

AGL-Score: Algebraic Graph Learning Score for Protein–Ligand Binding Scoring, Ranking, Docking, and Screening

Duc Duy Nguyen[†] and Guo-Wei Wei^{*,†,‡,§,¶}

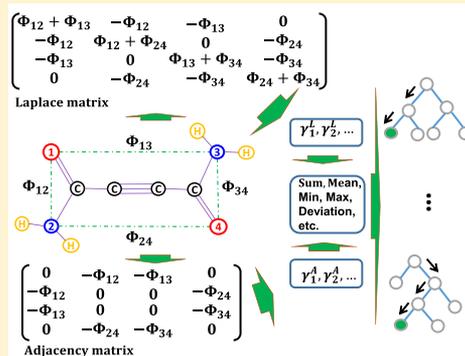
[†]Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States

[‡]Department of Biochemistry and Molecular Biology Michigan State University, East Lansing, Michigan 48824, United States

[§]Department of Electrical and Computer Engineering Michigan State University, East Lansing, Michigan 48824, United States

S Supporting Information

ABSTRACT: Although algebraic graph theory-based models have been widely applied in physical modeling and molecular studies, they are typically incompetent in the analysis and prediction of biomolecular properties, confirming the common belief that “one cannot hear the shape of a drum”. A new development in the century-old question about the spectrum–geometry relationship is provided. Novel algebraic graph learning score (AGL-Score) models are proposed to encode high-dimensional physical and biological information into intrinsically low-dimensional representations. The proposed AGL-Score models employ multiscale weighted colored subgraphs to describe crucial molecular and biomolecular interactions in terms of graph invariants derived from graph Laplacian, its pseudo-inverse, and adjacency matrices. Additionally, AGL-Score models are integrated with an advanced machine learning algorithm to predict biomolecular macroscopic properties from the low-dimensional graph representation of biomolecular structures. The proposed AGL-Score models are extensively validated for their scoring power, ranking power, docking power, and screening power via a number of benchmark datasets, namely CASF-2007, CASF-2013, and CASF-2016. Numerical results indicate that the proposed AGL-Score models are able to outperform other state-of-the-art scoring functions in protein–ligand binding scoring, ranking, docking, and screening. This study indicates that machine learning methods are powerful tools for molecular docking and virtual screening. It also indicates that spectral geometry or spectral graph theory has the ability to infer geometric properties.



1. INTRODUCTION

Graph theory is a prime subject of discrete mathematics that considers graphs as mathematical structures for modeling pairwise relations between vertices, nodes, or points. Such pairwise relations define graph edges. There are many different graph theories, such as geometric graph theory, algebraic graph theory, and topological graph theory. Geometric graphs admit geometric objects as graph nodes or vertices. Algebraic graph theory, particularly spectral graph theory, studies the algebraic connectivity via characteristic polynomial, eigenvalues, and eigenvectors of matrices associated with graphs, such as an adjacency matrix or a Laplacian matrix. Topological graph theory concerns the embeddings and immersions of graphs and the association of graphs with topological spaces, such as abstract simplicial complexes. Mathematically, graphs are useful tools in geometry and certain parts of topology such as knot theory and algebraic topology.

Like topology, graph theory also emphasizes connectivity. The geometric connectivity of a graph refers to pairwise relations among graph nodes and is often analyzed on the basis of the “topological index”,^{1,2} contact map,^{3,4} and graph centrality.^{5–7} The algebraic connectivity of a graph refers to the second-smallest eigenvalue of the Laplacian matrix of the graph and is also known as the Fiedler value or Fiedler

eigenvalue, which has many applications, including the stability analysis of dynamical systems.⁸ In contrast, topological connectivity refers to the connectedness of the entire system rather than pairwise ones as in the geometric graph theory. Topological connectivity is an important property for distinguishing topological spaces.

Over a century ago, Hermann Weyl investigated whether geometric properties of a bounded domain could be determined from the eigenvalues of the Laplace operator on the domain. This question was phrased as “Can one hear the shape of a drum?” by Mark Kac.⁹ An interesting question is: Can eigenvalues describe protein–ligand binding?

Graph theory has been widely applied in physical, chemical, biological, social, linguistic, computer, and information sciences. Many practical problems can be represented and analyzed by graphs. In chemistry and biology, a graph makes a natural model for a molecule, where graph vertices represent atoms and graph edges represent possible bonds. Graphs have been widely used in chemical analysis^{10–12} and biomolecular modeling,¹³ including normal-mode analysis (NMA)^{14–17} and elastic network model (ENM),^{3,18–22} for modeling protein

Received: April 22, 2019

Published: June 18, 2019

flexibility and long-time dynamics. Some of the most popular ENMs are the Gaussian network model (GNM)^{3,19,23} and the anisotropic network model (ANM).²⁰ In these methods, the diagonalization of the interaction Laplacian matrix is a required procedure to analyze protein flexibility, which has the computational complexity of $O(N^3)$, with N being the number of matrix elements. Graph theory has also been used to represent the structures of molecules and biomolecules, resulting in a popular approach for chemical datasets^{2,10,11,24–26} and biomolecular datasets^{3,27–32} in the past few decades.

Graph theories, especially geometric graph theories, are relatively intuitive and easy to use. Indeed, a great portion of previous graph theory-based study in molecular and biomolecular systems was qualitative and descriptive. Despite the intensive effort in the past, graph theory-based quantitative methods are often not as competitive as other quantitative approaches in the analysis and prediction of biomolecular properties from massive and diverse datasets. For example, graph-signature-based prediction of protein stability changes upon mutation³³ was not as accurate as some other methods.^{34–36} Additionally, the average Pearson correlation coefficients in protein *B*-factor predictions using spectral graph theory-based GNM were less than 0.6 in all of three datasets.³⁷ These situations may be attributed to the following reasons. First, most graph theory-based models do not distinguish different chemical element types in a molecule or biomolecule, which leads to a severe loss of critical chemical and biological information. Second, in many molecular graphs, edges are used to represent covalent bonds, while non-covalent interactions are often ignored, which under-represents the physical interactions of many biomolecular datasets. Finally, many graph-based models approximate the distance between a pair of atoms by the number of covalent bonds between them, which leads to a major error in describing their interaction strength.

In the past few years, we have developed a number of graph theory approaches to address the aforementioned problems. For example, weighted graphs were proposed in terms of the flexibility–rigidity index (FRI) to represent graph edges by radial basis functions.^{38–41} Physically, we assume that protein interactions, including those with its environment, fully determine its structure at the equilibrium. Protein structure and its environment, in turn, fully determine protein flexibility and function. As a consequence, one does not need to invoke a high-dimensional model that is subject to modeling errors to analyze protein flexibility and function when the native structure of the protein and its environment are known. Mathematically, our approach assumes a complete graph while it weights the importance of graph edges by scaling their Euclidean distances in radial basis functions so that the nearest neighbors in the sense of the Euclidean metric have the strongest edges. Additionally, multiscale FRI is a multigraph approach which is permitted to have multiple edges.^{40,42} Similar to persistent homology,^{43,44} this multi-edge technique allows a given molecular graph to be analyzed in multiscale, capturing the multiscale interactions in macromolecules.⁴⁰ Graph coloring, or, more generally, graph labeling, is an important graph theory technique allowing graph vertices or edges to be treated differently. This method enables the encoding of chemical and biological information into molecular graphs.^{45,46} Subgraphs constructed from vertex-labeled graphs and edge-labeled graphs give rise to powerful

graph representations of intermolecular and intramolecular interactions, such as hydrogen bonds, electrostatics, van der Waals interactions, hydrophilicity, hydrophobicity, etc.^{45,46} Our multiscale weighted colored graph is over 40% more accurate than GNM in protein *B*-factor predictions.⁴⁶

The importance of protein–ligand binding in living organisms cannot be overstated. A wide variety of biological processes, such as transmitter-mediated signal transduction, hormone- and growth-factor-regulated metabolic pathways, and stimulus-initiated gene expression, enzyme production, and cell secretion, are triggered by ligand–receptor agonist binding. Therefore, the understanding of protein–ligand interactions is a central issue in biochemistry, biophysics, and molecular biology. It is commonly believed that protein–ligand binding involves synergistic protein–ligand corporation, molecular recognition, and conformational changes for both protein and ligand.

Various scoring functions (SFs) have been developed for understanding protein–ligand binding. Among them, physics-based SFs uniquely offer mechanistic understanding and do not depend on existing data. Empirical SFs use physical submodels to fit existing data.^{47–49} Knowledge-based SFs take advantage of available protein–ligand binding datasets and can be used without further training.^{50–52} Finally, machine learning-based SFs are data-driven, and their performance strongly depends on the training set, in addition to their descriptors and machine learning algorithms.^{53–58} On the positive side, machine learning-based SFs can easily handle large and diverse datasets, as well as nonlinear correlations in the data. They often turn out as the winners for many standard benchmarks^{36,45,53,57,58} and community-wide competitions.⁵⁹

Notably, RF-Score was the first machine learning-based SF that impressively outperformed other SFs in 2010.⁵³ Since then, there has been much skepticism about machine learning-based SFs. For example, Gabel et al. have shown that RF-Score is unable to enrich virtual screening hit lists in true actives upon docking experiments of 10 reference DUD-E datasets.⁶¹ This comes as no surprise. All machine learning-based SFs are data-driven methods and do not work without structural and/or sequence similarity in training and prediction datasets, as shown by Li and Yang.⁶¹ It can be hard to decide what training set should be used, while Kramer et al. argued that leave-cluster-out cross-validation is appropriate for SFs derived from diverse protein datasets.⁶² Recently, Wang and Zhang have generated their own training sets (unfortunately, these useful training sets are not publicly available) to show that machine learning models can do very well in docking and screening tests.⁵⁸ It is highly important to design common benchmarks^{63–66} and/or blind grand challenges so that various SFs can be assessed on an equal footing without bias and prejudice.

