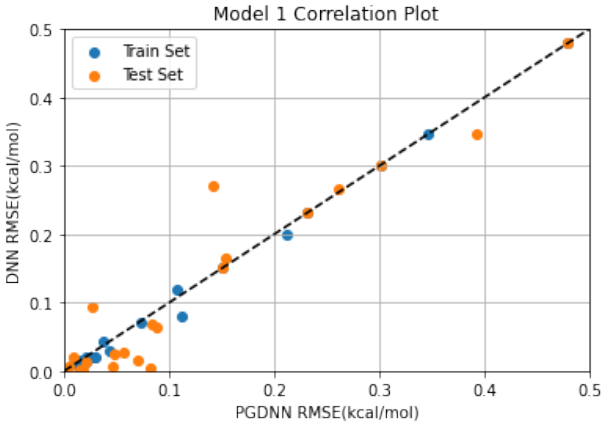
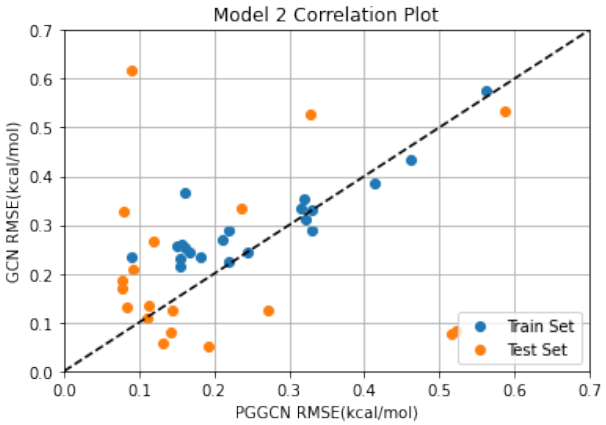


Fig. 9. Model II Correlation Plot across both the train and test sets of data.



Earlier works look at the application of hybrid models applied to binding free energy estimated by purely looking at binding enthalpy - our work differs by looking at the application of hybrid models on binding entropy calculations. This work surveys various binding entropy methods, selects a candidate method, and proposes an architecture to combine data-driven machine learning models with the selected physics-based method.

The publicly available SAMPL and benchmark datasets provided experimental data needed to run simulations for binding entropy estimation. Initially, we started with a host-guest system from SAMPL8 [22], surveying the computational complexity and accuracy of several methods. The accuracy of the NMA method on the entropy contribution is directly tied to how much of the conformal space is captured during the conformer search. In the BEERT model, the local well near the energy minima was considered, and the model worked very well with minor computational overhead. In the OLE regression model, entropy predictions generated from selected structural features and weights trained on many small structures performed accurately with minor computational overhead as well. Finally, the VM2 method was found to be the most accurate method, with the highest computational overhead of

the methods surveyed.

The primary reason for selecting VM2 in this work was due to its accuracy resulting from a comprehensive conformal search. When evaluating the different simulation modes of VM2, the outcome appeared to have a dependence on the input conformer. The full production run of VM2, in either random search or fixed ligand, converged to similar results, highlighting the importance of a long conformation search. Although VM2 analysis needs more computational time, the top-down approach of finding conformers via TORK produces an accurate entropy contribution calculation. Training data with rigid host-guest systems can guide VM2 closer to the experimental data. Here we implement this idea by introducing a physics-guided neural network. This hybrid model fuses VM2 with training data to close the accuracy gap in binding entropy calculation.

There are many improvements that could be made to this work, such as extending work to larger structures, adding additional entropy-specific features, using other fingerprinting methods, or investigating other network typologies. Another area is utilizing other physics-based models; for example, Verachem has developed a version of its mining minima algorithm that utilizes ab-inito calculations for specific active areas to further increase accuracy. Or alternatively, other conformations sampling techniques could be used and compared. The key component of future work is understanding how to maximize the knowledge contained in the input data. While this work identified that the minimized structure contained significant information and the addition of features to allow machine learning models to utilize, there may be faster, more accurate, or more information-dense calculations or features available. Ultimately, coupling data obtained through physics computation with modern data-driven methods offer an excellent opportunity to accurately and reliably predict binding entropy - resulting in a more informed drug screening process.

Data and Model Availability

Both the data and model information can be found publicly available online at https://github.com/allesrebel/cd_set1_gaff_vcharge and https://github.com/allesrebel/physics-guided-entropy.

Acknowledgment

We thank the Extreme Science and Engineering Discovery Environment (XSEDE) for providing computational resources used in this research. We appreciate Verachem LLC for providing us access to VM2 entire software suite to perform experiments, validation of published works, and technical support. This research has been partially funded by NIH R16GM146633 and NSF 2216858 awarded to N.F.

References

[1] S. A. DiMasi J.A., Feldman L. and W. A., "Trends in risks associated with new drug development: success rates for investigational drugs," *Clin Pharmacol Ther*, vol. 87, pp. 272–277, 2010.

[2] G. H. DiMasi J.A. and H. R.W, "Innovation in the pharmaceutical industry: new estimates of r&d costs," *Health Econ*, vol. 47, pp. 20–33, 2016.

[3] W.L.Jorgensen, "The many roles of computation in drug discovery," *Science*, vol. 303, no. 5665, pp. 1813–1818, 2004.

[4] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of Chemical Information and Modeling*, vol. 50, no. 5, pp. 742–754, 2010. PMID: 20426451.

