Biophysical Journal
Article

Biophysical Society

Further exploration of the quantitative distance-energy and contact number-energy relationships for predicting the binding affinity of protein-ligand complexes

Yong Xiao Yang[1,*] and Bao Ting Zhu[1,2,*]
[1]Shenzhen Key Laboratory of Steroid Drug Discovery and Development, School of Medicine, The Chinese University of Hong Kong, Shenzhen, Guangdong, China and [2]Shenzhen Bay Laboratory, Shenzhen, Guangdong, China

ABSTRACT Accurate estimation of the strength of the protein-ligand interaction is important in the field of drug discovery. The binding strength can be determined by using experimental binding affinity assays which are both time and labor consuming and costly. Predicting the binding affinity/energy in silico is an alternative approach, particularly for virtual screening of large data sets. In general, the distance-based terms such as electrostatic and van der Waals interactions are among the key determinants of binding energy. In this work, the distance-binding energy relationships, i.e., $E \propto -d^{-k}$, are further explored, extended, and developed for protein-ligand binding affinity prediction. The contributions of different atom-type pairs were considered synthetically and jointly. Additionally, the contact number-energy relationships ($E \propto -n^k$) were also explored for protein-ligand binding affinity prediction. Significantly, the power exponents of the distances or contact numbers in the energy functions are not restricted by the existing theories concerning van der Waals and electrostatic energies (expressed as $\frac{a}{r^6} - \frac{b}{r^{12}}$ and $\frac{c}{r}$). The performances of the new distance-based or contact number-based models are better than the performances of those sophisticated non-machine-learning-based scoring functions developed before. The exploration and extension of the distance-energy and contact number-energy relationships may offer insights into the development of more effective methods for predicting the protein-ligand binding affinity accurately and analyzing the protein-ligand interactions rationally.

SIGNIFICANCE In this work, the quantitative distance-binding energy relationships ($E \propto -d^{-k}$) and contact number-energy relationships ($E \propto -n^k$) are further explored anew and are developed for protein-ligand binding affinity prediction. The contributions of different atom-type pairs are considered synthetically, and the coefficients (parameters) of the distances and contact numbers with different power exponents are determined jointly. The ideas and strategies may be extended in the future for exploring other molecular interactions and the underlying theories governing physical interactions (especially the geometry-energy relationships).

INTRODUCTION

Biochemical experiment-based screening of bioactive compounds in the process of drug discovery is extremely time consuming and labor intensive (1). As an auxiliary means, computer-aided drug design (CADD) has been developed for nearly 50 years (2) since the first work about designing hemoglobin ligands, which was published in 1976 by Beddell et al. (3). One of the main issues in CADD is the accurate estimation of the strength of the protein-ligand interactions, i.e., calculating the binding energies/affinities for the protein-ligand complexes, which is directly related to the scoring power and ranking power of a scoring function for protein-ligand interactions (4). The scoring power refers to the ability to accurately predict the binding affinities of different protein-ligand complexes, and the ranking power refers to the ability to correctly rank different ligands with the same binding site of a protein based on the predicted binding affinities (4). Ideally, a perfect scoring function can predict the binding affinity nearly the same as the experimentally determined binding affinity for any protein-ligand complex. In practice, the non-perfect scoring functions may have relatively good performances (Pearson's correlation coefficient >0.7 or 0.8), but the ranking power often fluctuates dramatically for different protein targets,

Submitted November 12, 2024, and accepted for publication February 24, 2025.

*Correspondence: yangyongxiao@cuhk.edu.cn or btzhu@cuhk.edu.cn

Editor: Liang Hong.
https://doi.org/10.1016/j.bpj.2025.02.021

and the average ranking power of the practical scoring functions for these protein targets is weaker than expected.

The scoring functions can be categorized into four types: physics-based methods, knowledge-based potentials, empirical scoring functions, and descriptor-based scoring functions (5). In the scoring functions, the commonly used scoring terms or variables include van der Waals, electrostatic and hydrogen bonding energies, number of atom contacts and rotatable bonds, and others (6–9). Most of the energy terms are distance-based variables and in direct proportion to $-d^{-k}$ (d and k being the distance and power exponent, respectively) (6–8). The atomic contacts between the protein and ligand are also usually defined based on distance (9). In the contact number-based models, the binding energy (E) is in direct proportion to $-n^k$ (n and k being the contact number and power exponent, respectively) (9). The effectiveness of the distance-based or distance-related variables has been proved, but there is still room for improvement in the accuracy of predicting the protein-ligand binding affinity.

In this work, the quantitative distance-energy and contact number-energy relationships are further explored for the prediction of the binding energies/affinities in protein-ligand complexes. The new linear distance-based or contact number-based scoring functions are superior to the sophisticated non-machine-learning-based scoring functions developed previously.

MATERIALS AND METHODS

Data sets

The information about the protein-ligand complexes with known structures and experimental binding affinities were obtained from PDBbind (10,11) and CASF-2016 (4). In PDBbind v.2020 (https://www.pdbbind-plus.org.cn/) there are 19,443 protein-ligand complexes in the general set, which includes the 5316 complexes in the refined set. In CASF-2016 there are 285 structures of the protein-ligand complexes, and the proteins correspond to 57 targets and binding sites (4). All the experimental structures were downloaded anew from the Protein Data Bank (PDB) (https://www.rcsb.org/) (12). The ligands were defined based on the structural information from PDBbind (10,11) and CASF-2016 (4). The complexes were filtered based on four criteria: 1) there is a definite experimentally measured binding affinity, i.e., K_i or K_d value (K_i or K_d is the equilibrium dissociation constant for an inhibitor or a ligand, respectively); 2) the number of heavy atoms in the ligand is $\leq$70 and the total number of atoms in the ligand is $\leq$150; 3) no heavy atom clashes exist in the structure (all the distances between the heavy atom pairs in the complex are $>$1 Å); and 4) no unrecognized or unusual atom types (such as B, Du, Ni, Cd, As, Ir, Y, Yb, V, Ru, Te, Pr, Cs, Be, Sr, Rh, Re, Rb, Pt, In, Ga, Eu, or Au) exist in the ligand binding pocket of the structure. After filtering, there are 5171 protein-ligand complexes from the general set (*SET-1*), 4958 protein-ligand complexes from the refined set that are not included in the general set (*SET-2*), and 285 protein-ligand complexes in CASF-2016 (*SET-3*) (4).