Recently, we have developed various machine learning-based SFs using one of three types of descriptors, namely physics-based descriptors which consist of electrostatics binding free energies and atomic Coulombic and van der Waals interactions,⁵⁷ geometric graph theory-based descriptors,⁴⁵ and algebraic topology-based descriptors.³⁶ Geometric graph theory and algebraic topology-based predictions of free energies and their rankings (Kendall's tau) were ranked 1st for Set 1 (Stage 2) of D3R Grand Challenge 2 and for 10 of a total of 26 contests in D3R Grand Challenge 3.⁵⁹ Conceptually, our approaches are built upon the fundamental hypothesis that the intrinsic physics of interest lies in low-dimensional subspaces or manifolds embedded in a high-

dimensional data space. While the hypothesis is quite well-known in the manifold learning field, the major challenge is how to encode crucial physical information contained in a high-dimensional space into the desirable low-dimensional representation of molecules, biomolecules, and their complexes. Our multiscale weighted colored subgraphs address this challenge. The major advantages of our multiscale weighted colored subgraph approach are its low-dimensionality, simplicity, and robustness. For example, the only required data inputs for the binding affinity prediction are atomic names and coordinates. Indeed, it bypasses complicated data processing and parametrization. It does not need any molecular mechanical (MM) force fields, namely, charges, polarization assignments, bond lengths and angles, van der Waals well depths, dielectric constants, surface tension, electronegativity, etc. As such, it avoids errors associated with the parametrization. In fact, our geometric graph theory approach is also simpler than our algebraic topology approach mathematically and computationally, while it performs as well as our topological approach.⁶⁷

The objective of the present work is to develop multiscale weighted labeled algebraic subgraphs for representing protein–ligand interactions. For a given protein–ligand complex, there are many ways to construct corresponding algebraic (sub)graphs. The three most commonly used algebraic graphs are the graph Laplacian matrix, its pseudo-inverse, and adjacency matrices. The eigenvalues and eigenvectors computed from these matrices can be used to describe molecules, biomolecules, and their interactions in many different ways. We examine a few common approaches in this work. The resulting method, called algebraic graph learning score (AGL-Score), is critically assessed on its protein–ligand binding scoring power, ranking power, docking power, and screening power with a variety of benchmark datasets.

2. METHODS AND ALGORITHMS

2.1. Multiscale Weighted Labeled Geometric Subgraphs. We propose to develop systematical, scalable, accurate graph theory descriptors of protein–ligand binding interactions from massive and diverse datasets. However, the proposed method can be applied to other problems such as the predictions of toxicity, solubility, solvation, partition coefficient, mutation-induced protein folding stability change, and protein–nucleic acid interactions. In the present work, we target pairwise non-covalent interactions in our subgraph theory description. For a given dataset, we first perform a statistical analysis to identify a set of commonly occurring chemical element types, say $C = \{H, C, N, O, S, P, F, Cl, Br, \dots\}$. For a given molecule or biomolecule in the dataset, let us denote

$$\mathcal{V} = \{(\mathbf{r}_j, \alpha_j) | \mathbf{r}_j \in \mathbb{R}^3; \alpha_j \in C; j = 1, 2, \dots, N\} \quad (1)$$

a subset of N atoms (i.e., subgraph vertices) that are members of C . Note that the i th atom is labeled both by its element type α_i and its position \mathbf{r}_i . The classification of atoms into chemical element types is a graph coloring, which is important for encoding different types of interactions and gives rise to a basis for the collective coarse-grained description of the dataset. We assume that all the pairwise non-covalent interactions between element types C_k and $C_{k'}$ in a molecule or molecular complex can be represented by fast-decay weight functions

$$\begin{aligned} \mathcal{E} &= \{\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}) | \alpha_i = C_k, \alpha_j = C_{k'}; \\ & i, j = 1, 2, \dots, N; \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma\} \end{aligned} \quad (2)$$

where $\|\mathbf{r}_i - \mathbf{r}_j\|$ is the Euclidean distance between the i th and j th atoms, r_i and r_j are the atomic radii of i th and j th atoms, respectively, and σ is the mean value of the standard deviations of r_i and r_k in the dataset. The distance constraint ($\|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma$) excludes covalent interactions. Here $\eta_{kk'}$ is a characteristic distance between the atoms, and Φ is a subgraph weight, chosen to have the following properties:³⁹

$$\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}) = 1 \quad \text{as } \|\mathbf{r}_i - \mathbf{r}_j\| \rightarrow 0 \quad \text{and} \quad (3)$$

$$\begin{aligned} \Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}) &= 0 \quad \text{as } \|\mathbf{r}_i - \mathbf{r}_j\| \rightarrow \infty, \\ \alpha_i &= C_k, \alpha_j = C_{k'} \end{aligned} \quad (4)$$

Although most radial basis functions can be used, generalized exponential functions and generalized Lorentz functions were shown to work very well for biomolecules.³⁹ We, therefore, have a weighted colored subgraph $G(\mathcal{V}, \mathcal{E})$. To construct element-level collective molecular descriptors, we propose the multiscale weighted colored subgraph rigidity between k th element type C_k and k' th element type $C_{k'}$:

$$\begin{aligned} \text{RI}^G(\eta_{kk'}) &= \sum_i \mu_i^G(\eta_{kk'}) = \sum_i \sum_j \Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}), \\ \alpha_i &= C_k, \alpha_j = C_{k'}; \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma \end{aligned} \quad (5)$$

where $\mu_i^G(\eta_{kk'})$ is a geometric subgraph centrality for the i th atom, which offers accurate protein B -factor predictions.⁴⁶ The physical interpretation of eq 5 is straightforward—the summation over $\mu_i^G(\eta_{kk'})$ in eq 5 leads to the total interaction strength for the selected pair of element types C_k and $C_{k'}$, which provides the element-level coarse-grained description of molecular level properties. The above formulation is a generalization of the successful bipartite subgraph used in our earlier predictions of protein–ligand binding affinities and free energy ranking.⁴⁵ For a bipartite subgraph, each of its edges connects one atom in the protein and another atom in the ligand. The graph coloring, i.e., element-specific descriptions, and subgraph weight are designed to capture hydrogen bonds, polarization, electrostatics, van der Waals interactions, hydrophilicity, hydrophobicity, etc.

The different selections of characteristic distance $\eta_{kk'}$ give rise to a multiscale description of intermolecular and intramolecular interactions. By appropriate selections of element combinations k and k' , the characteristic distance $\eta_{kk'}$, and subgraph weight Φ , we systematically construct a family of collective, scalable, multiscale graph-based molecular and biomolecular descriptors. The proposed multiscale weighted colored subgraph rigidity is simple and robust—the only required data input is atomic names and coordinates. Indeed, it bypasses complicated data processing, parametrization, and MM force fields, such as charges, high-order polarizations, van der Waals well depths, dielectric constants, surface tensions, and electronegativity, and their associated errors in many physical models. Consequently, our graph theory approaches are very fast.³⁹ Our fast algorithm has the computational complexity of $O(N)$ and is able to predict B -factors for α -carbons of an HIV virus capsid (313 236 residues) in less than 30 s on a single processor.³⁹

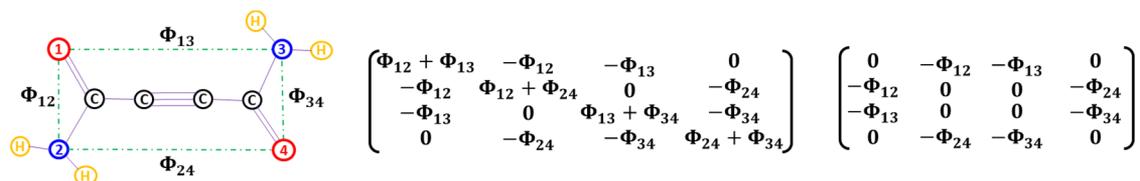


Figure 1. Illustration of weighted colored subgraph G_{NO} (Left), its Laplacian matrix (Middle), and adjacency matrix (Right) for cellocidin molecule ($C_4H_4N_2O_2$). Graph vertices, namely oxygen (i.e., atoms 1 and 4) and nitrogen (i.e., atoms 2 and 3), are labeled in red and blue colors, respectively. Here, graph edges (i.e., Φ_{ij}) are labeled by green-dashed lines which are *not* covalent bonds. Here, Φ_{ij} are distance-weighted edges. Note that there are nine other nontrivial subgraphs for this molecule (i.e., G_{CC} , G_{CN} , G_{CO} , G_{CH} , G_{NN} , G_{NH} , G_{OO} , G_{OH} , and G_{HH}).

2.2. Multiscale Weighted Labeled Algebraic Subgraphs. Our earlier work has demonstrated how to construct powerful geometric graph descriptors for analyzing and predicting biomolecular datasets. Mathematically, it is extremely interesting to understand whether there exist equally powerful algebraic graph or spectral graph formulations. Biologically, it is important to develop alternative graph tools for describing molecules, biomolecules, and their interactions since each method has its own advantages and potentials. For a given subgraph, its matrix representations provide a straightforward description of the interaction between subgraph elements, which can be easily expressed by matrices. Two of the most important matrices are the Laplacian matrix and the adjacency matrix.

2.2.1. Multiscale Weighted Labeled Laplacian Matrix. We consider a subgraph $G_{kk'}$ for each pair of element types, C_k and $C_{k'}$, and propose an element-level weighted labeled Laplacian matrix $L(\eta_{kk'})$ with elements

$$L_{ij}(\eta_{kk'}) = \begin{cases} -\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}) & \text{if } i \neq j, \alpha_i = C_k, \alpha_j = C_{k'} \\ & \text{and } \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma; \\ -\sum_j L_{ij} & \text{if } i = j \end{cases} \quad (6)$$

Mathematically, our element-level weighted labeled Laplacian matrix is symmetric, diagonally dominant and positive-semidefinite, and thus all of its eigenvalues are nonnegative. Since every row sum or column sum of $L(\eta_{kk'})$ is zero, the first eigenvalue value is zero. There can be more than one zero eigenvalue and the number of zero eigenvalues of the Laplacian is the rank of the zero-dimensional topological invariant, reflecting the number of the connected components in the graph. The first non-zero eigenvalue of $L(\eta_{kk'})$ is the so-called algebraic connectivity (or Fiedler value) of $G_{kk'}$. It is interesting to note a certain connection between geometric graph formulation and algebraic graph matrix:

$$RI^G(\eta_{kk'}) = \text{Tr } L(\eta_{kk'})$$

where Tr is the trace. Denote $\lambda_j^L, j = 1, 2, \dots$ and $\mathbf{u}_j^L, j = 1, 2, \dots$ the eigenvalues and eigenvectors of $L(\eta_{kk'})$. We define an atomic descriptor for the i th atom ($\mathbf{r}_i, \alpha_i = C_k$):

$$\mu_i^L(\eta_{kk'}) = \sum_l (\lambda_l^L)^{-1} [\mathbf{u}_l^L(\mathbf{u}_l^L)^T]_{ii} \quad (7)$$

where T is the transpose. We further propose element-level weighted labeled Laplacian matrix-based molecular descriptors,

$$RI^L(\eta_{kk'}) = \sum_i \mu_i^L(\eta_{kk'}) \quad (8)$$

Note that $\mu_i^L(\eta_{kk'})$ is a weight subgraph generalization of GNM⁴ or a subgraph generalization of our earlier generalized multiscale FRL.⁴² Therefore, $\mu_i^L(\eta_{kk'})$ can be used to represent atomic properties, such as protein B-factors. Additionally, we can construct a set of element-level weighted labeled Laplacian matrix-based molecular descriptors by the statistics of $\mu_i^L(\eta_{kk'})$, i.e., sum, mean, maximum, minimum, and standard deviation of $\mu_i^L(\eta_{kk'})$.

