TopP–S: Persistent Homology-Based Multi-Task Deep Neural Networks for Simultaneous Predictions of Partition Coefficient and Aqueous Solubility

Kedi Wu [a], Zhixiong Zhao,[b] Renxiao Wang,[c] and Guo-Wei Wei *[a,d,e]

Aqueous solubility and partition coefficient are important physical properties of small molecules. Accurate theoretical prediction of aqueous solubility and partition coefficient plays an important role in drug design and discovery. The prediction accuracy depends crucially on molecular descriptors which are typically derived from a theoretical understanding of the chemistry and physics of small molecules. This work introduces an algebraic topology-based method, called element-specific persistent homology (ESPH), as a new representation of small molecules that is entirely different from conventional chemical and/or physical representations. ESPH describes molecular properties in terms of multiscale and multicomponent topological invariants. Such topological representation is systematical, comprehensive, and scalable with respect to molecular size and composition variations. However, it cannot be literally translated into a physical interpretation. Fortunately, it is readily suitable for machine learning methods, rendering topological learning algorithms. Due to the inherent correlation between solubility and partition coefficient, a uniform ESPH representation is developed for both properties, which facilitates multi-task deep neural networks for their simultaneous predictions. This strategy leads to a more accurate prediction of relatively small datasets. A total of six datasets is considered in this work to validate the proposed topological and multitask deep learning approaches. It is demonstrated that the proposed approaches achieve some of the most accurate predictions of aqueous solubility and partition coefficient. Our software is available online at http://weilab.math.msu.edu/TopP-S/. © 2018 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.25213

Introduction

The partition coefficient, denoted P and defined to be the ratio of concentrations of a solute in a mixture of two immiscible solvents at equilibrium, is of great importance in pharmacology. It measures the drug-likeness of a compound as well as its hydrophobic effect on human body. The logarithm of this coefficient, i.e., $\log P$, has proved to be one of the key parameters in drug design and discovery. Optimal $\log P$ along with low molecular weight and low polar surface area plays an important role in governing kinetic and dynamic aspects of drug action. In particular, Hansch and coworkers[1] gave a detailed description of how lipophilicity impacted pharmacodynamics. This being said, surveys show that approximately half of the drug candidates fail to reach the market due to unsatisfactory pharmacokinetic properties or toxicity,[2] which indeed makes $\log P$ predictions even more important.

The extent of existing reliable experimental $\log P$ data is negligible compared to tremendous compounds whose $\log P$ data are practically needed. Therefore, computational prediction of partition coefficient is an indispensable approach in modern drug design and discovery. Since the pioneering work of Hansch and coworkers,[3–5] a large variety of octanol–water partition coefficient predictors has been developed over the past few decades.[6] Many methods are generally called as quantitative structure–activity relationship (QSAR) models. In general, these models can be categorized into atomic-based additive methods, fragment/compound-based methods, and property-based methods. One of the atomic-based additive methods, which was first proposed by Ghose and Crippen[7] is essentially purely additive and effectively a table look-up per atom. Later on, XLOGP3, a refined version of the atomic-based additive methods, was developed.[6] This approach considers various atom types, contributions from neighbors, as well as correction factors which help overcome known difficulties in purely atomistic additive methods. However, additivity may fail in some cases, where unexpected contributions to $\log P$ occur, especially for complicated structures. Fragment/compound-based predictors, instead of employing information from a

[a] K. Wu, Guo-Wei Wei
Department of Mathematics, Michigan State University, East Lansing, Michigan, 48824
E-mail: wei@math.msu.edu
[b] Z. Zhao
School of Medicine, Foshan University, Foshan, Guangdong, 528000, People's Republic of China
[c] R. Wang
State Key Laboratory of Bioorganic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, 200032, People's Republic of China
[d] Guo-Wei Wei
Department of Electrical and Computer Engineering, Michigan State University, Michigan, 48824
[e] Guo-Wei Wei
Department of Biochemistry and Molecular Biology, Michigan State University, Michigan, 48824
Contract grant sponsor: NSF; Contract grant numbers: DMS-1721024 and IIS-1302285; Contract grant sponsor: MSU Center for Mathematical Molecular Biosciences Initiative

single atom, are built at compounds or fragments level. Compounds or fragments are then added up with correction factors. Popular fragment methods include KOWWIN,[8,9] CLOGP,[10,11] ACD/LOGP,[12,13] KLOGP.[14,15] A major challenge for fragment/compound-based methods is the optimal classification of "building blocks." The number of fragments and corrections involved in current methods ranges from hundreds to thousands, which could be even larger if remote atoms are also taken into account. This fact may lead to technical problems in practice and may also cause overfitting in modeling. The third category is property-based. Basically property-based methods determine partition coefficient using properties, empirical approaches, three dimensional (3D) structures (e.g., implicit solvent models, molecule dynamics (MD) methods), and topological or electrostatic indices. Most of these methods are modeled using statistical tools such as associative neural networks (ALOGPS).[16,17] It is worthy to mention that property-based methods are relatively computationally expensive, and depend largely on the choice of descriptors and accuracy of computations. This to some extent results in a preference of methods in the first two categories over those in the third.

Another closely related chemical property is aqueous solubility, denoted by S, or its logarithm value $\log S$. In drug discovery and other related pharmaceutical fields, it is of great significance to identify molecules with undesirable water solubility on early stages as solubility affects absorption, distribution, metabolism, and elimination processes (ADME).[18,19] QSPR models, along with atom/group additive models,[20–23] have been developed to predict solubility. For example, QSPR models assume that aqueous solubility correlates with experimental properties such as aforementioned partition coefficient and melting point,[24] or molecular descriptors such as solvent accessible area. However, due to the difficulty of experimentally measuring solubility for certain compounds, the experimental data can contain errors up to 1.5 log units[25,26] and no less than 0.6 log units.[27] Such a high variability brings a challenge to solubility prediction.

Both partition coefficient and aqueous solubility reveal how a solute dissolves in a solvent. Therefore, it is reasonable to assume that there exists a shared feature representation across these two related tasks. In machine learning theory, multitask (MT) learning is designed to take the advantage of shared feature representations of correlated properties. It learns the so-called "inductive bias" from related tasks to improve accuracy using the same representation.[28] In other words, MT learning aims at learning a shared and generalized feature representation from multiple tasks and has brought new insights into the study of bioinformatics. Successful applications include splice-site and MHC-I binding prediction[29] in sequence biology, gene expression analysis, and system biology.[30] MT learning becomes more efficient when it is incorporated into deep learning (DL) strategies. DL has successfully achieved state-of-the-art results in signal and information processing fields, such as speech recognition[31,32] and natural language processing,[33,34] as well as toxicity prediction[35–39] and aqueous solubility prediction.[40]

Geometric descriptors are commonly used in machine learning to represent small molecules. In fact, geometric representation of molecules, particularly macromolecules, often involves too much structural detail and thus may become intractable for large and complex biomolecular datasets. In contrast, topology offers the highest level of abstraction and truly metric free representations of molecules. However, traditional topology incurs too much geometric reduction to be practically useful for molecules. Persistent homology bridges classical geometry and topology, offering a multiscale representation of molecular systems.[41,42] In doing so, it creates a family of topological spaces via a filtration parameter, which leads to a one-dimensional list of topological invariants, i.e., barcodes of Betti numbers. The physical interpretations of Betti-0, Betti-1, and Betti-2 barcodes are isolated components, circles, and cavities, respectively. Persistent homology has been applied to the modeling and prediction of nanoparticles, proteins, and other biomolecules.[43–47] Nonetheless, it was found that primitive persistent homology has very limited predictive power in machine learning-based classification of biomolecules,[48] which motivate us to introduce element-specific persistent homology (ESPH) to retain crucial biological information during the topological simplification of geometric complexity.[49–51] ESPH has found its success in the predictions of protein–ligand binding affinities,[50,51] mutation-induced protein stability changes[49,51] and virtual screening.[52] However, the representational and predictive power of persistent homology and ESPH for the prediction of small molecular properties is essentially unknown. Unlike proteins, small molecules involve a wide range of chemical elements and their properties are very sensitive to their chemical constitutions, symmetry, and stereochemistry. Therefore, it is not clear whether persistent homology and ESPH are suitable descriptors for small molecular properties.