Atom types and interacting pairs

The atom types in the structures of the protein-ligand complexes were assigned using Open Babel 2.4.1 (13). There are 33 heavy atom types in the binding pocket of the receptor (O.2, C.3, C.2, N.am, O.3, C.ar, O.co2, N.ar, N.pl3, S.3, C.cat, N.3, N.2, Zn, N.4, Mg, P.3, C.1, Ca, Cl, Mn, Na, S.O, Se, Fe, N.1, K, F, S.O2, Co.oh, S.2, I, and Br) and 25 heavy atom types in the ligand (C.3, C.ar, O.2, C.2, O.3, N.am, N.ar, N.3, N.pl3, O.co2, P.3, N.2, S.3, S.O2, Cl, F, S.2, C.1, C.cat, N.4, N.1, Br, I, Fe, and S.O). The definitions of the atom types are listed in Table S1. Two heavy atoms from the receptor and ligand separately are defined as an interacting pair, regarded as an atom contact if the distance between them is $\leq$10 Å and $>$1 Å. The interacting strength of a given atom pair will decay as the distance increases gradually. Here, the cutoff (10 Å) is adopted to reduce the computational amounts and define the binding pockets. For the standard amino acids, there are 18 heavy atom types (O.2, C.3, C.2, N.am, O.3, C.ar, O.co2, N.ar, N.pl3, S.3, C.cat, N.3, N.2, N.4, C.1, N.1, S.2, and S.O). The atom type of oxygen atom in the water molecule is O.3. If a pair is composed of two atoms with different atom types (atom type 1 and atom type 2), there would be two situations: atom type 1 from the receptor and atom type 2 from the ligand; and atom type 2 from the receptor and atom type 1 from the ligand. Here, the two situations are considered separately because the environments or backgrounds from the receptor and ligand are different.

To investigate the influence of different components of the receptor in the affinity prediction, the receptor was divided into three parts: standard amino acids, water molecules, and others (ions and other molecules). There are four situations when treating protein-ligand interactions: all the heavy atoms in the receptor are considered (situation S1); only the heavy atoms of standard amino acids in the receptor are considered (situation S2); the heavy atoms of standard amino acids and others in the receptor are considered (situation S3); and the heavy atoms of standard amino acids and water molecules in the receptor are considered (situation S4). There are 738 interacting atom-type pairs in situations S1 and S3, and 415 interacting atom-type pairs in situations S2 and S4.

Linear models for affinity prediction

The pre-linear models were generated based on distance-based or contact number-based variables using the least-square method (14). The binding energy (E) is in direct proportion to log(K), i.e., E = constant × log(K). The formation of the pre-linear models is expressed as follows:

$$\log(K) = a_1 \sum d_{1j}^{-k} + \ldots + a_i \sum d_{ij}^{-k} + \ldots a_m \sum d_{mj}^{-k} + b$$
$$k = 1, 2, \ldots, 18, \quad (1)$$

$$\log(K) = a_1 n_1^k + \ldots + a_i n_i^k + \ldots + a_m n_m^k + b$$
$$k = 1, 2, \ldots, 18. \quad (2)$$

Here, K is the inhibition or dissociation constant for a given protein-ligand complex, a_1, a_2, …, a_m are the coefficients, b is the constant term, d_{ij} is the distance of the j^{th} atom pair for the i^{th} atom-type pair, k is the power exponent, $\sum d_{ij}^{-k}$ is the sum of d_{ij}^{-k} of all the j^{th} atom pairs for the i^{th} atom-type pair, n_i is the contact number of the i^{th} atom-type pair, and m is the total number of atom-type pairs (m = 738 or 415). Therefore, there are 738 or 415 distance-based or contact number-based variables, respectively, in each linear model.

To explore the upper bound of the performance of the variables, there are seven kinds of subsets used to train the pre-linear models. In *SUBSET-1*, 4000 complexes are stochastically selected from *SET-1*; in *SUBSET-2*, 4000 complexes are stochastically selected from *SET-2*; in *SUBSET-3*, 200 complexes are stochastically selected from *SET-3*; *SUBSET-4* is composed of *SUBSET-1* and *SUBSET-2*; *SUBSET-5* is composed of *SUBSET-1* and *SUBSET-3*; *SUBSET-6* is composed of *SUBSET-2* and

SUBSET-3; and *SUBSET-7* is composed of *SUBSET-1*, *SUBSET-2* and *SUBSET-3*. The stochastic selection process was repeated 10,000 times. According to the affinity predictive performances in the three main sets, one representative distance-based (contact number-based) pre-linear model was selected from the 70,000 distance-based (contact number-based) pre-linear models for each situation (from S1 to S4) and power exponent (from 1 to 18).

A large number of new linear models was then generated based on 18 representative distance-based or contact number-based pre-linear models. The formations of the new linear models are expressed as follows:

$$\log(K) = a_1 dlm_1 + \ldots + a_i dlm_i + \ldots + a_n dlm_n + b$$
$$n = 1, 2, \ldots, 18 \quad (3)$$

$$\log(K) = a_1 nlm_1 + \ldots + a_i nlm_i + \ldots + a_n nlm_n + b$$
$$n = 1, 2, \ldots, 18 \quad (4)$$

Here, K is the equilibrium dissociation constant for a given protein-ligand complex, a_1, a_2, …, a_n are the coefficients, b is the constant term, and dlm_i and nlm_i are the i^{th} selected distance-based and contact number-based pre-linear models, respectively. Each new linear model contains no more than 18 selected pre-linear models. All the linear combinations of pre-linear models were considered, and the training sets were generated in the way used for the pre-linear models. The stochastic selection process was repeated 100 times. There were 183,500,100 ($(2^{18} - 1) \times 700$) linear models generated for affinity prediction for each situation. According to the performances in the three main sets, one representative distance-based (contact number-based) linear model was selected for each situation.