Alternatively, we can directly construct another set of element-level weighted labeled Laplacian matrix-based molecular descriptors by the statistics of nontrivial eigenvalues $\{\lambda_j^L\}_{j=2,3,\dots}$. In this case, the Fiedler value is included as the minimum. The performances of these two sets of molecular descriptors constructed from element-level weighted labeled subgraph Laplacian matrix will be examined and compared.

2.2.2. Multiscale Weighted Labeled Adjacency Matrix. The element-level weighted labeled adjacency matrix is equally important and can be easily constructed for subgraph $G_{kk'}$ by

$$A_{ij}(\eta_{kk'}) = \begin{cases} -\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}) & \text{if } i \neq j, \alpha_i = C_k, \alpha_j = C_{k'} \\ & \text{and } \|\mathbf{r}_i - \mathbf{r}_j\| > r_i + r_j + \sigma; \\ 0 & \text{if } i = j \end{cases} \quad (9)$$

Mathematically, adjacency matrix $A(\eta_{kk'})$ is a symmetric non-negative matrix and it contains the same amount of information as the corresponding Laplacian matrix, although its eigenvalues $\lambda_j^A, j = 1, 2, \dots$ and eigenvectors $\mathbf{u}_j^A, j = 1, 2, \dots$ behave very differently from those of corresponding Laplacian matrix. Figure 1 illustrates the Laplacian and adjacency matrices for the weighted colored subgraph G_{NO} in cellocidin molecule ($C_4H_4N_2O_2$). The spectrum of the proposed element-level weighted colored adjacency matrix is real. For each eigenvalue, its opposite is also an eigenvalue. Therefore, only positive eigenvalues will be used in our description. The Perron–Frobenius theorem states that the greatest eigenvalue, i.e., the spectral radius $\rho(A)$, is bounded above by the maximal diagonal element of the corresponding Laplacian matrix $\min_i \sum_j A_{ij} \leq \rho(A) \leq \max_i \sum_j A_{ij}$. The values of Laplacian matrix elements depend on the scale parameter $\eta_{kk'}$ and have many zeros at a characteristic scale parameter for hydrogen bonds or van der Waals interactions. However, the characteristic scale for electrostatic and hydrophobic interactions can be very large.⁶⁸ In such as case, spectral radius $\max_i \sum_j A_{ij} \approx n - 1$, with n being the number of atoms in the subgraph $G_{kk'}$.

Assume that all eigenvalues and eigenvectors of $A_{ij}(\eta_{kk'})$ are given by $\{\lambda_j^A\}$ and $\{\mathbf{u}_j^A\}$, respectively. In the present work, we

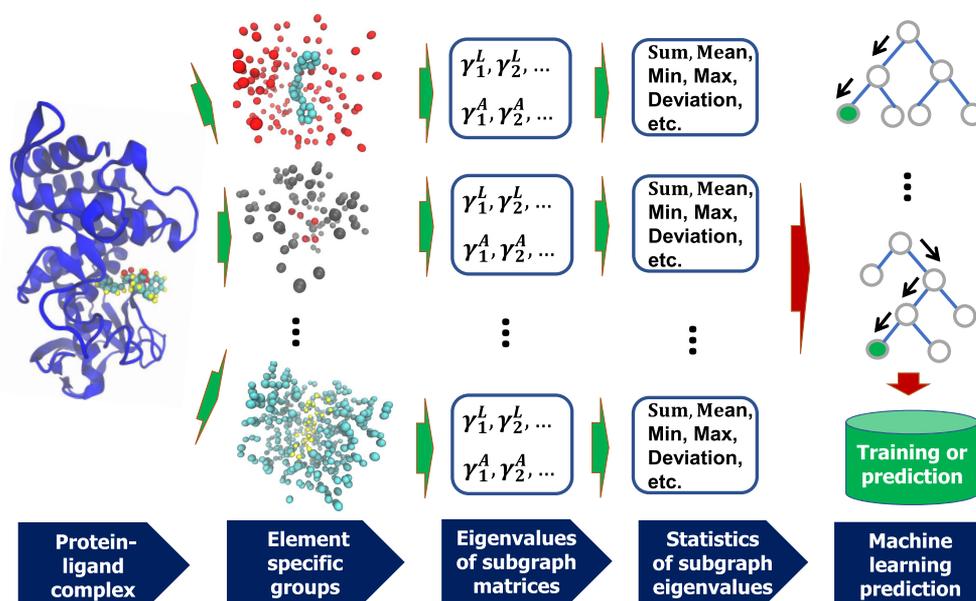


Figure 2. Illustration of algebraic graph learning strategy using IOSO (first column). In the second column, element specific groups are, from top to bottom, OC, NO, and CH, respectively. Their corresponding weighted labeled graph Laplacian and adjacency eigenvalues are shown in the third column. The statistics of these eigenvalues (fourth column) are used in gradient boosting trees for training and prediction (last column).

use a set of statistical values, namely, the sum of all positive eigenvalues, the mean of all positive eigenvalues, the largest (i.e., the principal) and the smallest positive eigenvalues, and the standard deviation of all positive eigenvalues as element-level molecular descriptors of molecules, biomolecules, and their interactions.

In principle, we can also construct atomic descriptors from $\{\lambda_j^A\}$ and $\{u_j^A\}$. Let us define a square matrix Q whose columns are n linearly independent eigenvectors of A : $Q = [u_1^A u_2^A \dots u_n^A]$ and a diagonal matrix Λ where each diagonal element Λ_{ii} is the eigenvalue associated with the i th column of Q . Then, a set of atomic descriptors can be obtained as

$$\mu_i^A(\eta_{kk'}) = \sum_j [Q \Lambda Q^{-1}]_{ij} \quad (10)$$

However, the method given in eq 10 is not a computationally efficient approach for describing atoms in molecules.

It might appear that the proposed algebraic graph theory-based method depends on eigenvalue analysis, which is normally very expensive. However, there are two facts that make the proposed method computationally efficient. First, only atoms in a small neighborhood of the protein–ligand binding site are involved in matrix constructions. Additionally, the element-specific selections further reduce the number of atoms involved in each matrix construction. As a result, one just needs to deal with many small matrices, rendering an efficient spectral approach for protein–ligand binding affinity analysis.

2.3. Graph Learning. To predict molecular and biomolecular properties, statistics of eigenvalues generated from the proposed weighted labeled subgraph Laplacian matrix or adjacency matrix will be combined with a machine learning algorithm. We assume the dataset is labeled and the problem is either a classification or a regression. From the machine learning point of view, we employ a supervised learning algorithm involving a training set and a test set. Denote X_i the dataset from the i th molecule or molecular complex in the

training dataset and let $G(X_i; \zeta)$ be a function that maps the geometric information into suitable graph representations with a set of parameters ζ consisting of kernel parameters. To set up a machine learning model, we cast the training into a minimization problem,

$$\min_{\zeta, \theta} \sum_{i \in I} \mathcal{L}(y_i, G(X_i; \zeta); \theta) \quad (11)$$

where \mathcal{L} is a scalar loss function to be minimized and y_i is the collection of labels in the training set. Here θ is the set of machine learning parameters to be optimized and depends on machine learning algorithms chosen. The loss function \mathcal{L} can be chosen according to the nature of the problem, i.e., regression and classification. Many machine learning algorithms, such as random forest, gradient boosting trees, artificial neural networks, and convolutional neural networks, can be employed in conjunction with the present graph descriptors. However, as our goal in the present work is to examine the descriptive power of the proposed algebraic graph features, let us focus on a relatively simple while still powerful machine learning algorithm, gradient boosting trees (GBTs). GBTs are very robust against overfitting³⁶ and their performance is quite similar to that of random forest. Figure 2 illustrates the proposed graph learning strategy.

Throughout this work, we choose GradientBoostingRegressor module implemented in the scikit-learn v0.19.1 package with parameters $n_estimators = 10000$, $max_depth = 7$, $min_samples_split = 3$, $learning_rate = 0.01$, $loss = ls$, $subsample = 0.3$, and $max_features = sqrt$. Changes in these parameters do not significantly affect the prediction results.

3. DATASETS AND EVALUATION METRICS

3.1. Datasets. In this work, we validate our proposed model against three commonly drug-discovery related benchmark datasets, namely, CASF-2007,⁶³ CASF-2013,⁶⁴ and CASF-2016.⁶⁶ These benchmarks are collected in the PDBbind database and have been used to evaluate the general

performance of a scoring function on a diverse set of protein–ligand complexes. The statistical information on these datasets is provided in Table 1. There is a wide variety of SFs in the

Table 1. Summary of PDBbind Datasets Used in the Present Work

	training set complexes	test set complexes
CASF-2007 benchmark	1105	195
CASF-2013 benchmark	3516	195
CASF-2016 benchmark	3772	285

binding affinity prediction and in pose scoring. In general, one can classify them into four groups:⁶⁹ (a) force-field-based or physical-based SFs, (b) empirical or linear regression-based SFs, (c) potential of the mean force (PMF) or knowledge-based SFs, and (d) machine learning-based SFs. The present method falls into the last category.

Note that for docking power and screening power assessments, additional data information is given for CASF-2007⁶⁴ and CASF-2013^{64,65} as described in the next section.

3.2. Evaluation Metrics. In this work, we assess the general performance of the proposed model by four metrics, namely, scoring power, ranking power, docking power, and screening power.^{63,65} These metrics and their associated datasets are briefly summarized below.