The objective of this work is to explore the representationability and predictive power of ESPH for small molecular properties. To this end, we focus on the analysis and prediction of small molecular solubility and partition coefficient. Due to their relevance to drug design and discovery, relatively large datasets have been collected in the literature for these problems, which give rise to good datasets for validation of topological descriptors. To overcome the difficulty of small available datasets for certain problems, we construct topological learning by integrating ESPH and multitask deep learning for partition coefficient and aqueous solubility predictions. We show that ESPH provides a competitive description of relatively small drug-like molecules. In addition, the inherent correlation between partition coefficient and aqueous solubility makes the multitask strategy a viable approach in joint $\log P$ and $\log S$ predictions.

The rest of this paper is structured as follow. In the "Datasets and Methods" section, we give an introduction to ESPH and the construction of element-specific topological descriptor (ESTD). The underlying motivation for ESTD is discussed and a concrete example is presented in the "Element specific topological descriptors (ESTD)" subsection. In the "Topological learning algorithms" subsection, we provide an

overview of classic ensemble methods, multitask learning (MT) and deep neural network (DNN). The MT–DNN architecture is carefully formulated and illustrated in the "Multi-task learning and deep neural networks" subsection. In the "Results" section, we first give an overview of datasets. The predictions of MT–DNN models, as well as other methods for partition coefficients and aqueous solubility, are presented. Finally, we wrap up the paper with some discussions in the "Discussion" section.

Datasets and Methods

This section is devoted to datasets, topological methods, machine learning algorithms, and evaluation metrics.

An overview of datasets

The primary work of this paper is to explore the proposed topology-based multitask methods for learning partition coefficient and aqueous solubility simultaneously. Datasets can naturally be divided into two parts—one for partition coefficient prediction and the other for aqueous solubility prediction. Note that for partition coefficient prediction there are multiple test sets while the training set remains the same.

Partition coefficient datasets. The training set used for partition coefficient prediction was originally compiled by Cheng et al.[6] and consists of 8199 compounds, which is based on Hansch et al.'s compilation.[53] These compounds are considered to have reliable experimental log *P* values by Hansch (marked with * or ✓). In addition, three sets were chosen as test sets. The first test set, which is completely independent of the training set, contains 406 small-molecule organic drugs approved by the Food and Drug Administration (FDA) of the United States and represents a variety of organic compounds of pharmaceutical interests. This set was also compiled by Cheng et al.[6] The remaining two test sets, Star set and Non-star set, were publicly available and originated from a monograph of Avdeef.[54] Star set comprises 223 compounds that are part of BioByte Star set and have been widely used to develop log *P* prediction method. The Non-star set contains 43 compounds that represent relatively new chemical structures and properties. The compound list and corresponding partition coefficient are available for download at http://ochem.eu/article/17434. We also made an attempt to expand our training set by searching the NIH database as other software packages use a large number of molecules for supervised learning. In this way, more than 3000 additional molecules were added to the training set.

Aqueous solubility datasets. To develop and validate prediction models for aqueous solubility, several well-defined aqueous solubility datasets were used. First, a diverse dataset of 1708 molecules proposed by Wang et al.[23] was used to verify the predictive power of descriptors. Both leave-one-out and 10-fold cross-validation were carried out on this set. Furthermore, we also tested our models on a relatively small set which has independent training and testing sets.[21] As Hou et al.[21] suggested, we also removed some molecules from the training set to ensure that training set and test set have no overlapping molecules.

Table 1. Summary of log*P* and log*S* datasets used.

log*P* data	Number of molecules	log*S* data	Number of molecules
log*P* train set	8199	Wang's 1708 set[23]	1708
FDA test set	406	Hou's training set[21]	1290
Star test set	223	Klopman's test set[21]	21
Nonstar test set	43	Zhu's test set[21]	120

Statistics of datasets. As a reference, a summary of datasets used for the proposed models is given in Table 1.

Element-specific topological descriptors (ESTD)

A brief introduction is given to persistent homology and ESPH, followed by a detailed example to illustrate the persistent homology characterization of small molecules. A refined version of ESPH and corresponding ESTD construction are also discussed.

Persistent homology. Persistent homology is a branch of algebraic topology that defines topological spaces in terms of algebraic structures. It is the main workhorse for topological data analysis, which offers topological simplification of data complexity. Unlike conventional physical or chemical approaches, persistent homology captures the underlying topological connectivity of small molecules directly from atomic coordinates, i.e., point cloud data in $\mathbb{R}^n$. Mathematically, isolated atoms of a molecule are 0-simplices. The connectivity among atoms defines high-dimensional simplexes. For example, linked two atoms (a line segment) give rise to a 1-simplex and mutually linked three atoms in a triangular shape are called a 2-simplex. A mutually linked four-atom tetrahedron is 3-simplex and so on and so forth.

Appropriate collection of simplices forms a simplicial complex, which is a topological space consisting of vertices (points), edges (line segments), triangles, and their high-dimensional counterparts. Simplicial homology can be defined on the basis of simplicial complex and can be used to analyze topological invariants, i.e., Betti numbers. Physically, Betti-0, Betti-1, and Betti-2 describe numbers of independent components, rings and cavities, respectively. However, it is important to note that topological connectivity among a set of atoms in a molecule does not follow their physical relations, i.e., covalent bonds, hydrogen bonds, and van der Waals bonds. Instead, it is defined by a filtration parameter or artificial ball radius (not the atomic radius) for each atom. Therefore, at a given filtration radius, one obtains a set of simplices, and thus Betti numbers for a molecule. In persistent homology, the filtration radius varies continuously from zero to a large number until no meaningful topological structure can be created further.[42,55] Therefore, persistent homology computes topological invariants of a given molecule at different spatial scales which correspond to different geometric shapes (simplices) and thus different topological connectivities. The persistence