Metrics for evaluating the performance of the pre-linear and linear models

The scoring power of a given scoring function is frequently evaluated using the Pearson's correlation coefficient (R), which is calculated as follows (4):

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}. \quad (5)$$

Here, n is the number of protein-ligand complexes in a given set, x_i (or y_i) is the experimental (or predicted) binding affinity of the i^{th} complex, and $\overline{x}$ (or $\overline{y}$) is the average of the experimental (or predicted) binding affinities.

The ranking power is usually evaluated using the Spearman's rank correlation coefficient (ρ), which is expressed as follows (4,15):

$$\rho = \frac{\sum_{i=1}^{n}(rx_i - \overline{rx})(ry_i - \overline{ry})}{\sqrt{\sum_{i=1}^{n}(rx_i - \overline{rx})^2 \sum_{i=1}^{n}(ry_i - \overline{ry})^2}}. \quad (6)$$

Here, n is the number of protein-ligand complexes for a given protein target and binding site, rx_i (or ry_i) is the rank based on the experimental (or predicted) binding affinity of the i^{th} complex, and $\overline{rx}$ (or $\overline{ry}$) is the average rank based on the experimental (or predicted) binding affinities. There are 57 protein targets in CASF-2016, and five ligands bind to the same binding site in each target (4). The average Spearman's rank correlation coefficient (ρ) for the 57 targets is used to assess the ranking power of a given scoring function.

RESULTS

The procedure of this research is shown in Fig. 1, which includes the following general steps: 1) filtering the data obtained from PDBbind (10,11) and CASF-2016 (4); 2) assigning the atom types using Open Babel 2.4.1 (13); 3) calculating the distance-based and contact number-based variables; and 4) training the pre-linear and linear models for protein-ligand affinity prediction.

Performances of the distance-based and contact number-based variables for different atom-type pairs in affinity prediction

Without any prior training, the basic distance-based and contact number-based variables, i.e., $-\sum {d_{ij}}^{-k}$ in Eq. 1, $-\sum\sum {d_{ij}}^{-k}$ (negative sum of $\sum {d_{ij}}^{-k}$ for all the i^{th} atom-type pairs), $-{n_i}^k$ in Eq. 2, and $-\sum {n_i}^k$ (negative sum of ${n_i}^k$ for all the i^{th} atom-type pairs), were directly employed to predict the binding affinity. The scoring power and ranking power of these variables were evaluated using Pearson's correlation coefficient (R) and the average Spearman's rank correlation coefficient (ρ).

The performances of these variables are stored in Data S1 and S2. There are several main points worthy of mentioning.

1) Some variables for single atom-type pair ($-\sum {d_{ij}}^{-k}$ in Eq. 1, $-{n_i}^k$ in Eq. 2) have relatively better performances ($R \geq 0.30$ and average $\rho \geq 0.30$ in CASF-2016 (*SET-3*)). The atom-type pairs include ligand C.3-receptor O.2, ligand C.3-receptor C.3, ligand C.3-receptor C.2, and ligand C.3-receptor N.am.
2) For the two types of variables, i.e., $-\sum\sum {d_{ij}}^{-k}$ and $-\sum {n_i}^k$, the predictive power decreases as the power exponent (k) increases (Fig. 2). The best performances are $R \approx 0.39$ in *SET-1* (5171 complexes) (Fig. 2 *A*), $R \approx 0.51$ in *SET-2* (4958 complexes) (Fig. 2 *B*), $R \approx 0.63$ in *SET-3* (CASF-2016, 285 complexes) (Fig. 2 *C*), and average $\rho \approx 0.60$ in CASF-2016 (Fig. 2 *D*). They can be achieved by both the distance-based and contact number-based variables in all four situations (S1–S4) when the power exponent (k) is equal to 1 ($-\sum\sum {d_{ij}}^{-1}$ and $-\sum n_i$). There are no apparent differences between the contact number-based variables ($-\sum {n_i}^k$) for different situations (S1–S4) (Fig. 2). The performances of the contact number-based variables ($-\sum {n_i}^k$) decrease more dramatically than those of the distance-based variables when the power exponents (k) are small (Fig. 2). The ranking power of the contact number-based variables ($-\sum {n_i}^k$) are not changed when the power exponents (k) are ≥ 4, whereas the average ρ of the distance-based variables ($-\sum\sum {d_{ij}}^{-k}$) in CASF-2016 still decreases (Fig. 2 *D*; the standard deviations are too big and not shown in the figure). With the increase of the power exponent (k), the differences between the contact number-based variables ($-\sum {n_i}^k$) of different