3.2.1. Scoring Power. This assessment validates a scoring function's ability to predict binding affinities that have a linear correlation with experimental data. The evaluation metric used in this task is the standard Pearson's correlation coefficient (R_p), given by

$$R_p = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (12)$$

where x_i and y_i are, respectively, predicted binding affinity and experimental data for the i th complex. Here \bar{x} and \bar{y} are the average of all predicted values and experimental values in the dataset, respectively. In this work, three datasets, CASF-2007, CASF-2013, and CASF-2016, are employed to test AGL-Score's scoring power.

3.2.2. Ranking Power. This assessment validates a scoring function's ability to rank binding affinities of protein–ligand complexes in each cluster.^{63,65} Both benchmarks CASF-2007 and CASF-2013 have 65 clusters of complexes, and each cluster has three complexes formed by the same protein but with different ligands. In the so-called high-level success measurement, the binding affinities of three complexes in each cluster are to be correctly ranked. While in the so-called low-level success measurement, a SF only needs to pick the structure with the highest binding affinity. The ranking power is evaluated by the percentage of successful identifications in a given benchmark.

In fact, the ranking power metric could be improved. Currently, it only counts the correct order of binding affinities of three native ligands for each target receptor in the core set. Also, it may not reflect the realistic setting of a real virtual screening process, in which a considerable number of ligands can bind to the same target. Additionally, more robust evaluation metrics, such as Kendall's tau or Spearman correlation coefficient, might be used.

3.2.3. Docking Power. This assessment validates a scoring function's ability to identify the “native” pose from docking

software generated poses.⁶³ In the benchmark, a pose is considered to be a native one if its root-mean-square deviation with respect to the true binding pose is less than 2 Å. In CASF-2007 benchmark, each ligand was given a total of 100 poses generated from docking software packages, namely, GOLD,^{70,71} Surflex,^{72,73} FLEX⁷⁴ and LigandFit.⁷⁵ In CASF-2013, 100 poses for each ligand were generated from three docking programs, namely, GOLD v5.1 (<https://www.ccdc.cam.ac.uk>), Surflex-Dock provided in SYBYL v8.1 (<https://www.certara.com/>), and MOE v2011 (<https://www.chemcomp.com/>). Note that the RMSD values in CASF-2007 were referred to as the standard RMSD algorithm.⁶³ However, a method may fail to report correct RMSDs on structures with a certain symmetry. Therefore, property-matched RMSD (RMSD^{PM}) values were provided in CASF-2013.^{64,65} In both benchmarks, there can be more than one “native” pose for each given ligand in the given dataset. Then if a method can identify any one of these native poses, it will be regarded as successful for the ligand. Docking power is evaluated by the number of ligands whose “native” poses are correctly identified.

3.2.4. Screening Power. This assessment validates a scoring function's ability to discriminate a target protein's true binders from decoy structures. There are a total of 65 different proteins in benchmark CASF 2013. Each receptor has at least three true binders. In fact, some of the 195 ligands in CASF 2013 dataset might bind to more than one protein.⁶⁴ Indeed, by searching through the ChEMBL database, one can show that 12 proteins have more than three true binders.⁶⁴ Fortunately, for each target protein, each of the 195 ligands is labeled as either a true binder or a decoy in the CASF dataset. Furthermore, for each target protein, the best true binder judged by the highest experimental binding affinity is specified in the CASF dataset.

There are two kinds of screening power measurements. The task of the first screening power measurement is to find out the enrichment factor (EF) in $x\%$ top-ranked candidates:

$$EF_{x\%} = \frac{\text{no. of true binders among } x\% \text{ top-ranked candidates}}{\text{total no. of true binders of the given target protein}} \quad (13)$$

Here, top-ranked candidates are defined as those SF-predicted candidates that have high binding affinities. The average of all EF values over 65 target proteins is used to assess the screening power of a scoring function.

The task of the second screening power measurement is to identify the best true binder. The success rate is given by the percentage of the best binders of 65 receptors being found among $x\%$ top-ranked candidates.

4. RESULTS AND DISCUSSION

Herein we assess the scoring power, ranking power, docking power, and screening power of the proposed algebraic graph learning (AGL) approach using aforementioned benchmark datasets and evaluation metrics.

4.1. Model Parametrization. For the sake of convenience, we use the notation $AGL_{\Omega, \beta, \tau}^M$ to indicate the algebraic graph learning features generated by using interactive matrix type M with kernel type Ω and corresponding kernel parameters β and τ . As such, $M = \text{Adj}$, $M = \text{Lap}$, and $M = \text{Inv}$ represent adjacency matrix, Laplacian matrix, and pseudo inverse of

Laplacian matrix, respectively. Here, $\Omega = E$ and $\Omega = L$ refer to generalized exponential and generalized Lorentz kernels, respectively. Additionally, β is the kernel order such that $\beta = \kappa$ if $\Omega = E$, and $\beta = \nu$ if $\Omega = L$. Finally, τ is used such that $\eta_{kk'} = \tau(\bar{r}_k + \bar{r}_{k'})$, where \bar{r}_k and $\bar{r}_{k'}$ are the van der Waals radii of element type k and element type k' , respectively.

We propose an AGL representation in which multiple kernels are parametrized at different scale (η) values. In this work, we consider at most two kernels. As a straightforward notation extension, two kernels can be parametrized by $AGL_{\Omega, \beta, \tau}^{M_1, M_2}$.

4.2. Hyperparameter Optimization. As a rule of thumb, the machine learning models achieve the best performance when their essential parameters are properly optimized. To this end, 5-fold cross-validation (CV) is carried out to tune the kernel hyperparameters M , Ω , β , and τ in the proposed model $AGL_{\Omega, \beta, \tau}^M$. For simplicity, we perform the kernel parameter optimization on CASF-2007 benchmark's training data ($N = 1105$), the smallest training set among three benchmarks. Based on our previous work,^{42,45} a reasonable range of hyperparameters is listed in Table 2. Specifically, the scale

Table 2. Ranges of Model Hyperparameters for Five-Fold Cross-Validations

parameter	domain
τ	{0.5, 1.0, ..., 6}
β	{0.5, 1.0, ..., 6} \cup {10, 15, 20}
M	{Adj, Lap, Inv}

factor τ and power parameters $\beta = \kappa$ or ν are chosen in [0.5, 6] with an increment of 0.5 so that our model can effectively represent the interactions between protein and ligand in a

complex. In addition, high values of the power order such as $\beta \in \{10, 15, 20\}$ are also taken into account to approximate the ideal low-pass filter (ILF).⁴² There are 40 element interactive pairs formed by the combinations of 4 commonly occurring atom types in proteins, i.e., {C, N, O, S}, and 10 commonly occurring atom types in ligands, i.e., {H, C, N, O, F, P, S, Cl, Br, I}. For adjacency matrices, we consider only their positive eigenvalues. (Note that Laplacian matrices are positive semidefinite.) From the resulting set of eigenvalues or corresponding atomic descriptors, one can compute nine descriptive statistical values, namely the sum, minimum (i.e., the Fiedler value for Laplacian matrices or the half band gap for adjacency matrices), maximum, mean, median, standard deviation, and variance of all eigenvalues. Additionally, we also utilize the number of eigenvalues and the sum of the second power of eigenvalues. This gives rise to a total of 360 features.

For a given interaction matrix type M and a given kernel type Ω , we carry out five-fold cross validations on the training data of CASF-2007 to search for the optimal parameters β and τ based on the averaged Pearson correlation coefficient value (R_p). Figure S1 in the supplement material reports the best models with associated R_p in this experiment. The optimal models are ($AGL_{E, 6, 2.5}^{Adj}$, $R_p = 0.748$), ($AGL_{E, 10, 3.5}^{Lap}$, $R_p = 0.74$), ($AGL_{E, 1.5, 4.5}^{Inv}$, $R_p = 0.708$), ($AGL_{L, 3.5, 1.5}^{Adj}$, $R_p = 0.749$), ($AGL_{L, 1.5, 3}^{Lap}$, $R_p = 0.740$), and ($AGL_{L, 3.5, 4}^{Inv}$, $R_p = 0.706$). Among them, $AGL_{L, 3.5, 1.5}^{Adj}$ is the best model and $AGL_{L, 3.5, 4}^{Inv}$, $R_p = 0.706$ is the worst one. This finding is no surprise. In fact, adjacency matrix is the simplest one but still effectively captures all the interactions between protein and ligand atoms. Since the GNM-style matrix, i.e., $M = Inv$, involves the Moore–Penrose inverse, it likely admits errors from the numerical evaluation of large eigenvalues.

It is reported in the literature that the multiscale information can boost predictor's performance.^{40,45} Thus, on top of the

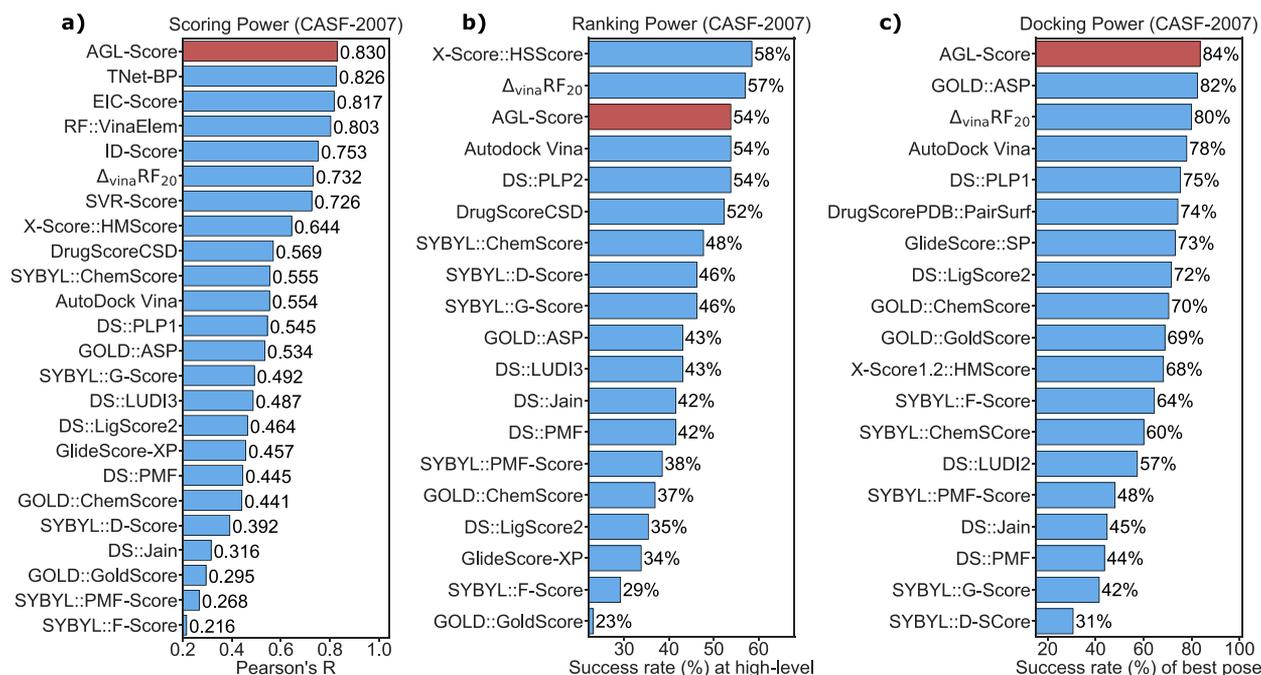


Figure 3. Performance comparison of different scoring functions on CASF-2007 benchmark: (a) scoring power measured by Pearson correlation coefficient, (b) ranking power evaluated by the high-level success measurement, and (c) docking power in terms of the rate of successfully identifying the “native” pose from 100 poses for each ligand. The proposed algebraic graph learning-based scoring function, AGL-Score, is plotted in the red color. The results of other methods, taken from refs 36, 53, 58, 63, and 76–78, are in the blue color.