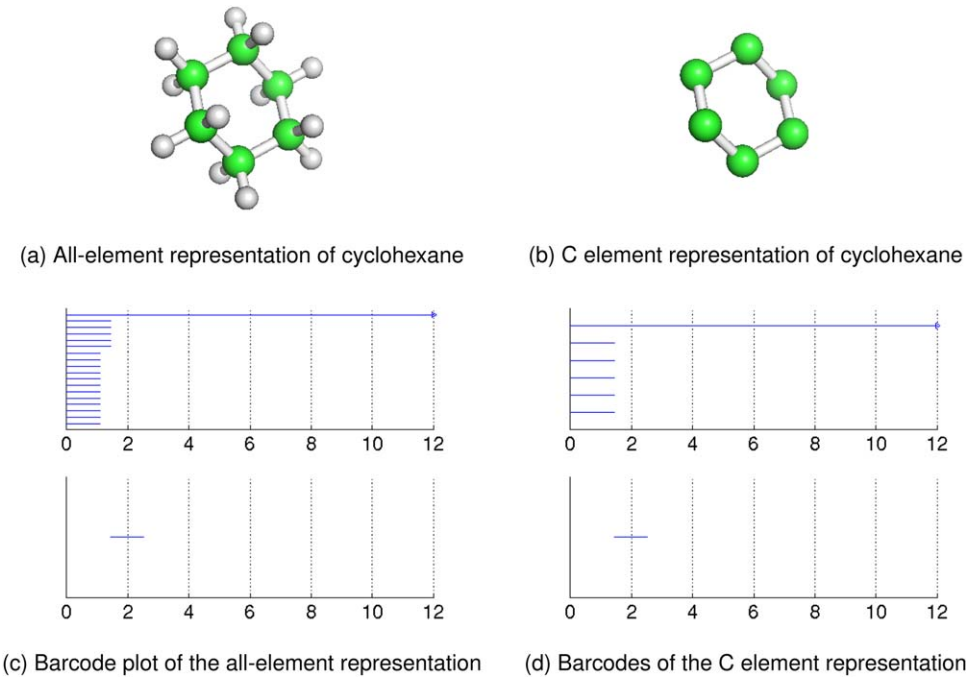

Figure 1. Cyclohexane and its persistent homology barcode plots. In subfigure a) and b), cyclohexane is shown with all elements and carbon element selected, respectively. In subfigure c) and d), from top to bottom, the results are for Betti-0 and Betti-1, respectively. a) All-element representation of cyclohexane. b) C element representation of cyclohexane. c) Barcode plot of the all-element representation. d) Barcodes of the C element representation. [Color figure can be viewed at wileyonlinelibrary.com]

of topological invariants over the filtration for a given molecule can be recorded in barcodes or persistent diagrams.[56,57] The barcode representation of persistent homology is utilized in this work to construct ESTDs. Readers are referred to Ref. 43 for a detailed while the still simple introduction of persistent homology.

The necessity of ESPH and an example. Primitive persistent homology treats all atoms in an equal footing, which neglects chemical and physical properties of molecules. To obtain an accurate representation of a given molecule, it is necessary to at least distinguish different element types and construct element-specific topological descriptors (ESTDs).[49,50] Figure 1 is a detailed example of how our ESTDs are calculated and how they can reveal the structure information of cyclohexane. An all-element representation of cyclohexane is given in Figure 1a, where carbon atoms are in green and hydrogen atoms are in white. As we can see from its barcode plot (Fig. 1c), there are 18 Betti-0 bars that correspond to 18 atoms at the very beginning, 12 of which disappear when the filtration value gets 1.08 Å. It indicates that each carbon atom has merged with its closest 2 hydrogen atoms as the filtration value becomes larger than the length of C¯H bond and these three atoms are regarded as a single connected component. When the filtration value increases to 1.44 Å, a Betti-1 bar emerges which means that a hexagonal carbon ring is captured and there is only one connected component left. As the filtration value eventually exceeds the radius of the hexagon, the ring structure disappears and that is why there is no Betti-1 bar. The longest Betti-0 bar corresponds to the existence of the connected component. When only carbon atoms are selected (C element), it is relatively straightforward to interpret the barcode plot. The cutoff where 5 Betti-0 bars disappear corresponds to the C¯C bond length and the Betti-1 bar represents the existence of the hexagonal carbon ring.

The challenge for primitive persistent homology. Aforementioned persistent homology, although is able to capture the information such as covalent bonds between different atom types easily as shown in Figure 1, does not necessarily reflect intramolecular interactions such as hydrogen bonds and van der Waals interaction, which is not ideal for the purpose of small molecule modeling. In other words, the Betti-0 bar between two atoms with certain hydrogen bonding or van der Waals cannot be captured since there already exist shorter Betti-0 bars between them (essentially covalent bonds). Thus, it is important to redefine the distance between atom i at (x_i, y_i, z_i) and atom j at (x_j, y_j, z_j) as following:[52]

$$M_{i,j} = \begin{cases} d_{i,j}, & \text{if } d_{i,j} \geq r_i + r_j + \sigma \\ d_\infty, & \text{otherwise} \end{cases} \quad (1)$$

where r_i and r_j are the atomic radius of atoms i and j, respectively, and σ is the bond length deviation in the dataset. Here, d_∞ is a large number which is set to be greater than the maximal filtration value, and $d_{i,j}$ is the Euclidean distance between atom i and atom j, or equivalently,

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \quad (2)$$

By setting the distance between two close atoms to a sufficiently large number, we should be able to capture

intramolecular interactions since a connection longer than filtration value is automatically neglected during persistent homology computation.

ESTD construction. Inspired by classic atom-additive models for partition coefficient prediction, we utilize a total of 61 basic element types calculated by antechamber[58,59] using general amber force field (GAFF).[58] Atoms of given atom type and their appropriate combinations are selected to construct Vietoris–Rips complex and ESTDs are subsequently calculated.

It is also important to construct ESTDs via a small bin size. As the example above shows, barcodes at different cutoffs give rise to a variety of information such as covalent and non-covalent bonds. Specifically, we can divide the barcodes into several small bins and then extract features from each bin. A complete list of ESTDs used in this study is shown in Ref. 2. Group 1 ESTDs focuses on different atom types, group 2 is to capture the occurrences of non-covalent bonding and group 3 mainly highlights the strength of non-covalent bonding and van der Waals interactions. Note that statistics of birth or death values in group 3 refer to maximum, minimum, mean, and summation. With such construction, we have 61 ESTDs from group 1, 60 ESTDs from group 2, and 24 ESTDs from group 3. Specifically, for group 2, we have 3 two-element combinations {(C,O), (C,N), (N,O)}, with each combination having 2 ESTDs (birth and death values) for 10 small bins, which gives rise to 60 (3*2*10) ESTDs. Similarly, for group 3, we have also 3 two-element combinations with each combination having 4 statistical values for the collection of birth or death values, resulting in 24 (3*4*2) ESTDs.

The essence of ESTDs is to offer new insight to small molecule modeling by topological modeling and topological learning. By constructing a topological feature vector $\mathbf{x}_i^t$ for the ith molecule of task t, we are readily to combine topological learning with advanced machine learning algorithms, with further details to be discussed in the "Topological learning algorithms" section.

Additional molecular descriptors

In addition to ESTDs introduced in the previous subsection, we generate 633 2D molecule descriptors by ChemoPy[60] for each molecule. The feature pool contains feature groups such as E-state descriptors. There are a total of 13 categories for these 2D molecule descriptors—30 molecular constitutional descriptors, 35 topological descriptors, 44 molecular connectivity indices, 7 Kappa shape descriptors, 64 Burden descriptors, 245 E-state indices, 21 Basak information indices, 96 autocorrelation descriptors, 6 molecular property descriptors, 25 charge descriptors, and 60 MOE-type descriptors. A more detailed description of features and the ChemoPy software is available on line at https://code.google.com/archive/p/pychem/downloads.