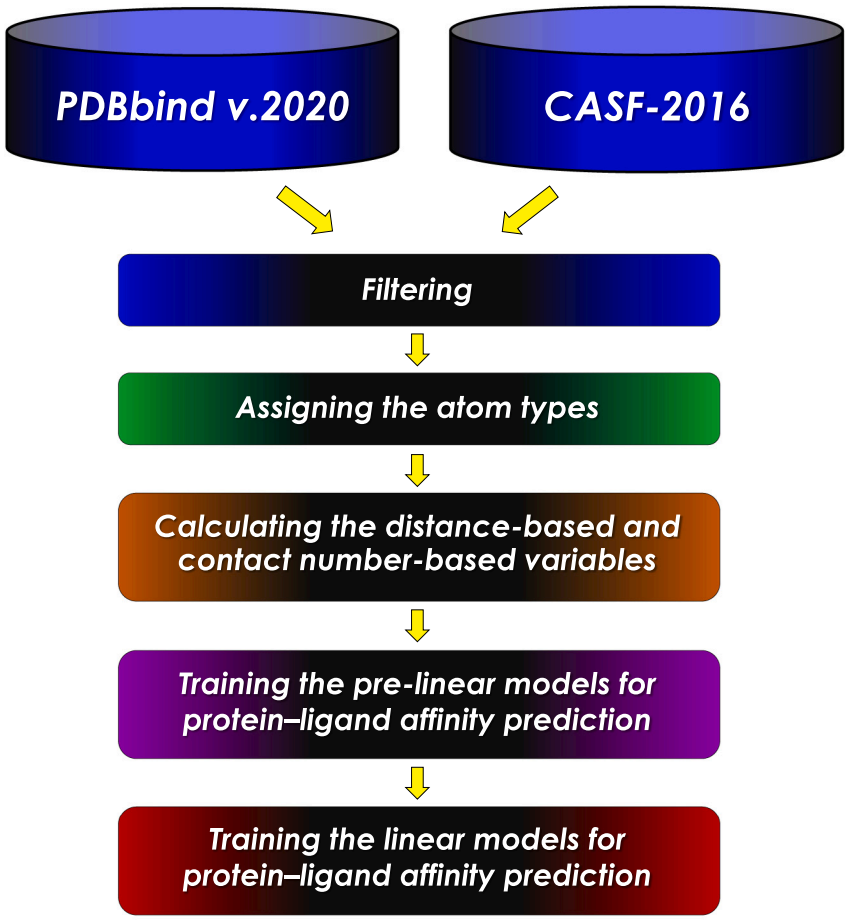

FIGURE 1 Flowchart illustrating the general concept and approach in exploring the protein-ligand binding affinity prediction. The steps include filtering the data, assigning the atom types, calculating the variables, and training pre-linear and linear models.

complexes also increase, but the differences between the distance-based variables ($-\sum\sum {d_{ij}}^{-k}$) of different complexes decrease. While increased difference in contact number-based variables ($-\sum {n_i}^{k}$) does not affect the ranking, the net effect of a reduced difference in distance-based variables ($-\sum\sum {d_{ij}}^{-k}$) does change the ranking.

3) The best performances of the two types of variables, i.e., $-\sum\sum {d_{ij}}^{-k}$ and $-\sum {n_i}^{k}$, are comparable to some of the sophisticated traditional non-machine-learning scoring functions (in the CASF-2016 data set (4)) according to the reported results in previous works (7,8).

When calculating the two types of variables ($-\sum\sum {d_{ij}}^{-k}$, $k = 1, 2, \ldots, 18$; $-\sum {n_i}^{k}$, $k = 1$) without discrimination of different atom-type pairs, it is not necessary to assign the atom types. Hence, it is a faster approach to compare the relative binding affinities of the protein-ligand complexes in a given set using the two types of variables when no trained linear or machine-learning-based non-linear models are available. In particular, the two variables, i.e., $-\sum\sum {d_{ij}}^{-1}$ (negative sum of the reciprocals of the distances of all the interacting atom pairs between the receptor and ligand) and $-\sum n_i$ (negative sum of the atom contact numbers between the receptor and ligand) in situation S2, are recommended to be used in practice. It should be noted that in situation S2, only the heavy atoms of standard amino acids in the receptor are considered, and there are 415 atom-type pairs between the receptor and ligand within a distance cutoff of 10 Å.

Performances of the representative pre-linear models in affinity prediction

Several hundred thousand pre-linear models were generated based on the distance-based and contact number-based variables. According to the performances, one representative pre-linear model was selected for each power exponent, each type of variable, and each situation. The coefficients and constant terms of the selected pre-linear models are stored in Data S3 and S4, and the scoring and ranking powers of the selected pre-linear models are shown in Fig. 3 (Data S5). There are some contact number-based pre-linear models for which the coefficients are set to -1 and the constant term are set at 0 because of the weaker performances of the trained pre-linear models based on the statistically selected training data (Data S4). The best performances of the pre-linear models are $R \approx 0.60$ in *SET-1* (5171 complexes) (Fig. 3 *A*), $R \approx 0.66$ in *SET-2*