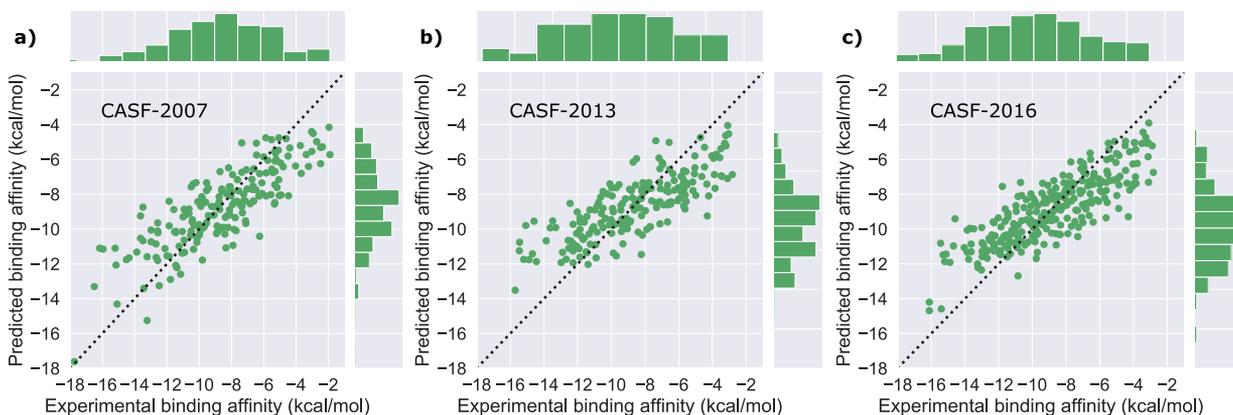


Figure 4. Correlation between AGL-Score predictions and experimental data for various benchmarks. (a) CASF-2007: Pearson correlation coefficient $R_p = 0.83$ and RMSE = 1.864 kcal/mol. (b) CASF-2013: Pearson correlation coefficient $R_p = 0.792$ and RMSE = 1.973 kcal/mol. (c) CASF-2016: Pearson correlation coefficient $R_p = 0.833$ and RMSE = 1.733 kcal/mol.

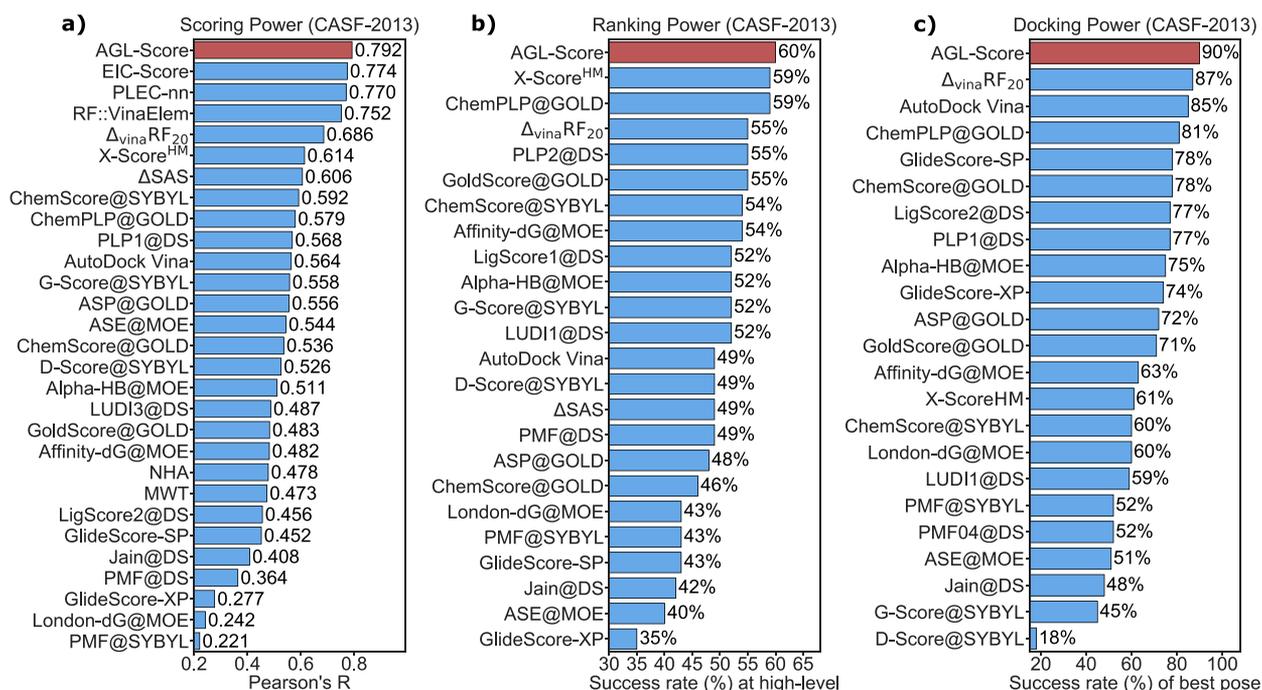


Figure 5. Performance comparison of different scoring functions on CASF-2013 benchmark: (a) scoring power measured by Pearson correlation coefficient, (b) ranking power evaluated by the high-level success measurement, and (c) the docking power in terms of the rate of successfully identifying the “native” pose from 100 poses for each ligand. The proposed algebraic graph learning-based scoring function, AGL-Score, is plotted in the red color. The results of other methods, taken from refs 58, 64, and 78–80, are in the blue color.

optimal one-scale model, we impose another kernel with a different parametrization. We also carry out a similar grid-search procedure as we did for the single-scale model to explore optimal parameters. Based on Figure S2, the best two-kernel models are found at the following ($AGL_{E,6,2,5;E,4,2}^{Adj}$, $R_p = 0.75$), ($AGL_{E,10,3,5;E,5,1,5}^{Lap}$, $R_p = 0.745$), ($AGL_{E,1,5,4,5;E,4,5,2}^{Inv}$, $R_p = 0.714$), ($AGL_{L,3,5,1,5;L,15,0,5}^{Adj}$, $R_p = 0.751$), ($AGL_{L,1,5,3;L,6,1}^{Lap}$, $R_p = 0.745$), and ($AGL_{L,3,5,4;L,10,1}^{Inv}$, $R_p = 0.715$). It is clear that models involving the adjacency matrix, i.e., $AGL_{E,6,2,5;E,4,2}^{Adj}$ and $AGL_{L,3,5,1,5;L,15,0,5}^{Adj}$ still outperform the rest. Finally, we form a consensus model named AGL-Score that is defined by the mean of the predicted values produced by those two aforementioned AGL models.

4.3. Performance and Discussion. 4.3.1. Scoring Power.

First of all, we validate the scoring power of the proposed

AGL-Score using the CASF-2007 benchmark. We train two AGL models, namely $AGL_{E,6,2,5;E,4,2}^{Adj}$ and $AGL_{L,3,5,1,5;L,15,0,5}^{Adj}$ on the refined set ($N = 1105$) of the PDBbind v2007 excluding the test set ($N = 195$) of CASF-2007 benchmark. For the prediction task, we repeat each AGL model up to 50 times. The average of all the predicted values is used as the predicted binding affinity of the AGL model. It is noted that the energy unit in the PDBbind database is pK_d . For the kcal/mol unit conversion, we multiply the predicted values by -1.3633 .⁵⁷ In addition, we are interested in comparing the predictive power of our AGL-Score with various state-of-the-art SFs introduced in the literature.^{36,53,58,63,76–78} Figure 3a illustrates such a comparison. Clearly, the proposed model is one of the most accurate SFs in this benchmark with a Pearson correlation coefficient value $R_p = 0.830$ and RMSE = 1.864 kcal/mol. The

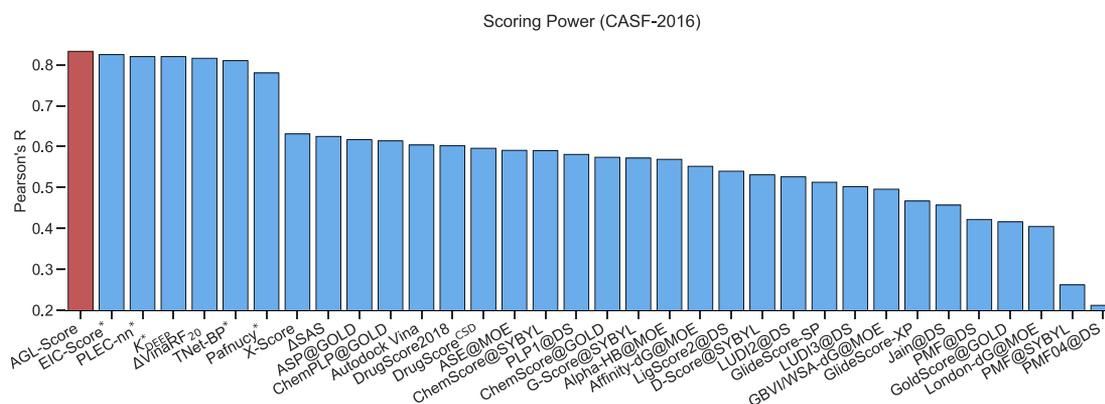


Figure 6. Performance comparison of different scoring functions on CASF-2016. The Pearson correlation coefficients of other methods are taken from refs 36, 66, 78, and 80–82. The proposed algebraic graph learning-based scoring function, AGL-Score, achieves $R_p = 0.833$ and RMSE = 1.733 kcal/mol. Note that scoring functions marked with * use PDBbind v2016 core set ($N = 290$).

runner-up is TNet-BP with reported $R_p = 0.826$.³⁶ The geometric graph approach has a slightly lower performance with $R_p = 0.825$.⁴⁵ This comparison confirms the scoring power of the present model. Furthermore, the correlation between our predicted values and the experimental data is depicted in Figure 4a.