To improve the overall performance, we also combine these features with ESTDs to create ESTD$^+$. For consistency reasons, only molecules whose features can be calculated by both our ESTD software and ChemoPy software are used for training purpose. It is worthy to mention that our ESTD approaches are applicable to all molecules whereas ChemoPy has difficulty in dealing with some molecules. Separate results and discussions for different sets of descriptors will also be conducted in later sections.

Topological learning algorithms

In this section, we present methods and algorithms of the topology-based multitask learning for simultaneous predictions of partition coefficient and aqueous solubility. Ensemble methods, including random forest (RF) and gradient boosting decision tree (GBDT), and neural network architectures are discussed. A detailed description of multitask neural networks is also provided.

Ensemble methods. Ensemble methods have been widely used to solve QSAR problems and have achieved some state-of-the-art results. They naturally handle correlations between descriptors and are generally insensitive to parametrization and feature selection due to bagging operations. In this work, we choose gradient boosting tree as our baseline method and implement it using the scikit-learn package[61] (version 0.13.1). The number of estimators for the random forest is set to 8000 because a further increase in the number does not essentially improve accuracy. For each set, 50 runs were done and the average of 50 predictions is taken as the final prediction.

Multitask learning and deep neural networks. *Multitask learning.*

The idea of multitask learning (MTL) is to learn "inductive bias" from related tasks to improve accuracy, using the same representation. In other words, MTL aims at learning a shared, generalized feature representation for multiple tasks and potentially gives better predictions. In our work, we assume that the underlying molecular mechanism of partition coefficient and aqueous solubility shares commonalities and differences, which can be learned jointly.

Multitask deep neural network (MT–DNN). A deep neural network has a wider and deeper architecture as compared to the traditional artificial neural network—it consists of more layers and more neurons in each layer and reveals the facets of input features at different levels. As for this multitask framework, different tasks share the first few dense layers, on top of which individual predictor is attached to for each specific task (one for partition coefficient and one for aqueous solubility). An illustration of the multitask deep neural network is shown in Figure 2.

In this study, we have a total of 2 tasks—one for log$P$ prediction and the other for log$S$ prediction. Suppose that there are N_t molecules in tth task and the ith molecule for tth task can be represented by a topological feature vector $\mathbf{x}_i^t$. Given the training data $(\mathbf{x}_i^t, y_i^t)_{i=1}^{N_t}$, where $t=1,2$, $i=1,\ldots,N_t$, with $\mathbf{y}_i^t$ being experimental value (log$P$ or log$S$) for the ith molecule of task t, the objective topological learning is to minimize the function for different tasks:

$$\sum_{i=1}^{N_t} L(y_i^t, f^t(\mathbf{x}_i^t; \{\mathbf{W}^t, \mathbf{b}^t\})) \tag{3}$$

where f^t is a functional of topological feature vectors to be learned, parametrized by weight matrix $\mathbf{W}^t$ and bias term $\mathbf{b}^t$,

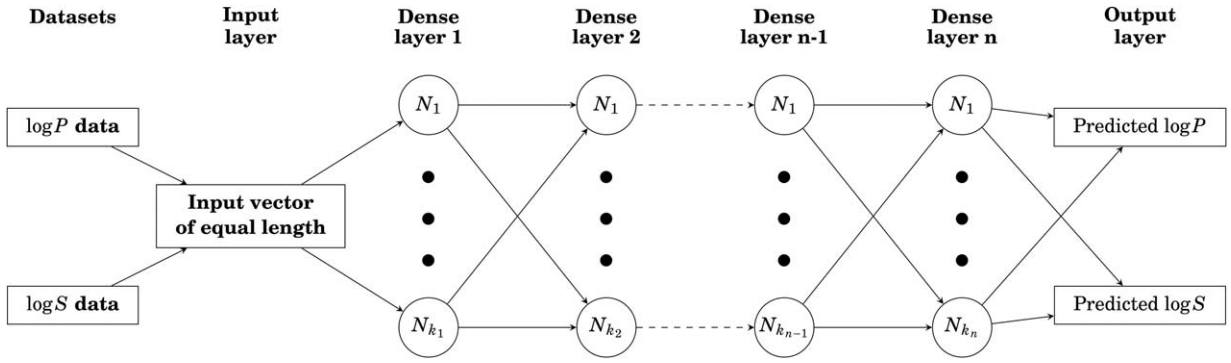

Figure 2. An illustration of the MT–DNN architecture.

and L can be cross entropy loss for classification and mean squared error for regression. Since log$P$ and log$S$ are measured quantitatively, the loss function ($t = 1$ or 2) to be minimized is defined as:

$$\text{Loss of Task } t = \frac{1}{2}\sum_{i=1}^{N_t}(y_i^t - f(\mathbf{x}_i^t; \{\mathbf{W}^t, \mathbf{b}^t\})^2 \quad (4)$$

A wide range of parameters for MT–DNN has been tested and following parameters in Table 2 are used for training and testing in this study although our parameter search is by no means exhaustive.

Note that Adam (adaptive momentum estimation)[62] is used as gradient update method. For stability purpose during training, all data were normalized with zero-mean and unit variance.

Remark. Although the units for these two tasks are different (partition coefficient uses log unit as units, and solubility prediction use $\log_{10}$mol/L as units), we can still train an MT–DNN models for them simultaneously-shared layers learn their inherent commonalities while individual layers deal with their differences.

Evaluation metrics

Several different evaluation metrics, including root mean squared error (RMSE), Pearson correlation coefficient (R), and mean unsigned error (MUE), were used to evaluate the performances of different models. These methods are defined as below:

$$R = \frac{\sum_{i=1}^{N}(\log X_i^{\text{Pred}} - \overline{\log X^{\text{Pred}}})(\log X_i^{\text{Expl}} - \overline{\log X_i^{\text{Expl}}})}{\sqrt{\sum_{i=1}^{N}(\log X_i^{\text{Pred}} - \overline{\log X^{\text{Pred}}})^2}\sqrt{\sum_{i=1}^{N}(\log X_i^{\text{Expl}} - \overline{\log X^{\text{Expl}}})^2}}, \quad (5)$$

Table 2. Parameters for MT–DNN.

# of hidden layers	7
# of neurons on each dense layer	1000 for first 4 layers and 100 for the next 3 layers
# of neurons on output layer	2 with linear activation
Learning rate	0.0001

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\log X_i^{\text{Pred}} - \log X_i^{\text{Expl}}\right)^2} \quad (6)$$

and

$$\text{MUE} = \frac{1}{N}\sum_{i=1}^{N}|\log X_i^{\text{Pred}} - \log X_i^{\text{Expl}}|, \quad (7)$$

where log X represents log P or log S, N is the total number of molecules in the test set, $\log X_i^{\text{Expl}}$ and $\log X_i^{\text{Pred}}$ stand for the experimental and predicted value for the ith molecule, respectively, $\overline{\log X^{\text{Pred}}}$ and $\overline{\log X^{\text{Expl}}}$ is the average of predicted and experimental value for the entire test set, respectively.

In addition, Tetko and coworkers[63] proposed and an additional metric based on the difference between experimental and predicted log P ($\Delta\log P$) was proposed. The percentage within each error range was considered.

- If $|\Delta\log P| < 0.5$, prediction is considered to be "acceptable;"
- If $0.5 \le |\Delta\log P| < 1.0$, prediction is considered to be "disputable;"
- If $|\Delta\log P| \ge 1.0$, prediction is considered to be "unacceptable."