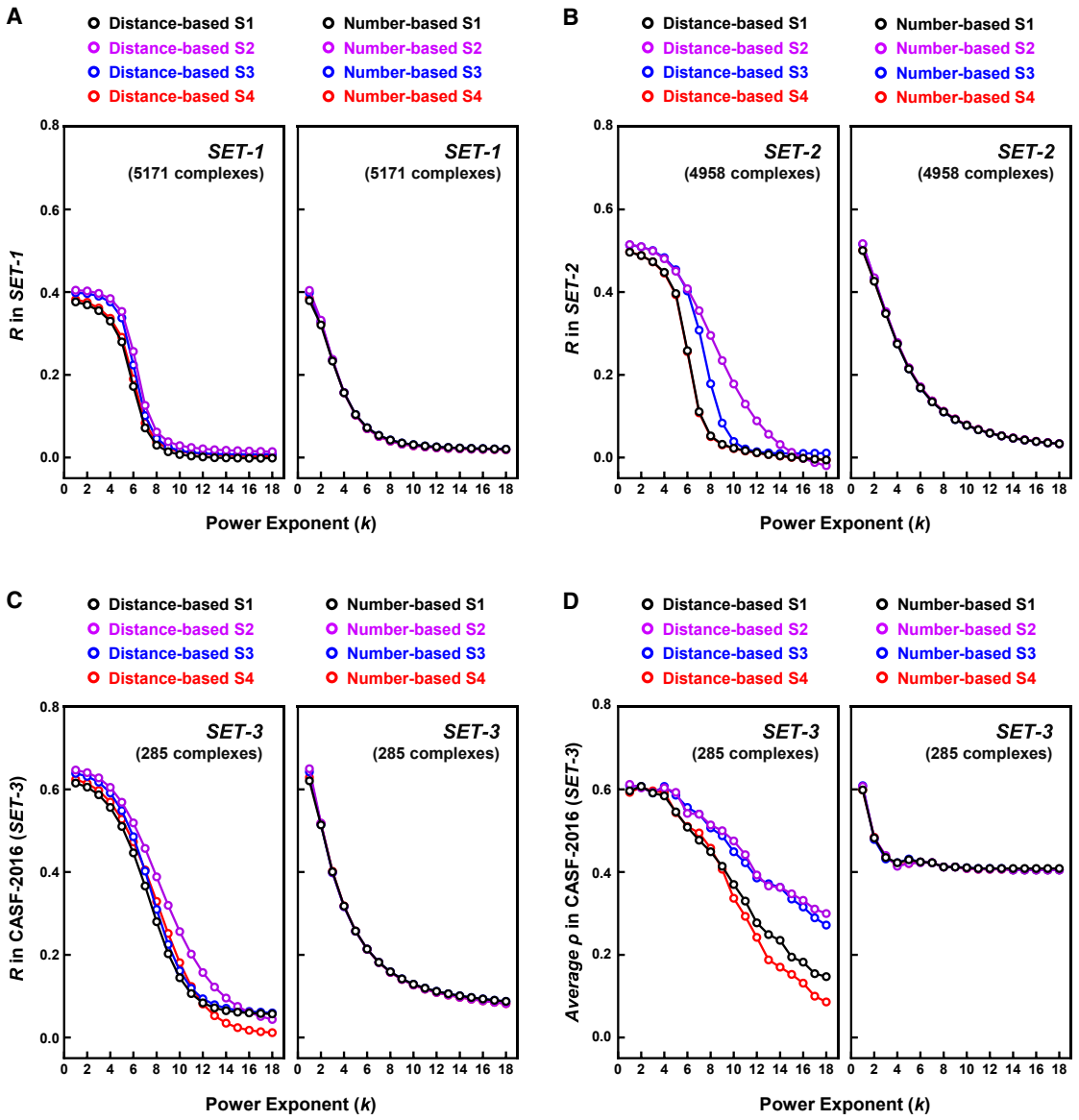

FIGURE 2 Performances of two types of distance-based and contact number-based variables ($-\sum\sum d_{ij}^{-k}$ and $-\sum n_i^k$) for different power exponents in affinity prediction without any training. (*A*) *R* (Pearson's correlation coefficient) in *SET-1* (5171 complexes). (*B*) *R* in *SET-2* (4958 complexes). (*C*) *R* in *SET-3* (CASF-2016) (285 complexes). (*D*) Average ρ (Spearman's rank correlation coefficient) in *SET-3* (CASF-2016) (57 targets, 285 complexes). S1–S4 refer to situations S1–S4, respectively.

(4958 complexes) (Fig. 3 *B*), $R \approx 0.73$ in *SET-3* (CASF-2016, 285 complexes) (Fig. 3 *C*), and average $\rho \approx 0.57$ in CASF-2016 (Fig. 3 *D*). Except for the contact number-based pre-linear models in situations 1 and 3, the scoring power can be achieved by all the other types of pre-linear models with adjusted power exponents. Although the scoring powers are improved significantly with the use of the linear regression technique ($R \approx 0.39 \rightarrow 0.60$ in *SET-1*; $R \approx 0.51 \rightarrow 0.66$ in *SET-2*; $R \approx 0.63 \rightarrow 0.73$ in *SET-3* (CASF-2016)), the ranking powers decrease slightly (average $\rho \approx 0.60 \rightarrow 0.57$ in CASF-2016). In view of the different improvements in scoring power and ranking power when the linear regression technique is used, it is advised that more emphasis should be considered for improving the ranking power, as improvements in the ranking often are of practical utility.

Overall, the predictive powers of the pre-linear models decrease when the power exponents (k) increase (Fig. 3). The performances of the contact number-based pre-linear models appear to decrease more rapidly than the performances of the distance-based pre-linear models when the values of the power exponents (k) are small (Fig. 2). With the power exponent (k) increasing from 1 to 5, the distance-based variables are more robust and useful than the contact number-based variables to evaluate the interacting strength for affinity prediction.