In the second benchmark, i.e., CASF-2013, its training data ($N = 3516$) is compiled on the basis of the refined set of PDBbind v2015 excluding its test set ($N = 195$). We carry out a similar prediction procedure as of the previous one. Interestingly, the proposed model is able to outperform the state-of-the-art SFs adopted from refs 58, 64, and 78–80, as seen in Figure 5b. Specifically, our AGL-Score attains Pearson correlation coefficient value $R_p = 0.792$ and RMSE = 1.973 kcal/mol followed by EIC-Score model with $R_p = 0.774$.⁷⁸ Our previous geometric graph model achieves similar performance to AGL-Score with $R_p = 0.782$.⁴⁵ In addition, Figure 4b provides a scatter plot to illustrate the correlation between our predicted values and experimental results.

CASF-2016 is the last benchmark considered in this work. It is also the latest CASF released by PDBbind database.⁶⁶ We train the AGL model on the basis of the refined set of PDBbind v2016 excluding the benchmark's test set. In this experiment, both training data ($N = 3772$) and test data ($N = 285$) are slightly larger than their predecessor, CASF-2013. The performance of a number of SFs has been reported for this benchmark. Specially, K_{DEEP} ,⁸¹ Pafnucy,⁸² and PLEC-nn⁸⁰ SFs make use of deep learning architectures. Also, numerous SF models have been tested by the PDBbind team.⁶⁶ We compare the proposed AGL-Score to the aforementioned methods in Figure 6. Note that, K_{DEEP} , Pafnucy, PLEC-nn, and EIC-Score all carry out the predictions on the original PDBbind v2016 core set with $N = 290$. Table S1 provides the discrepancy information between the PDBbind v2016 core set and CASF-2016 test set. Our AGL-Score is still superior to its counterparts with $R_p = 0.833$ (0.835) and RMSE = 1.733 (1.732) kcal/mol for CASF-2016 test set (PDBbind v2016 core set). The second best approaches in the chart are K_{DEEP} and PLEC-nn both having $R_p = 0.82$. The features and machine learning algorithms of these SFs are listed in Table 3. Moreover, the algebraic graph model's performance on this benchmark is still superior to that of the earlier geometric graph approach with $R_p = 0.815$.⁴⁵ This result confirms the accuracy and reliability of the AGL model for diversified

Table 3. Description of Some Machine Learning-Based Scoring Functions

name	feature	machine learning algorithm
TNet-BP ³⁷	algebraic topology	convolution neural network (CNN)
EIC-Score ⁷⁹	differential geometry	GBT
RI-Score ⁴⁶	geometric graph	random forest
AGL-Score	algebraic graph	GBT
PLEC-nn ⁸¹	extended connectivity fingerprint	neural networks
K_{DEEP} ⁸²	3D voxel representation	CNN
Pafnucy ⁸³	3D voxel representation	CNN
RF::VinaElem ⁷⁸	intermolecular contacts and Autodock Vina features	random forest
$\Delta_{\text{vina}}\text{RF}_{20}$ ⁵⁹	Autodock Vina and other physical features	random forest

binding affinity datasets. Finally, the comparison between the predicted affinities of AGL model and the experimental values is depicted in Figure 4c.

4.3.2. Ranking Power. We use predicted binding affinities on the scoring power task to validate the performance of our proposed scoring function on the ranking power. Among two models, namely generalized exponential kernel model $\text{AGL}_{E,6,2.5;E,4,2}^{\text{Adj}}$ and generalized Lorentz kernel model $\text{AGL}_{L,3.5,1.5;L,15,0.5}^{\text{Adj}}$, the generalized exponential model produces better results on both benchmarks (see Figures S3 and S4). Thus, its results are used in Figures 3b and 5b. In CASF-2013, AGL-Score achieved the best performance with the high-level success measurement being 60%, followed by X-Score^{HM} with the success rate being 59%. In CASF-2007, the AGL-Score's rate for the high-level success measurement is 54%, which is ranked the third, following X-Score::HSScore (success rate = 58%)⁶⁴ and $\Delta_{\text{vina}}\text{RF}_{20}$ (success rate = 57%). One can see that top ranking SFs for the scoring power also perform quite well in the ranking SFs power assessment. For example, in both CASF-2007 and CASF-2013, our AGL-Score model is among the top ranking SFs on both scoring power and ranking power. If the machine learning models are excluded in the scoring tests, one can conclude that $\Delta_{\text{vina}}\text{RF}_{20}$ and X-Score^{HM} are good performers on these assessments.

4.3.3. Docking Power. In order for our scoring function-based machine learning method to be able to recognize the “native” pose among the computer-generated 100 poses, there

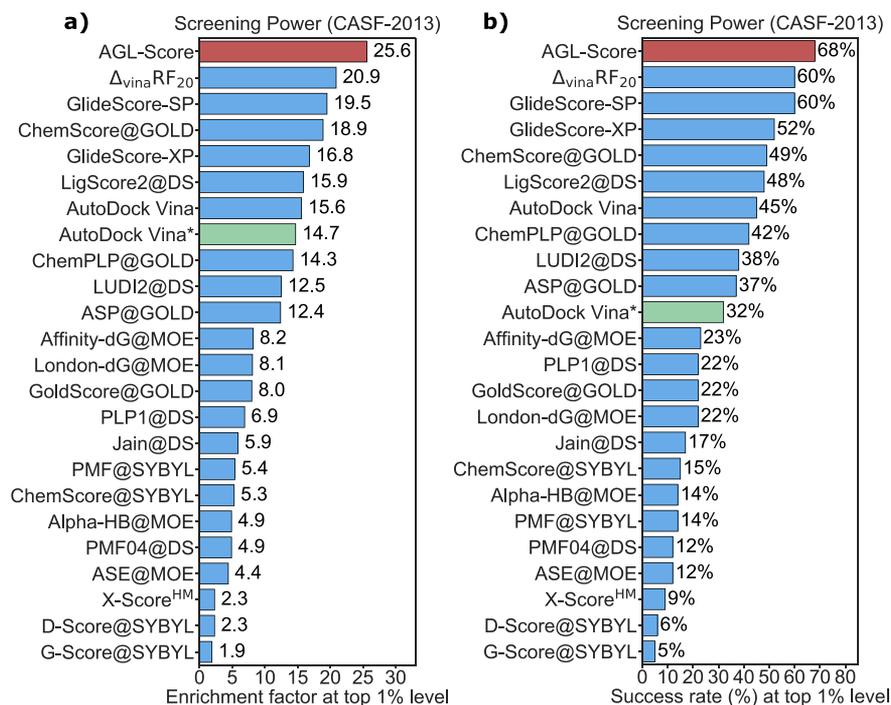


Figure 7. Performance comparison of different scoring functions on the screening power for CASF-2013 benchmark evaluated by (a) enrichment factor and (b) success rate at the top 1% level. The proposed algebraic graph learning-based scoring function, AGL-Score, is in the red color. The results of AutoDock Vina computed in the present work (AutoDock Vina*) are in the green color. The results of other methods, taken from refs 58 and 64, are in the blue color.

is a need to create a training set which includes diverse conformers. To this end, for each given target ligand binding to a specific receptor, we use GOLD v5.6.3⁷¹ to generate a set of 1000 training poses. The parameters of the GOLD software are chosen as the following: autoscale = 1.5, early_termination = 0, and gold_fitfunc_path = plp. For each of CASF-2007 and CASF-2013, we have created a total of 365 000 training poses for machine learning analysis. These pose structures and their scores reported by GOLD are available at <https://weilab.math.msu.edu/AGL-Score>.

For each target ligand, we retrain AGL-Score model on the target-specifically generated 1000 poses and use docking software's scores as their labels. To reduce the calculation time cost, we employ single exponential kernel AGL-Score $AGL_{E,6,2,5}^{Adj}$. Figures 3c and 5c reveal that our AGL-Score is the top performer on the docking power test. Specifically, the AGL-Score's success rate in identifying the best pose as the native pose is 84% on CASF-2007, followed by GOLD::ASP (82%)⁶⁴ and $\Delta_{vina}RF_{20}$ (80%).⁵⁹ In CASF-2013, our AGL-Score is still ranked at the first place with its success rate = 90%, followed by $\Delta_{vina}RF_{20}$ (87%)⁵⁹ and AutoDock Vina (85%).⁵⁹

Moreover, we are interested in examining the quality of the training data generated by GOLD using ChemPLP score (ChemPLP@GOLD) in term of the success rate of the best pose. Specifically, ChemPLP@GOLD scores 67% and 82% on the datasets generated for CASF-2007 and CASF-2013, respectively. By employing such datasets, our AGL-Score improves the success rates to 84% and 90% on CASF-2007 and CASF-2013, respectively. These statistical results strongly affirm that our AGL-Score is able to capture the real physical interactions in protein–ligand binding complexes and

significantly improve the performance of existing docking software.

To do well on both scoring power and docking power tests is still a major challenge in the SF development.^{60,83,84} Most reputable docking software packages might offer reliable accuracy on pose predictions but perform poorly on binding affinity predictions. For example, for CASF-2007, GOLD software with ASP as a SF attains 82% in the success rate for the docking power task. However, it performs unsatisfactorily on the binding affinity prediction with Pearson correlation coefficient (R_p) being 0.534. On the other hand, RF-IChem,⁶⁰ a SF-based machine learning, achieves a more respectable $R_p = 0.791$ on the scoring power test. However, it fails desperately on identifying the native pose task with the success rate being less than 30%. Recently, another machine learning model named $\Delta_{vina}RF_{20}$ was developed.⁵⁸ It produces a great performance on docking power task with the success rate being 80%, while its Pearson correlation coefficient ($R_p = 0.732$) is also better than other traditional SFs. However, if one includes machine learning SFs, $\Delta_{vina}RF_{20}$ is still quite behind the best in the literature, i.e., TNet-BP ($R_p = 0.826$).³⁷ Our AGL-Score not only achieves significant accuracy in the binding affinity prediction in various benchmarks ($R_p = 0.83$ in CASF-2007) but also attains the top place in the docking power assessment (success rate = 84% in CASF-2007). These results again rigorously confirm that the proposed eigenvalue features of multiscale weighted colored graphs have an ability to accurately encode the physical and biological information of protein–ligand complexes.