[5] M. Goethe, J. Gleixner, I. Fita, and J. M. Rubi, "Prediction of protein configurational entropy (popcoen)," *Journal of Chemical Theory and Computation*, vol. 14, no. 3, pp. 1811–1819, 2018. PMID: 29351717.

[6] N. Forouzesh and N. Mishra, "An effective mm/gbsa protocol for absolute binding free energy calculations: a case study on sars-cov-2 spike protein and the human ace2 receptor," *Molecules*, vol. 26, no. 8, p. 2383, 2021.

[7] N. Mishra and N. Forouzesh, "Protein-ligand binding with applications in molecular docking," in *Algorithms and Methods in Structural Bioinformatics*, pp. 1–16, Springer, 2022.

[8] N. Forouzesh, "Binding free energy of the novel coronavirus spike protein and the human ace2 receptor: An mmgb/sa computational study," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 1–7, 2020.

[9] S. Decherchi, M. Masetti, I. Vyalov, and W. Rocchia, "Implicit solvent methods for free energy estimation," *European Journal of Medicinal Chemistry*, vol. 91, pp. 27–42, 2015. Molecular Dynamics: New Advances in Drug Discovery.

[10] N. Forouzesh, S. Izadi, and A. V. Onufriev, "Grid-based surface generalized born model for calculation of electrostatic binding free energies," *Journal of chemical information and modeling*, vol. 57, no. 10, pp. 2505–2513, 2017.

[11] A. Onufriev, "Continuum electrostatics solvent modeling with the generalized born model," *Modeling Solvent Environments: Applications to Simulations of Biomolecules*, pp. 127–165, 2010.

[12] A. V. Onufriev and S. Izadi, "Water models for biomolecular simulations," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 2, p. e1347, 2018.

[13] R. A. Friesner, "Ab initio quantum chemistry: Methodology and applications," *Proceedings of the National Academy of Sciences*, vol. 102, no. 19, pp. 6648–6653, 2005.

[14] A. Kawano and G. G. Maisuradze, "Reducing the computational cost of ab initio potential energy calculations by using a dual-level approach," *Journal of Biological Physics and Chemistry*, pp. 83–90, 2006.

[15] L. Chan, G. M. Morris, and G. R. Hutchison, "Understanding conformational entropy in small molecules," *Journal of Chemical Theory and Computation*, vol. 17, no. 4, pp. 2099–2106, 2021.

[16] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (pgnn): An application in lake temperature modeling," 2017.

[17] S. Cain, A. Risheh, and N. Forouzesh, "Calculation of protein-ligand binding free energy using a physics-guided neural network," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2487–2493, 2021.

[18] S. Cain, A. Risheh, and N. Forouzesh, "A physics-guided neural network for predicting protein-ligand binding free energy: From hostndash;guest systems to the pdbbind database," *Biomolecules*, vol. 12, no. 7, 2022.

[19] H. L. Morgan, "The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service.," *Journal of Chemical Documentation*, vol. 5, no. 2, pp. 107–113, 1965.

[20] Y. G. D. M. Martin Amezcua, Jeffry Setiadi, "An overview of the sampl8 host-guest binding challenge," *ChemRxiv*, 2022.

[21] D. L. Mobley, G. Heinzelmann, N. M. Henriksen, and M. K. Gilson, "Predicting binding free energies: Frontiers and benchmarks (a perpetual review)," 2017.

[22] D. L. Mobley and amezcum1, "samplchallenges/sampl8: Sampl8 cb8 "drugs of abuse" challenge inputs," Sep 2020.

[23] D. Case, H. M. Aktulga, K. Belfon, I. Ben-Shalom, J. Berryman, S. Brozell, D. Cerutti, T. Cheatham, G. A. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Götz, R. Harris, S. Izadi, S. Izmailov, and P. Kollman, "Amber 2022," 04 2022.

[24] V. Alexandrov, U. Lehnert, N. Echols, D. Milburn, D. Engelman, and M. Gerstein, "Normal modes for predicting protein motions: A comprehensive database assessment and associated web tool," *Protein Science*, vol. 14, no. 3, pp. 633–643, 2005.

[25] VeraChem, "Free Energy Calculations by Mining Minima¶." https://www.verachem.com/vm2_free_energy_calculation_by_mining_minima/, 2019.

[26] M. K. G. CHIA-EN CHANG, "Tork: Conformational analysis method for molecules and complexes," *Journal of Computational Chemistry*, vol. 24, no. 16, pp. 1987–1998, 2003.

[27] I. Y. Ben-Shalom, S. Pfeiffer-Marek, K.-H. Baringhaus, and H. Gohlke, "Efficient approximation of ligand rotational and translational entropy changes upon binding for use in mm-pbsa calculations," *Journal of chemical information and modeling*, vol. 57 2, pp. 170–189, 2017.

[28] B. Ramsundar, P. Eastman, P. Walters, V. Pande, K. Leswing, and Z. Wu, *Deep Learning for the Life Sciences*. O'Reilly Media, 2019. https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837.

[29] DeepChem, "Grid Hyperparameter Optimization¶." https://deepchem.readthedocs.io/en/latest/api_reference/hyper.html#grid-hyperparameter-optimization, 2018.

[30] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds.), vol. 28, Curran Associates, Inc., 2015.

[31] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: moving beyond fingerprints," *Journal of Computer-Aided Molecular Design*, vol. 30, pp. 595–608, aug 2016.