As a result, we use this metric along with R^2 to evaluate model performances on Star set and Non-star set described below.

Results

In this section, we present the results of the proposed ESPH methods in conjugation with random forest and multitask deep neural networks for a variety of datasets, including partition coefficient and solubility test sets. Otherwise stated, different tasks are trained together on the same network. Besides, we would like to introduce some notations for easier reference. ESTD-1 contains 61 Betti-0 bar-based ESTDs (group 1 in Table 3), ESTD-2 contains all ESTDs listed in Table 3 (groups 1–3) (a total of 145 ESTDs). When molecular descriptors calculated by ChemoPy software[60] are combined with ESTD-1 and ESTD-2 respectively, two more descriptor sets are created and they are denoted as ESTD$^+$-1 and ESTD$^+$-2, respectively.

Table 3. Detailed ESTD descriptions.

Feature group	Element used	Descriptors
Group 1	One element: e where $e \in \{\text{GAFF}_{61}\}$	Counts of Betti-0 bars for each element type with a total of 61 different element types calculated with GAFF[58]
Group 2	Two element types: $\{a_i, b_j\}$, where $a_i, b_j \in$ {C, O, N}, $a_i \neq b_j$ and $i < j$	Counts of Betti-0 bars with birth or death values falling within each bin $B_i = [0.5i - 0.5, 0.5i], i = 1, \ldots, 10$
Group 3	Two element types: $\{a_i, b_j\}$, where $a_i, b_j \in$ {C, O, N}, $a_i \neq b_j$ and $i < j$	Statistics of birth or death values for all Betti-0 bars (consider all birth and death values)

Partition coefficient prediction

Training set cross-validation. To have an idea of how our topological representation would work for partition coefficient, a 10-fold cross-validation is performed using baseline method GBDT. Note that 50 runs were done to achieve the final results as randomness is involved and the results are summarized in Table 4. Notice that ChemoPy descriptors were not used here. It can be seen that our descriptors perform better than XLOGP3 software[6] given the same training data, and thus demonstrates great predictive power. In addition, we also provide the fluctuation of 10-fold cross-validation of Table 5, even though such statistics is not available for XLOGP3 software. It is clear that our ESTDs give quite consistent predictions and their performances are independent of random fold generations since RMSDs of R^2, RMSE and MUE do not essentially fluctuate.

Thus, it would be very interesting to see the performances of our MT–DNN compared to XLOGP3 and GBDT.

FDA set. The first test set that we would like to apply our model to is the FDA test set. A molecule that contains Hg was dropped due to the difficulty of computation. A major challenge of this set is that its structures are more complex than that of the training set, and the partition coefficient range spans over nearly 12 units. A series of prediction methods,[6] including our multitask neural networks, are applied to this set and their results are summarized in Table 6 for a comparison with ours.

As we can see from Table 6, our multitask model gives the best prediction in terms of R^2, RMSE, and MUE. Specifically, the small MUE of our model indicates that our predictions are less biased than other methods tested, except for some outliers. Also, note that the training set is completely independent of the test set which shows the applicability of our multitask architecture. We also build models with the same architecture when additional molecules gathered from NIH-database are included as there is no guarantee that ALOGPS are completely independent of the test set. It turns out that the accuracy can be greatly improved. For instance, the performance of ESTD$^+$-1 can be improved by more than 10% in terms of RMSE (0.60 log units to 0.53 log units). It demonstrates the potential of our MT–DNN architecture when more data become available and it will be more carefully discussed in later section.

Star set and Non-star set. Star set and Non-star set were proposed by Tetko and coworkers[63] as two benchmark sets for evaluating partition coefficient models. Over 20 different models were tested on these two sets. It should be emphasized that for these sets, different models are trained on different training sets and their overlap with the test sets is unknown. Thus, it makes more sense to merge our 8199 training set with additional molecules in NIH database and see how additional training data can benefit the overall performances. Results of different models on these two sets can be found in Table 7. Notice that models trained with additional data from NIH database are labeled with (-AD).

For Star set, we achieve RMSE of 0.49 log units with other popular commercial software packages such as ACD/log$P$ and CLOGP, in addition to a high acceptable prediction percentage (77%, rank 2). For Non-star set, most methods do not give accurate predictions as the structures in this set are relatively new and complex. Our 51% acceptable rate ranks number 2 among all predictors, though RMSE is relatively high due to a few large outliers. The results are satisfactory, especially when considering commercial software packages generally use a much larger training set than that ours. In general, when there exist more overlapped molecules in the training set, the test results will be significantly improved. Thus as a baseline comparison, it would be more meaningful if we compare our results with XLogP3 software. As Table 7 indicates, our MT-ESTD$^+$ models achieve a substantial improvement over XLOP3 for Star set, while XLogP3 achieves a lower RMSE for Non-star set. It may be due to XLogP3's corrections terms with relatively new structures. The performances of our MT-ESTD$^+$ models suggest that our models are able to predict log P accurately. We also would like to know the predictive power can potentially be further improved once more molecules are incorporated into the training set. Thus, we extend our original 8199 training set by adding molecules in both Star set and Non-star

Table 4. Results of 10-fold cross validation on the partition coefficient training set, $N = 8199$.

Method	R^2	RMSE	MUE
GBDT-ESTD-2	0.923	0.45	0.32
GBDT-ESTD-1	0.912	0.48	0.35
XLOGP3-AA[6]	0.904	0.50	0.39

Table 5. Fluctuations of fifty 10-fold cross-validation tests.

Method	Mean R^2 (RMSD)	Mean RMSE (RMSD)	Mean MUE (RMSD)
GBDT-ESTD-2	0.923 (0.001)	0.45 (0.003)	0.32 (0.002)
GBDT-ESTD-1	0.912 (0.001)	0.48 (0.002)	0.35 (0.001)

Table 6. Results of different log*P* prediction methods on 406 FDA-approved drugs,[6] ranked by R^2. Two molecules were dropped for our model evaluation due to feature generation failure of ChemoPy.

Method	R^2	RMSE	MUE
GBDT-ESTD$^+$-2-AD	0.935	0.51	0.24
GBDT-ESTD$^+$-1-AD	0.932	0.52	0.23
MT-ESTD$^+$-1-AD	0.930	0.53	0.22
MT-ESTD-1-AD	0.929	0.54	0.26
MT-ESTD$^+$-2-AD	0.928	0.53	0.27
MT-ESTD-1	0.920	0.57	0.28
MT-ESTD-2-AD	0.912	0.59	0.37
GBDT-ESTD$^+$-1	0.910	0.60	0.28
GBDT-ESTD$^+$-2	0.910	0.60	0.30
MT-ESTD$^+$-1	0.909	0.60	0.27
MT-ESTD$^+$-2	0.909	0.60	0.34
ALOGPS	0.908	0.60	0.42
GBDT-ESTD$^+$-1	0.900	0.63	0.39
GBDT-ESTD-1	0.893	0.66	0.41
MT-ESTD-2	0.891	0.66	0.44
GBDT-ESTD$^+$-2	0.883	0.68	0.49
XLOGP3	0.872	0.72	0.51
GBDT-ESTD-2	0.848	0.78	0.57
XLOGP3-AA	0.847	0.80	0.57
CLOGP	0.838	0.88	0.51
TOPKAT	0.815	0.88	0.56
ALOGP98	0.80	0.90	0.64
KowWIN	0.771	1.10	0.63
HINT	0.491	1.93	1.30

set. As shown in Table 7, descriptor sets labeled with (-AD) (with additional data from NIH database) generally offer better performances on both Star set and Non-star set.