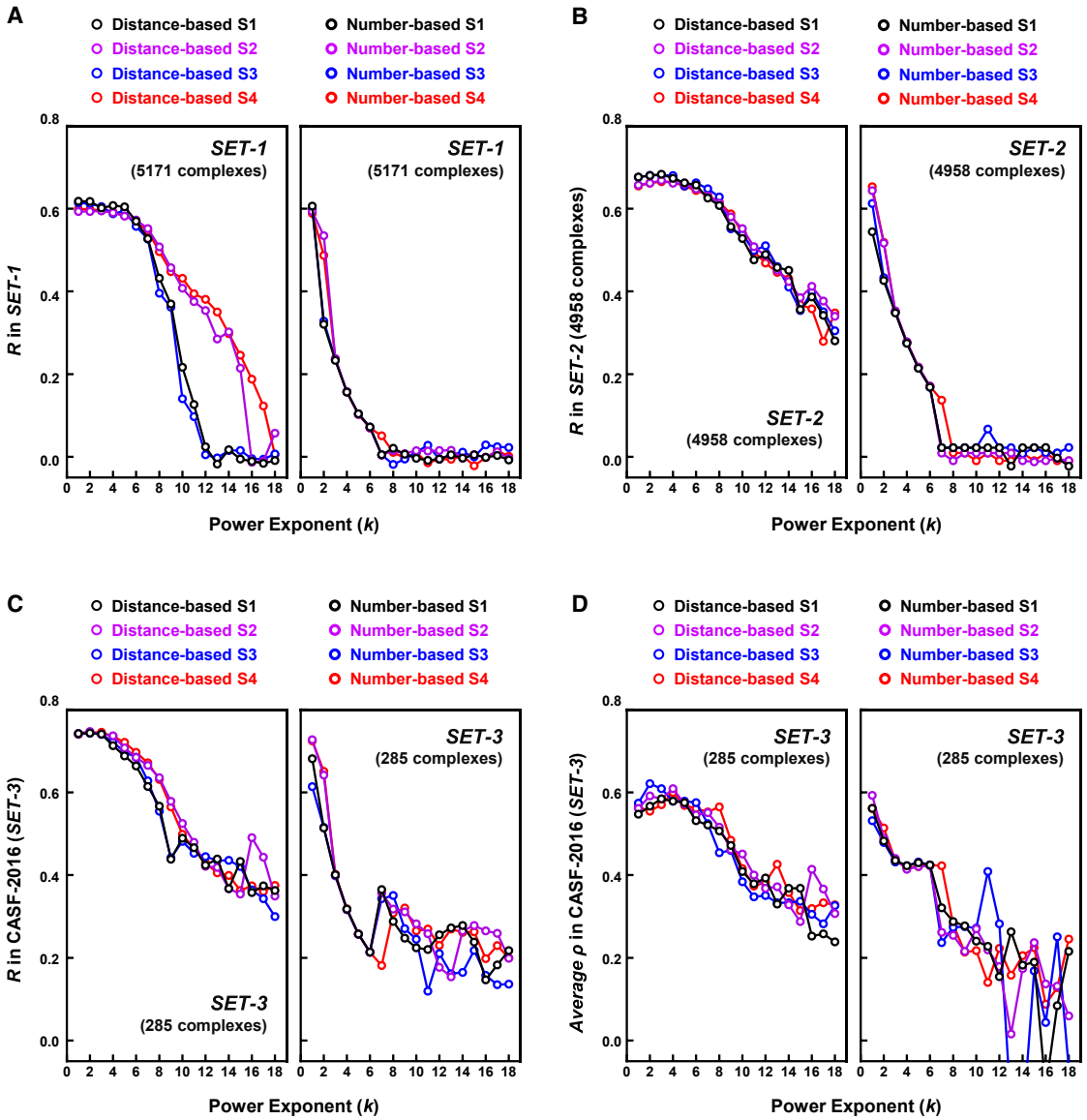

FIGURE 3 Performances of the representative pre-linear models in affinity prediction. (*A*) *R* (Pearson's correlation coefficient) in *SET-1* (5171 complexes). (*B*) *R* in *SET-2* (4958 complexes). (*C*) *R* in *SET-3* (CASF-2016) (285 complexes). (*D*) Average ρ (Spearman's rank correlation coefficient) in *SET-3* (CASF-2016) (57 targets, 285 complexes). S1–S4 refer to situations S1–S4, respectively.

Performances of the representative linear models in affinity prediction

To further improve their performance, a large number of linear models were trained using the selected pre-linear models. One representative linear model was selected for each type of variable and each situation. The coefficients and constant terms of the linear models are stored in Data S6. The performances of the representative linear models are shown in Table 1 (Data S5).

On average, from the single variable ($-\sum\sum d_{ij}^{-1}$ or $-\sum n_i$) to the pre-linear models, then to the linear models, the best performances are: $R \approx 0.39 \rightarrow 0.60 \rightarrow 0.61$ in *SET-1*; $R \approx 0.51 \rightarrow 0.66 \rightarrow 0.67$ in *SET-2*; $R \approx 0.63 \rightarrow 0.73 \rightarrow 0.73$ in *SET-3* (CASF-2016); and average $\rho \approx 0.60 \rightarrow 0.57 \rightarrow 0.60$ in CASF-2016. Unfortunately, the performances were not markedly improved as expected from the pre-linear models to the linear models. Even so, there are still two linear models with relatively good performances, namely the linear models DS3 and NS2. For the linear model DS3, $R = 0.64$ in *SET-1*, $R = 0.70$ in *SET-2*, $R = 0.75$ in *SET-3* (CASF-2016), and $\rho = 0.62 \pm 0.36$ in CASF-2016; and for the linear model NS2, $R = 0.60$ in *SET-1*, $R = 0.65$ in *SET-2*, $R = 0.73$ in *SET-3* (CASF-2016), and $\rho = 0.63 \pm 0.40$ in CASF-2016. Although the performances have not caught up with those of the machine-learning-based methods, the practical application value and explainable ability of the

TABLE 1 Performances of the representative linear models for affinity prediction

	Scoring power (Pearson's correlation coefficient, R)				Ranking power (average Spearman's rank correlation coefficient, ρ)
Model name	R in *SET-1* (5171 complexes)	R in *SET-2* (4958 complexes)	R in *SET-3* (285 complexes)	Mean R ± std. R in the three sets	Average ρ ± std. ρ for the 57 targets in CASF-2016 (*SET-3*)
Linear model DS1	0.64	0.70	0.75	0.70 ± 0.06	0.58 ± 0.38
Linear model DS2	0.61	0.68	0.75	0.68 ± 0.07	0.59 ± 0.43
Linear model DS3	0.64	0.70	0.75	0.70 ± 0.06	0.62 ± 0.36
Linear model DS4	0.61	0.67	0.75	0.68 ± 0.07	0.58 ± 0.47
Linear model NS1	0.61	0.67	0.70	0.66 ± 0.05	0.60 ± 0.41
Linear model NS2	0.60	0.65	0.73	0.66 ± 0.07	0.63 ± 0.40
Linear model NS3	0.60	0.65	0.64	0.63 ± 0.02	0.57 ± 0.46
Linear model NS4	0.59	0.65	0.73	0.66 ± 0.07	0.58 ± 0.41

non-machine-learning-based scoring functions are still noteworthy and substantial.

Comparison with other non-machine-learning-based scoring functions for protein-ligand binding affinity prediction

In the past 30 years, many scoring functions have been developed for protein-ligand binding affinity prediction (4–9). The distance-based and contact number-based linear models were compared with several sophisticated non-machine-learning-based methods such as X-Score (6), APBScore (7), AA-Score (8), and PRODIGY-LIG (9). Three methods, i.e., X-Score (6), APBScore (7) and AA-Score (8), involve distance-based energy terms, and PRODIGY-LIG (9) employs contact number-based terms. The binding energies in the protein-ligand complexes used in this work were calculated using X-Score (6). The binding affinities of the protein-ligand complexes in CASF-2016 (*SET-3*) were also predicted using PRODIGY-LIG (9). The performances of APBScore (7) and AA-Score (8) were taken from the original studies (7,8) for comparison.