4.3.4. Screening Power. In this assessment, our AGL-Score $AGL_{E,6,2,5}^{Adj}$ outperforms other SFs on both screening power measurements, namely enrichment factor (EF) and success rate at top 1% level on the CASF-2013 benchmark. Specifically,

the AGL-Score obtains the enrichment factor of 25.6. The methods at the second and third places are Δ vinaRF₂₀ (EF = 20.9)⁵⁹ and GlideScore-SP (EF = 19.5).⁶⁵ On the success rate, our model is also the top performer with the success rate being 68%. Δ vinaRF₂₀ and GlideScore-SP are tied for the second place with a success rate being 60%. Figure 7 depicts the performance detail of various methods.

Since screening power assessment involves the identification of true binders for each of 65 proteins, we need to retrain our AGL-Score model. To this end, we collect a training set which includes both poses and energy labels to set up an AGL-Score model for each protein. Then, the poses of the 195 ligands are generated with a docking procedure and their energies are predicted by our AGL-Score model. The ligands with high energies are regarded as predicted binders.

Here, we describe our procedure for the training dataset selection and generation for the screening power test. The training set for each target protein consists of all complex structures and their energy labels from PDBbind v2015 refine set, excluding the core (test) set complexes. Additionally, for a given target protein, additional poses and their labels in the training set are generated by using Autodock Vina.⁸⁵ For a given target protein, Autodock Vina is used to dock all the ligands in the PDBbind v2015 refined set, carefully excluding those in the core-set and true binders of the target protein. Autodock Vina is parametrized as follows: exhaustiveness = 10, num_modes = 10, and energy_range = 3. Note that the list of true binders for each protein is listed in CASF 2013 benchmark. For each ligand, we keep the pose with the highest energy for the training set. This procedure gives rise to a few thousands of additional training poses and associated energy labels for each target protein.

Also, to use Autodock Vina score as the energy label, we convert kcal/mol to pK_d unit by multiplying the Autodock Vina score with a conversion factor -1.3633.⁵⁷ Since those ligands in the refined set that do not bind to the given target protein are regarded as decoys,^{64,65} their binding energies should not be larger than those of the true binders of the given target protein. Therefore, if the decoy energy predicted by Autodock Vina is higher than the lower bound of true binders' energies in the refined set database, we relabel the decoy with the lower bound of the true binders. The additional training set, i.e., Autodock Vina generated poses and their energy labels, can be downloaded via link <https://weilab.math.msu.edu/AGL-Score>.

It is interesting to see that the performances of Autodock Vina are much lower than that of our AGL-Score. Specifically, EF of Autodock Vina's training set is 14.7 while that of AGL-Score is 25.6. In addition, the success rate at top 1% level of Autodock Vina's training set is 32% while AGL-Score gains more than 100% improvement with a success rate of 68%. These results illustrate the remarkable improvement of our AGL-Score in the screening power in comparison to the well-known docking software.

The screening power assessment is still a difficult task for many machine learning-based SF models.⁸⁴ This type of SFs often performs well on scoring power test but works poorly on the identification of the true binders among the decoys. Particularly, the machine learning model RF@ML reported in⁸⁴ achieves a reasonable $R_p = 0.704$ on the scoring test for CASF-2013. This result is better than that of Δ vinaRF₂₀ ($R_p = 0.686$). However, RF@ML's performance on the screening task is not comparable to Δ vinaRF₂₀ as well as top docking

SFs. Specifically, RF@ML attains EF at top 1% level as low as 2.15, and success rate at top 1% level as low as 6.45 in CASF-2013. GlideScore-SP, by contrast, produces much higher EF (19.5) and success rate (60%). By utilizing the superior algebraic graph representation features and appropriate training set, our AGL-Score is able to deliver top rankings in both screening power assessments without sacrificing the scoring accuracy.

5. CONCLUSION

Algebraic graph theories are commonly used in the study of molecular and biomolecular systems. However, most algebraic graph theory-based models are not as competitive as another predictive models on the same tasks.^{33–36} Even for some powerful methods on binding affinity predictions, they often perform poorly on other tasks involving unphysical and unreliable structures.⁶⁶ Motivated by our previous work on the multigraph approaches for B-factor predictions,⁴⁶ we propose a novel algebraic graph learning score (AGL-Score) for dealing with drug design related problems. The proposed AGL-Score model makes use of multiscale weight colored subgraphs to encode the essential physical and biological information, such as hydrogen bonds, electrostatics, van der Waals interactions, hydrophilicity, and hydrophobicity in the high-dimension space into the low-dimension representation of molecular and biomolecular structures. The constructions of three types of subgraphs are discussed in this work, namely adjacency matrix, Laplacian matrix, and pseudo-inverse of Laplacian matrix. The eigenvalues calculated from such matrices are used as features to characterize the biological and physical interactions of molecules and biomolecules.

In this work, we first investigate binding affinity datasets to demonstrate the robustness, accuracy, and reliability of the proposed model. To this end, three mainstream benchmark tests on scoring power, ranking power, docking power, and screening power assessments, namely CASF-2007,⁶³ CASF-2013,^{64,65} and CASF-2016,⁶⁶ are utilized. The results of benchmark tests reveal the superior performance of the proposed AGL-Score over other state-of-the-art methods. Additionally, we consider the ranking power, docking power and screening power assessments proposed in the literature.^{63,64} Additional training sets are generated to retrain the proposed AGL-Score for docking power and screening power assessments. Extensive numerical experiments rigorously confirm the top performance of the proposed AGL-Score model on benchmark protein–ligand binding datasets, containing both X-ray crystal and diverse decoy structures.

There has been widely spread scepticism or misunderstanding of the ability of machine learning-based SFs for docking and virtual screening.⁶⁰ The present work shows that our machine learning method is not only able to enrich virtual screening but also able to significantly improve the performance of standard docking software packages, such as AutoDock Vina.

In addition to the confirmed accuracy and reliability, another major advantage of the present AGL-Score model is its simplicity. Only raw structural inputs regarding atom types and coordinates are used for free energy predictions. There is no need for any molecular force field. Moreover, the present AGL-Score model is robust without invoking complicated data pre-processing and optimization procedures.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00334.

Cross-validation performance of AGL models on the training dataset of CASF-2007 benchmark, some additional results on ranking power assessments, and discrepancy between CASF-2016 test set and PDBbind v2016 core set, including Figures S1–S4 and Table S1 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: weig@msu.edu.

ORCID

Guo-Wei Wei: 0000-0002-5781-2937

Notes

The authors declare no competing financial interest.

The algebraic graph learning score (AGL-Score) model is implemented on our own hosted server at <https://weilab.math.msu.edu/AGL-Score>. Training sets for the docking power assessment and additional training sets for the screening power assessment are also available for download at <https://weilab.math.msu.edu/AGL-Score>.

■ ACKNOWLEDGMENTS

This work was supported in part by NSF Grants DMS-1721024, IIS1900473, DMS-1761320, and NIH grant GM126189. D.D.N. and G.-W.W. are also funded by Bristol-Myers Squibb and Pfizer.

■ REFERENCES

- (1) Hosoya, H. Topological Index. a Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (2) Hansen, P. J.; Jurs, P. C. Chemical Applications of Graph Theory. Part I. Fundamentals and Topological Indices. *J. Chem. Educ.* **1988**, *65*, 574.
- (3) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-parameter Harmonic Potential. *Folding Des.* **1997**, *2*, 173–181.
- (4) Yang, L. W.; Chng, C. P. Coarse-grained Models Reveal Functional Dynamics—I. Elastic Network Models—theories, Comparisons and Perspectives. *Bioinf. Biol. Insights* **2008**, *2*, 25–45.
- (5) Newman, M. *Networks: An Introduction*; Oxford University Press: New York, 2010.
- (6) Bavelas, A. Communication Patterns in Task-Oriented Groups. *J. Acoust. Soc. Am.* **1950**, *22*, 725–730.
- (7) Dekker, A. Conceptual Distance in Social Network Analysis. *J. Soc. Struct.* **2005**, *6*, online.
- (8) Wei, G. W.; Zhan, M.; Lai, C. H. Tailoring Wavelets for Chaos Control. *Phys. Rev. Lett.* **2002**, *89*, 284103.
- (9) Kac, M. Can One Hear the Shape of a Drum? *Am. Math. Mon.* **1966**, *73*, 1–23.
- (10) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983.
- (11) Schultz, H. P. Topological Organic Chemistry. I. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Model.* **1989**, *29*, 227–228.
- (12) Janežič, D.; Milčević, A.; Nikolić, S.; Trinajstić, N. *Graph-Theoretical Matrices in Chemistry*; CRC Press: Boca Raton, FL, 2015.
- (13) Angeleska, A.; Jonoska, N.; Saito, M. DNA Rearrangement Through Assembly Graphs. *Discrete. Appl. Math.* **2009**, *157*, 3020–3037.

(14) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of A Small Globular Protein in Terms of Low-frequency Vibrational Modes. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 3696–3700.

(15) Tasumi, M.; Takeuchi, H.; Ataka, S.; Dwivedi, A. M.; Krimm, S. Normal Vibrations of Proteins: Glucagon. *Biopolymers* **1982**, *21*, 711–714.

(16) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(17) Levitt, M.; Sander, C.; Stern, P. S. Protein Normal-mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *J. Mol. Biol.* **1985**, *181*, 423–447.

(18) Flory, P. J.; Gordon, M.; McCrum, N. G. Statistical Thermodynamics of Random Networks. *Proc. R. Soc. London A* **1976**, *351*, 351–380.

(19) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Proteins: Significance of Slow and Fast Modes in Relation to Function and Stability. *Phys. Rev. Lett.* **1998**, *80*, 2733–2736.

(20) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80*, 505–515.

(21) Hinsen, K. Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417–429.

(22) Tama, F.; Sanejouand, Y. H. Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng., Des. Sel.* **2001**, *14*, 1–6.