Aqueous solubility prediction

To evaluate the performances of solubility our models, several datasets are used, derived from Wang et al.[23] and Hou et al.[21] For leave-one-out validation, only the baseline method is used. For 10-fold cross-validation, the 9 remaining folds are trained together with the partition coefficient training set when evaluating the remaining fold with MT–DNN architecture.

Wang's 1708 set in ref. 23. For this dataset, both leave-one-out and 10-fold cross-validations are carried out to evaluate the performance of our models.

Leave-one-out. As MT–DNN requires a lot of computational resources, only baseline method GBDT is used for leave-one-out prediction. We use 4000 trees and 0.10 learning rate as training parameters to develop models and following results in Table 8 are achieved. Based on the results, our models are able to improve the R^2 and RMSE of AMS-LOGP by 3.8% and 18.2%, respectively.

10-Fold cross-validation. As MT–DNN and baseline method GBDT involves randomness, we run MT–DNN and GBDT 50 times and report mean performances for all metrics. The results are summarized in Table 9. It is observed that our models yield more accurate and robust predictions than ASMS and ASMS-LOGP models do, improving the R^2 from 0.884 to 0.925 (a 4.6% increase) and the RMSE from 0.699 to 0.568 (an 18.6% improvement). In addition, we also notice that there generally exists an improvement of MT-ESTD models over GBDT models, though, not as significant as what we see in the previous partition coefficient prediction.

Dataset in ref. 21. We test our models on dataset proposed by Hou et al.,[21] where training and test sets were predefined to cover a variety of molecules. Klopman's test set contains 21 commonly used compounds of pharmaceutical and environmental interest[64] and is to be trained on the original 1290 molecules. Zhu's test set contains 120 molecules that were

Table 7. Benchmark test results[63] of different log*P* prediction methods on both Star set molecules and Non-star set molecules.

	Star set ($N = 223$)				Non-star set ($N = 43$)			
		% of Molecules within error range				% of Molecules within error range		
Method	RMSE	<0.5	<1	>1	RMSE	<0.5	<1	>1
AB/Log*P*	0.41	84	12	4	1.00	42	23	35
S+log*P*	0.45	76	22	3	0.87	40	35	26
MT-ESTD$^+$-1-AD	0.49	77	16	7	0.98	49	19	33
MT-ESTD$^+$-2	0.49	74	21	5	0.97	49	23	28
MT-ESTD$^+$-2-AD	0.50	76	17	7	0.94	51	19	30
ACD/log*P*	0.50	75	17	7	1.00	44	32	23
GBDT-ESTD$^+$-1-AD	0.51	76	17	6	1.03	44	30	25
GBDT-ESTD$^+$-2-AD	0.51	75	17	7	1.04	41	30	27
CLOGP	0.52	74	20	6	0.91	47	28	26
VLOGP OPS	0.52	64	21	7	1.07	33	28	26
ALOGPS	0.53	71	23	6	0.82	42	30	28
MT-ESTD$^+$-1	0.53	75	17	8	0.97	47	28	26
MT-ESTD-1-AD	0.53	73	18	9	1.00	37	30	33
MT-ESTD-2-AD	0.53	71	19	9	1.01	47	19	35
MT-ESTD-1	0.55	72	18	10	1.01	33	28	40
MT-ESTD-2	0.56	66	23	11	1.06	35	33	33
MiLog*P*	0.57	69	22	9	0.86	49	30	21
GBDT-ESTD$^+$-2	0.58	75	16	8	1.06	44	25	30
GBDT-ESTD$^+$-1	0.60	74	15	9	1.02	46	23	30
XLOGP3	0.62	60	30	10	0.89	47	23	30
KowWIN	0.64	68	21	11	1.05	40	30	30
GBDT-ESTD-2-AD	0.65	62	26	11	1.15	46	16	37
CSLog*P*	0.65	66	22	12	0.93	58	19	23
GBDT-ESTD-1-AD	0.68	71	16	12	1.16	41	11	46
ALOGP	0.69	60	25	16	0.92	28	40	33
MolLog*P*	0.69	61	25	14	0.93	40	25	26
ALOGP98	0.70	61	26	13	1.00	30	37	33
GBDT-ESTD-1	0.71	63	22	13	1.07	34	20	44
OsirisP	0.71	59	26	16	0.94	42	26	33
VLOGP	0.72	65	22	14	1.13	40	28	33
GBDT-ESTD-2	0.73	52	30	17	1.23	44	16	39
TLOGP	0.74	67	16	13	1.12	30	37	30
ABSOLV	0.75	53	30	17	1.02	49	28	23
QikProp	0.77	53	30	17	1.24	40	26	35
QuantlogP	0.80	47	30	22	1.17	35	26	40
SLIPPER-2002	0.80	62	22	15	1.16	35	23	42
COSMOFrag	0.84	48	26	19	1.23	26	40	23
XLOGP2	0.87	57	22	20	1.16	35	23	42
QLOGP	0.96	48	26	25	1.42	21	26	53
VEGA	1.04	47	27	26	1.24	28	30	42
CLIP	1.05	41	25	30	1.54	33	9	49
LSER	1.07	44	26	30	1.26	35	16	49
MLOGP(Sim+)	1.26	38	30	33	1.56	26	28	47
NC+NHET	1.35	29	26	45	1.71	19	16	65
SPARC	1.36	45	22	32	1.70	28	21	49
HINTLOGP	1.80	34	22	44	2.72	30	5	65

Table 8. Leave-one-out test on the 1708 solubility dataset.

Method	R^2	RMSE	MUE
GBDT-ESTD$^+$-1-AD	0.931	0.543	0.389
GBDT-ESTD$^+$-2-AD	0.929	0.551	0.389
GBDT-ESTD$^+$-2	0.910	0.621	0.457
ASMS-LOGP[23]	0.897	0.664	0.505
GBDT-ESTD$^+$-1	0.893	0.683	0.494
ASMS[23]	0.884	0.707	0.547

used to develop Klopman and Zhu's group contribution model.[65] As Hou et al.[21] suggested, we remove 83 molecules that overlap with Zhu's test set from the training set to make prediction independent and unbiased. This reduces the size of the training set for Zhu's test set to 1207.

Klopman's test set. Table 10 shows the performances of different models on Klopman's test set. Our MT-ESTD models perform similarly to Drug-LOGS method while achieving improvement over Klopman and Zhu's MLR method[65] with ESTDs. It is also evident that the MT–DNN method has an edge over GBBT method, which is consistent with our previous experiments.

Zhu's test set. The results of Zhu's test set are summarized in Table 11. For this dataset, our MT-ESTD models give satisfactory results with a high Pearson correlation over 0.97 across all ESTD combinations. Such results indicate that our methods are applicable to a wide variety of molecules. Again, the MT–DNN method outperforms the GBBT method.

Discussion

In this section, we discuss the outcome of trained models from several aspects. The emphasis is put on the predictive power of our ESTDs in terms of MT–DNN and how topology-based multitask learning can improve partition coefficient and aqueous solubility predictions.