Overall, the scoring powers of the linear models (linear model DS3: $R = 0.64$ in *SET-1*, $R = 0.70$ in *SET-2*, $R = 0.75$ in *SET-3* (CASF-2016), and $\rho = 0.62 \pm 0.36$ in CASF-2016); linear model NS2: $R = 0.60$ in *SET-1*, $R = 0.65$ in *SET-2*, $R = 0.73$ in *SET-3* (CASF-2016), and $\rho = 0.63 \pm 0.40$ in CASF-2016) are better than those of the other four sophisticated non-machine-learning-based methods, and the ranking powers are comparable to that of the best one of them (Fig. 4). The performances of X-Score (6) are $R = 0.49$ in *SET-1* (Fig. 4 *A*), $R = 0.54$ in *SET-2* (Fig. 4 *B*), $R = 0.64$ in CASF-2016 (*SET-3*) (Fig. 4 *C*), and $\rho = 0.62 \pm 0.39$ in CASF-2016 (Fig. 4 *D*). The performances of the other three methods in CASF-2016 are $R = 0.67$ and average $\rho = 0.60$ for APBScore (7), $R = 0.69$ and average $\rho = 0.62$ for AA-Score (8), and $R = 0.53$ and average $\rho = 0.46$ for PRODIGY-LIG (9). To more conveniently compare the models of this work with the other scoring functions that may be developed in the future, the values of the negative sum ($-\sum\sum d_{ij}^{-1}$, $-\sum n_i$) and the predicted values of the representative pre-linear and linear models X-Score (6) and PRODIGY-LIG (9) are stored in Data S7.

DISCUSSION

Here, we briefly discuss some of the factors that affect the overall ability of the computational methods for predicting the protein-ligand binding affinity, which includes the quantitative X-affinity/energy relationships, the relationship between the protein-ligand binding affinity and the binding process, and the influence of the data-set quality and completeness on the predictive models being developed. Here, the X refers to the basic variables (such as distance, area, contact number, and number of rotatable bonds) or energy terms (such as van der Waals, electrostatic, and hydrogen bonding energies).

Quantitative X-affinity/energy relationships

In certain ways, the structure of a protein-ligand complex can be viewed mostly as geometric in nature, and the geometric structures in biochemical environments are reflections of energy adaption for protein-ligand interaction. The commonly used basic variables in the structure-based scoring functions include distance, area, and atomic contact numbers (4,6–9). The traditional non-covalent interacting energy terms such as van der Waals and electrostatic interactions were usually evaluated based on the variable distances (16). The contact numbers of different atom-type pairs (i.e., hydrophobic, hydrophilic, polar, and non-polar) were also frequently defined according to the distance cutoffs (4,9,17). The solvent-accessible surface areas and the buried polar surface areas between a protein and ligand were adopted in several scoring functions (6,18–20). Nearly all these basic variables are geometric in nature. Figuring out the geometry-energy relationship remains a basic issue in protein chemistry and biochemistry. At the core of the binding affinity/energy prediction essentially are the quantitative X-affinity/energy relationships as presented in the scoring functions. It is, therefore, important to find the