(23) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; Chapman and Hall/CRC: Boca Raton, FL, 2010.

(24) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic Press: Cambridge, MA, 1976.

(25) Foulds, L. R. *Graph Theory Applications*; Springer Science & Business Media: New York, 2012.

(26) Ozkanlar, A.; Clark, A. E. ChemNetworks: A Complex Network Analysis Tool for Chemical Systems. *J. Comput. Chem.* **2014**, *35*, 495–505.

(27) Di Paola, L.; Giuliani, A. Protein Contact Network Topology: A Natural Language for Allostery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 43–48.

(28) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L. A Graph-Theory Algorithm for Rapid Protein Side-Chain Prediction. *Protein Sci.* **2003**, *12*, 2001–2014.

(29) Ryslik, G. A.; Cheng, Y.; Cheung, K.-H.; Modis, Y.; Zhao, H. A Graph Theoretic Approach to Utilizing Protein Structure to Identify Non-Random Somatic Mutations. *BMC Bioinf.* **2014**, *15*, 86.

(30) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein Flexibility Predictions Using Graph Theory. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 150–165.

(31) Vishveshwara, S.; Brinda, K.; Kannan, N. Protein Structure: Insights from Graph Theory. *J. Theor. Comput. Chem.* **2002**, *1*, 187–211.

(32) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *arXiv preprint arXiv:1703.00564*, 2017; <https://arxiv.org/abs/1703.00564>.

(33) Pires, D. E. V.; Ascher, D. B.; Blundell, T. L. MCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* **2014**, *30*, 335–342.

(34) Quan, L.; Lv, Q.; Zhang, Y. STRUM: Structure-Based Prediction of Protein Stability Changes upon Single-Point Mutation. *Bioinformatics* **2016**, *32*, 2936–2946.

(35) Cang, Z. X.; Wei, G. W. Analysis and Prediction of Protein Folding Energy Changes upon Mutation by Element Specific Persistent Homology. *Bioinformatics* **2017**, *33*, 3549–3557.

(36) Cang, Z. X.; Wei, G. W. TopologyNet: Topology Based Deep Convolutional and Multi-task Neural Networks for Biomolecular

Property Predictions. *PLoS Comput. Biol.* **2017**, *13* (7), No. e1005690.

(37) Park, J. K.; Jernigan, R.; Wu, Z. Coarse Grained Normal Mode Analysis Vs. Refined Gaussian Network Model for Protein Residue-Level Structural Fluctuations. *Bull. Math. Biol.* **2013**, *75*, 124–160.

(38) Xia, K.; Wei, G.-W. A Review of Geometric, Topological and Graph Theory Apparatuses for the Modeling and Analysis of Biomolecular Data. *arXiv preprint arXiv:1612.01735*, 2016; <https://arxiv.org/abs/1612.01735>

(39) Opron, K.; Xia, K. L.; Wei, G. W. Fast and Anisotropic Flexibility-Rigidity Index for Protein Flexibility and Fluctuation Analysis. *J. Chem. Phys.* **2014**, *140*, 234105.

(40) Opron, K.; Xia, K. L.; Wei, G. W. Communication: Capturing Protein Multiscale Thermal Fluctuations. *J. Chem. Phys.* **2015**, *142*, 211101.

(41) Nguyen, D. D.; Xia, K. L.; Wei, G. W. Generalized Flexibility-Rigidity Index. *J. Chem. Phys.* **2016**, *144*, 234106.

(42) Xia, K. L.; Opron, K.; Wei, G. W. Multiscale Gaussian Network Model (mGNM) and Multiscale Anisotropic Network Model (mANM). *J. Chem. Phys.* **2015**, *143*, 204106.

(43) Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological Persistence and Simplification. *Discrete Comput. Geom.* **2002**, *28*, 511–533.

(44) Zomorodian, A.; Carlsson, G. Computing Persistent Homology. *Discrete Comput. Geom.* **2005**, *33*, 249–274.

(45) Nguyen, D. D.; Xiao, T.; Wang, M. L.; Wei, G. W. Rigidity strengthening: A mechanism for protein-ligand binding. *J. Chem. Inf. Model.* **2017**, *57*, 1715–1721.

(46) Bramer, D.; Wei, G. W. Weighted Multiscale Colored Graphs for Protein Flexibility and Rigidity Analysis. *J. Chem. Phys.* **2018**, *148*, 054103.

(47) Zheng, Z.; Merz, K. M., Jr Ligand Identification Scoring Algorithm (LISA). *J. Chem. Inf. Model.* **2011**, *51*, 1296–1306.

(48) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. I. Knowledge Based Ligand-Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus Protease Binding Affinity. *Protein Eng., Des. Sel.* **1995**, *8*, 677–691.

(49) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. the Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(50) Muegge, I.; Martin, Y. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(51) Velec, H.; Gohlke, H.; Klebe, G. DrugScore (CSD)-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* **2005**, *48*, 6296–303.

(52) Huang, S. Y.; Zou, X. An Iterative Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions: I. Derivation of Interaction Potentials. *J. Comput. Chem.* **2006**, *27*, 1866–1875.

(53) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169–1175.

(54) Kinnings, S. L.; Liu, N.; Tonge, P. J.; Jackson, R. M.; Xie, L.; Bourne, P. E. A Machine Learning Based Method to Improve Docking Scoring Functions and Its Application to Drug Repurposing. *J. Chem. Inf. Model.* **2011**, *51*, 408–419.

(55) Ashtawy, H. M.; Mahapatra, N. R. A Comparative Assessment of Ranking Accuracies of Conventional and Machine-Learning-Based Scoring Functions for Protein-Ligand Binding Affinity Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2012**, *9*, 1301–1313.

(56) Lee, A. A.; Brenner, M. P.; Colwell, L. J. Predicting Protein-Ligand Affinity with a Random Matrix Framework. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13564–13569.

(57) Wang, B.; Zhao, Z.; Nguyen, D. D.; Wei, G. W. Feature Functional Theory - Binding Predictor (FFT-BP) for the Blind Prediction of Binding Free Energy. *Theor. Chem. Acc.* **2017**, *136*, 55.

(58) Wang, C.; Zhang, Y. Improving Scoring-Docking-Screening Powers of Protein-Ligand Scoring Functions Using Random Forest. *J. Comput. Chem.* **2017**, *38*, 169–177.

(59) Nguyen, D. D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.-W. Mathematical Deep Learning for Pose and Binding Affinity Prediction and Ranking in D3R Grand Challenges. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 71–82.

(60) Gabel, J.; Desaphy, J.; Rognan, D. Beware of Machine Learning-Based Scoring Functions: On the Danger of Developing Black Boxes. *J. Chem. Inf. Model.* **2014**, *54*, 2807–2815.

(61) Li, Y.; Yang, J. Structural and Sequence Similarity Makes a Significant Impact on Machine-Learning-Based Scoring Functions for Protein-Ligand Interactions. *J. Chem. Inf. Model.* **2017**, *57*, 1007–1012.

(62) Kramer, C.; Gedeck, P. Leave-Cluster-out Cross-Validation Is Appropriate for Scoring Functions Derived from Diverse Protein Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 1961–1969.

(63) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.

(64) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717–1736.

(65) Li, Y.; Su, M.; Liu, Z.; Li, J.; Liu, J.; Han, L.; Wang, R. Assessing Protein-Ligand Interaction Scoring Functions with the CASF-2013 Benchmark. *Nat. Protoc.* **2018**, *13*, 666.

(66) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59*, 895.

(67) Cang, Z. X.; Mu, L.; Wei, G. W. Representability of Algebraic Topology for Biomolecules in Machine Learning Based Scoring and Virtual Screening. *PLoS Comput. Biol.* **2018**, *14* (1), No. e1005929.

(68) Cang, Z. X.; Wei, G. W. Integration of Element Specific Persistent Homology and Machine Learning for Protein-ligand Binding Affinity Prediction. *Int. J. Numer. Method. Biomed. Eng.* **2018**, *34* (2), e2914.

(69) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55*, 475–482.

(70) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.

(71) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(72) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499–511.

(73) Jain, A. N. Surflex-Dock 2.1: Robust Performance from Ligand Energetic Modeling, Ring Flexibility, and Knowledge-Based Search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.

(74) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(75) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.

(76) Li, G.-B.; Yang, L.-L.; Wang, W.-J.; Li, L.-L.; Yang, S.-Y. ID-Score: A New Empirical Scoring Function Based on a Comprehensive Set of Descriptors Related to Protein-Ligand Interactions. *J. Chem. Inf. Model.* **2013**, *53*, 592–600.

(77) Li, H.; Leung, K.-S.; Wong, M.-H.; Ballester, P. J. Improving AutoDock Vina Using Random Forest: The Growing Accuracy of Binding Affinity Prediction by the Effective Exploitation of Larger Data Sets. *Mol. Inf.* **2015**, *34*, 115–126.

(78) Nguyen, D. D.; Wei, G.-W. DG-GL: Differential Geometry-Based Geometric Learning of Molecular Datasets. *Int. J. Numer. Method. Biomed. Eng.* **2019**, *35*, No. e3179.

(79) Li, H.; Leung, K.-S.; Wong, M.-H.; Ballester, P. J. Low-Quality Structural and Interaction Data Improves Binding Affinity Prediction via Random Forest. *Molecules* **2015**, *20*, 10947–10962.

(80) Wójcikowski, M.; Kukielka, M.; Stepniewska-Dziubinska, M.; Siedlecki, P. Development of a Protein-Ligand Extended Connectivity (PLEC) Fingerprint and Its Application for Binding Affinity Predictions. *Bioinformatics* **2019**, *35*, 1334–1341.

(81) Jiménez, J.; Skalic, M.; Martínez-Rosell, G.; De Fabritiis, G. K. DEEP: Protein–Ligand Absolute Binding Affinity Prediction Via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287–296.

(82) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein-Ligand Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, 3666–3674.

(83) Plewczynski, D.; Łażniewski, M.; Augustyniak, R.; Ginalski, K. Can We Trust Docking Results? Evaluation of Seven Commonly Used Programs on PDBbind Database. *J. Comput. Chem.* **2011**, *32*, 742–755.

(84) Khamis, M. A.; Gomaa, W. Comparative Assessment of Machine-Learning Scoring Functions on PDBbind 2013. *Eng. Appl. Artif. Intell.* **2015**, *45*, 136–151.

(85) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461.