ESTDs for small molecules

As the previous results indicated, there exists a common feature representation for both partition coefficient and solubility prediction. Our features come from two different categories—one that is computed solely by ESPH, and the other one that has been widely used in the development of QSAR models. Although the number of ESTDs (121) is small, it turns out that our topological feature representation of molecules has a very strong predictive power as compared to baseline method. Our ESTDs highlight atom type information and keep track of the formation of the chemical bond as well as ring structures of a given molecule. Indeed, by constructing topological space of a molecule, ESTDs are able to extract useful features from ESPH computation. In fact, more detailed topological representations, which more precisely reflect the covalent bond length, hydrogen bond strength, and van der Waals interaction, can be constructed via a small bin size.[51]

Table 9. 10-fold cross-validation on the 1708 solubility dataset.

Method	Mean R^2 (RMSD)	Mean RMSE (RMSD)	Mean MUE (RMSD)
MT-ESTD$^+$-1	0.925 (0.001)	0.568 (0.005)	0.393 (0.003)
MT-ESTD$^+$-2	0.924 (0.003)	0.571 (0.010)	0.395 (0.004)
GBDT-ESTD$^+$-1	0.924 (0.002)	0.572 (0.006)	0.408 (0.005)
GBDT-ESTD$^+$-2	0.923 (0.002)	0.571 (0.006)	0.408 (0.005)
MT-ESTD-1	0.908 (0.002)	0.630 (0.005)	0.466 (0.003)
GBDT-ESTD-2	0.904 (0.002)	0.642 (0.008)	0.469 (0.005)
MT-ESTD-2	0.902 (0.002)	0.649 (0.007)	0.466 (0.005)
GBDT-ESTD-1	0.889 (0.003)	0.697 (0.009)	0.502 (0.005)
ASMS[23]	0.884 (0.021)	0.699 (0.054)	0.527 (0.034)
ASMS-LOGP[23]	0.869 (0.022)	0.742 (0.053)	0.570 (0.034)

Table 10. Results of Klopman's test set,[21] where MUE was not reported.

Method	R	RMSE
MT-ESTD$^+$-1	0.94	0.69
Drug-LOGS[21]	0.94	0.64
GBDT$^+$-2	0.94	0.71
MT-ESTD$^+$-2	0.93	0.75
GBDT$^+$-1	0.93	0.76
MT-ESTD-2	0.92	0.79
Klopman MLR[65]	0.92	0.86
GBDT-2	0.92	0.85
MT-ESTD-1	0.91	0.82
GBDT-1	0.84	1.07

Multitask learning

The goal of multitask learning is to learn commonalities between different tasks, and to simultaneously improve model performances. In this study, partition coefficient and aqueous solubility were trained jointly and substantial improvements over single-task models are observed. Our results suggest that there exists shared information across these two tasks that can benefit prediction accuracy. Indeed, the original motivation for predicting log$P$ and log$S$ is that both coefficients closely relate to the extent to which a compound dissolves in a solvent. By comparing our MT–DNN with gradient boosting tree, we find that it is beneficial to learn partition coefficient and aqueous solubility models together. Our MT-ESTD models achieve satisfactory results on various partition coefficient and aqueous solubility datasets, some of which are the state-of-the-art to our knowledge. Our ESTDs alone can give very accurate predictions, bringing us new insights by ESPH computations. In

Table 11. Results of Zhu's test set.

Method	R	RMSE	MUE
MT-ESTD$^+$-1	0.97	0.65	0.47
MT-ESTD$^+$-2	0.97	0.67	0.48
MT-ESTD-1	0.97	0.70	0.50
MT-ESTD-2	0.97	0.71	0.53
GBDT$^+$-2	0.97	0.73	0.50
GBDT$^+$-1	0.96	0.76	0.52
Drug-LOGS[21]	0.96	0.79	0.57
GBDT-2	0.96	0.79	0.60
GBDT-1	0.96	0.82	0.58
Group contribution[65]	0.96	0.84	0.70

addition to ESTDs, commonly used 2D descriptors also help to improve the overall accuracy. The attempt to learn these two related measurements together gives a boost over learning them separately and the improvement is validated on most datasets.

Predictive power for log*P* and log*S*

We have shown that a common set of ESTDs can be used to accurately predict log*P* and log*S*. However, we also notice that the performances of ESTDs on log*P* are generally better than those of log*S*. One major reason for such result is that the size of log*S* training set is small comparing to that of log*P* training set, and it is difficult to fulfill the potential of MT–DNN algorithms. Also, the (descriptor)/(training sample size) ration of log*S* is much lower than that of log*P*, and MT–DNN and GBDT models are likely to overfit due to the large number of fitting parameters. However, MT–DNN is still able to take advantage of log*P* prediction tasks—for Klopman's and Hou's test sets, MT–DNN achieves better results than GBDT. We believe that log*P* and log*S* can be predicted simultaneously using MT–DNN architectures and it could be beneficial for both prediction tasks.

Conclusion

Partition coefficient and aqueous solubility are among most important physical properties of small molecules and have significant applications to drug design and discovery in terms of lipophilic efficiency. Based on chemical and physical models, a wide variety of computational methods has been developed in the literature for the theoretical predictions of partition coefficient and aqueous solubility.

This work introduces an algebraic topology-based method, ESPH, for simultaneous partition coefficient and aqueous solubility predictions. ESPH offers an unconventional representation of small molecules in terms of multiscale and multicomponent topological invariants. Here, the multiscale representation is inherited from persistent homology, while the multicomponent formulation is developed to retain essential chemical information during the topological simplification of molecular geometric complexity. Therefore, the present ESPH gives a unique representation of small molecules that cannot be obtained by any other method. Although ESPH representation of molecules cannot be literally translated into a physical interpretation, it systematically and comprehensively encipher chemical and physical information of molecules into scalable topological invariants, and thus is ideally suited for machine learning/deep learning algorithms to decipher such information.

To predict partition coefficient and aqueous solubility, we integrate ESPH with advanced machine learning methods, including gradient boosting tree, random forest, and deep neural networks to construct topological learning strategies. Since partition coefficient and aqueous solubility are highly correlated with each other, we develop a common set of ESPH descriptors, called element-specific topological descriptors (ESTDs), to represent both properties. This approach enables us to perform simultaneous predictions of partition coefficient and aqueous solubility using a topology-based multitask deep learning strategy.

To test the representational of ESPH and the predictive power of the proposed topological multitask deep learning strategy, we consider some commonly used datasets, including two benchmark test sets, for partition coefficient, as well as additional solubility datasets. Extensive cross-validations and benchmark tests indicate that the proposed topological multitask strategy offers some of the most accurate predictions of the partition coefficient and aqueous solubility.

Availability

Our software is available as an online server at http://weilab.math.msu.edu/TopP-S/.

Keywords: persistent homology · partition coefficient · aqueous solubility · multitask learning · deep neural networks · topological learning

How to cite this article: K. Wu, Z. Zhao, R. Wang, G.-W. Wei. *J. Comput. Chem*. **2018**, *39*, 1444–1454. DOI: 10.1002/jcc.25213

[1] C. Hansch, A. Leo, D. Hoekman, Exploring QSAR: Hydrophobic, Electronic, and Steric Constants; American Chemical Society: Washington, DC, **1995**.
[2] H. Van, D. Waterbeemd, E. Gifford, *Nat. Rev. Drug Discov.* **2003**, *2*, 192.
[3] C. Hansch, T. Fujita, *J. Am. Chem. Soc.* **1964**, *86*, 1616.
[4] T. Fujita, J. Iwasa, C. Hansch, *J. Am. Chem. Soc.* **1964**, *86*, 5175.
[5] A. Leo, C. Hansch, D. Elkins, *Chem. Rev.* **1971**, *71*, 525.
[6] T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, L. Lai, *J. Chem. Inf. Model.* **2007**, *47*, 2140.
[7] A. K. Ghose, G. M. Crippen, *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21.
[8] W. M. Meylan, P. H. Howard, *J. Pharm. Sci.* **1995**, *84*, 83.
[9] W. M. Meylan, P. H. Howard, *Perspect. Drug Discov. Des.* **2000**, *19*, 67.
[10] A. J. Leo, *Chem. Rev.* **1993**, *93*, 1281.
[11] J. Chou, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 172.
[12] A. A. Petrauskas, E. A. Kolovanov, *Perspect. Drug Discov. Des.* **2000**, *19*, 99.
[13] M. J. Walker, *QSAR Comb. Sci.* **2004**, *23*, 515.
[14] H. Zhu, A. Sedykh, S. K. Chakravarti, G. Klopman, *Curr. Comput. Aided Drug Des.* **2005**, *1*, 3.
[15] A. Y. Sedykh, G. Klopman, *J. Chem. Inf. Model.* **2006**, *46*, 1598.
[16] I. V. Tetko, V. Y. Tanchuk, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136.
[17] I. V. Tetko, P. Bruneau, *J. Pharm. Sci.* **2004**, *93*, 3103.
[18] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
[19] L. Di, E. H. Kerns, *Drug Discov. Today* **2006**, *11*, 446.
[20] B.-G. Duan, Y. Li, J. Li, T.-J. Cheng, R.-X. Wang, *Acta Phys. Chim. Sin.* **2012**, *28*, 2249.
[21] T. Hou, K. Xia, W. Zhang, X. Xu, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 266.
[22] J. Wang, T. Hou, X. Xu, *J. Chem. Inf. Model.* **2009**, *49*, 571.
[23] J. Wang, G. Krudy, T. Hou, W. Zhang, G. Holland, X. Xu, *J. Chem. Inf. Model.* **2007**, *47*, 1395.
[24] S. H. Yalkowsky, S. C. Valvani, *J. Pharm. Sci.* **1980**, *69*, 912.
[25] J. C. Dearden, *Expert Opin. Drug Discov.* **2006**, *1*, 31.
[26] R. Dannenfelser, M. Paric, M. White, S. H. Yalkowsky, *Chemosphere* **1991**, *23*, 141.
[27] W. L. Jorgensen, E. M. Duffy, *Adv. Drug Deliv. Rev.* **2002**, *54*, 355.
[28] R. Caruana, In *Learning to Learn*; Springer, **1998**; pp. 95–133.

[29] C. Widmer, G. Rätsch, In ICML Unsupervised and Transfer Learning, **2012**; pp. 207–216.
[30] Q. Xu, Q. Yang, *J. Comput. Sci. Eng.* **2011**, *5*, 257.
[31] G. E. Dahl, D. Yu, L. Deng, A. Acero, *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 30.
[32] L. Deng, G. Hinton, B. Kingsbury, In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on IEEE, **2013**; pp. 8599–8603.
[33] R. Socher, Y. Bengio, C. D. Manning, In *Tutorial Abstracts of ACL 2012*; Association for Computational Linguistics, **2012**; pp. 5–5.
[34] I. Sutskever, O. Vinyals, Q. V. Le, In Advances in Neural Information Processing Systems, **2014**; pp. 3104–3112.
[35] G. E. Dahl, N. Jaitly, R. Salakhutdinov, *arXiv preprint arXiv:1406.1231*, **2014**.
[36] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, *J. Chem. Inf. Model.* **2015**, *55*, 263.
[37] A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter, *Front. Environ. Sci.* **2016**, *3*, 80.
[38] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, V. Pande, *arXiv preprint arXiv:1502.02072*, **2015**.
[39] K. Wu, G.-W. Wei, *J. Chem. Inf. Model.* **2018**, *58*, 520.
[40] A. Lusci, G. Pollastri, P. Baldi, *J. Chem. Inf. Model.* **2013**, *53*, 1563.
[41] H. Edelsbrunner, D. Letscher, A. Zomorodian, *Discrete Comput. Geom.* **2002**, *28*, 511.
[42] A. Zomorodian, G. Carlsson, *Discrete Comput. Geom.* **2005**, *33*, 249.
[43] K. L. Xia, G. W. Wei, *Int. J. Numer. Methods Biomed. Eng.* **2014**, *30*, 814.
[44] K. L. Xia, X. Feng, Y. Y. Tong, G. W. Wei, *J. Comput. Chem.* **2015**, *36*, 408.
[45] K. L. Xia, Z. X. Zhao, G. W. Wei, *J. Comput. Biol.* **2015**, *22*, 1.
[46] K. L. Xia, Z. X. Zhao, G. W. Wei, *J. Chem. Phys.* **2015**, *143*, 134103.
[47] B. Wang, G. W. Wei, *J. Comput. Phys.* **2016**, *305*, 276.
[48] Z. X. Cang, L. Mu, K. Wu, K. Opron, K. Xia, G.-W. Wei, *Mol. Based Math. Biol.* **2015**, *3*, 140.
[49] Z. X. Cang, G. W. Wei, *Bioinformatics* **2017**, *33*, 3549.
[50] Z. X. Cang, G. W. Wei, *Int. J. Numer. Methods Biomed. Eng.* **2017**, *34*, e2914.
[51] Z. X. Cang, G. W. Wei, *PLOS Comput. Biol.* **2017**, *13*, e1005690.
[52] Z. X. Cang, L. Mu, G. W. Wei, *PLOS Comput. Biol.* **2018**, *14*, e1005929.
[53] C. Hansch, A. Leo, D. Hoekman, et al., *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, Vol. *557*; American Chemical Society: Washington, DC, **1995**.
[54] A. Avdeef, *Absorption and Drug Development: Solubility, Permeability, and Charge State*; John Wiley & Sons; Hoboken, New Jersey, **2012**.
[55] H. Edelsbrunner, D. Letscher, A. Zomorodian, In Foundations of Computer Science, 2000. Proceedings 41st Annual Symposium on IEEE, **2000**; pp. 454–463.
[56] G. Carlsson, A. Zomorodian, A. Collins, L. J. Guibas, *Int. J. Shape Model.* **2005**, *11*, 149.
[57] R. Ghrist, *Bull. Am. Math. Soc.* **2008**, *45*, 61.
[58] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
[59] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247.
[60] D.-S. Cao, Q.-S. Xu, Q.-N. Hu, Y.-Z. Liang, *Bioinformatics* **2013**, *29*, 1092.
[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
[62] D. Kingma, J. Ba, *arXiv preprint arXiv:1412.6980*, **2014**.
[63] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *J. Pharm. Sci.* **2009**, *98*, 861.
[64] G. Klopman, S. Wang, D. M. Balthasar, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474.
[65] G. Klopman, H. Zhu, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 439.

Received: 6 December 2017
Revised: 15 January 2018
Accepted: 25 February 2018
Published online on 6 April 2018