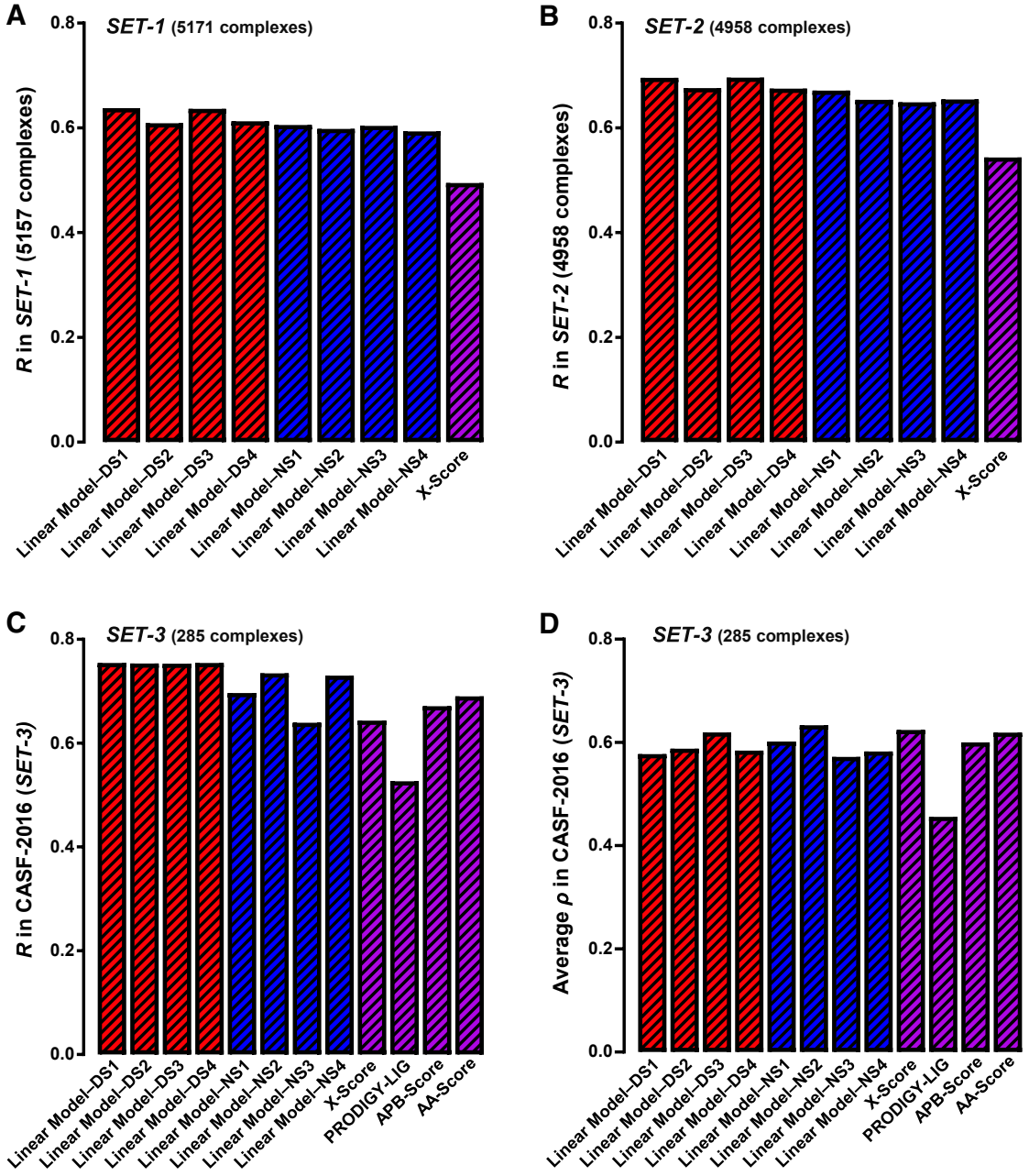

FIGURE 4 Performances of the representative linear models and other sophisticated non-machine-learning scoring functions in affinity prediction. (*A*) *R* (Pearson's correlation coefficient) in *SET-1* (5171 complexes). (*B*) *R* in *SET-2* (4958 complexes). (*C*) *R* in *SET-3* (CASF-2016) (285 complexes). (*D*) Average ρ (Spearman's rank correlation coefficient) in *SET-3* (CASF-2016) (57 targets, 285 complexes). S1–S4 refer to the situations S1–S4, respectively; DS1–DS4 are the distance-based linear models in situations S1–S4, respectively; NS1–NS4 are the contact number-based linear models in situations S1–S4, respectively.

appropriate formation of scoring functions and select the proper training methods to determine the coefficients and constants in the scoring functions.

Some of the traditional scoring functions employ a single class of variables, such as distance, contact numbers, and areas. For example, a traditional physics-based scoring function (D-Score@SYBYL) was used in early versions of DOCK (21). It is the sum of the van der Waals and electrostatic interaction energies between a protein and a ligand. The formations of van der Waals and electrostatic energies are expressed as $\frac{a}{r^6} - \frac{b}{r^{12}}$ and $\frac{c}{r}$, respectively. Here, r is the distance between the two atoms, and a, b, and c are constants. The distance-energy relationship is reflected in the scoring function composed of distance-based energy terms. In the "no-electrostatics protocol" of PRODIGY-LIG (9), the binding energy ($\Delta G_{\mathrm{noelec}}$) in the protein-ligand complex is a linear combination of the number of atomic contacts. It reflects the contact number-energy relationships. The

present work sought to further explore the distance-energy and contact number-energy relationships. The power exponents (k) are beyond those used in the traditional formations. Overall, the predictive powers of the pre-linear models are decreased with the increase of power exponents (k) (Fig. 3). The performances of the contact number-based pre-linear models appear to decrease more rapidly than the performances of the distance-based pre-linear models when the values of the power exponents (k) are small (Fig. 2). With the increase of the power exponents (k) from 1 to 5, the distance-based variables are robust and useful for evaluating the interaction strength for affinity prediction compared to the contact number-based variables. Here, it is of note that the information of interaction implied in the distance-based variables is quite different from the information implied in contact number-based variables. The contact number is defined on the basis of a distance cutoff, which discounts the varying contributions of different distances within the cutoff. It is perhaps the affluent information inherent to the distance-based variables that make the performances decrease more slowly when the power exponents (k) are small. Given the differential improvements in scoring power versus ranking power with the linear regression techniques (Figs. 2 and 3), it is advised that future emphasis should be placed on improving the ranking power, as this improvement often is of more practical utility.

Additionally, the area-energy relationship for protein-ligand interactions has not been explored in detail in the past. The buried solvent-accessible surface areas of the ligand upon binding (ΔSAS) are effective in protein-ligand binding affinity prediction ($R = 0.63$ and average $\rho = 0.59$ in CASF-2016) (4). Zhu et al. found that the buried solvent-accessible surface areas (SASAs) are better correlated with the binding affinities in a given set, and the performances of the other complicated methods would be better in the same set (22). In PDBbind (10,11), Pearson's correlation coefficient (R) between the buried SASA and the experimentally determined binding affinities is 0.39 (22). The area-energy relationship for protein-protein interactions were tentatively explored in our previous works about *AREA-AFFINITY* (23–25). In the future, new area-based methods may be developed for protein-ligand binding affinity prediction.

Explainability, generalization ability, and practical usefulness

It should be noted that the linear models with explicit mathematical expressions developed in this study are just preliminary explorations beyond the power exponents adopted in the classical energy terms such as the electrostatic and van der Waals energies. These preliminary explorations are mostly intended to help shed light on the potential underlying mechanisms of the predictive ability. The linear models developed here may have an inherent potential of improved explainability, as these linear models have relatively simple, explicit mathematical expressions, which may help explain the principles and mechanisms associated with the predictive ability. In comparison, the existing machine-learning-based scoring functions are usually not associated with explicit mathematical expressions and are, therefore, often viewed as more opaque in mechanistic terms.

The linear models developed in this study also have comparatively good practical values with acceptable predictive powers. The best linear models developed in this study are actually superior to the highly sophisticated non-machine-learning-based scoring functions developed previously. Notably, these traditional non-machine-learning-based models are still very commonly used in present-day virtual drug screening (7,8). As such, the best linear models developed in this study may be useful in virtual drug screening as a supplementary tool to the existing non-machine-learning-based scoring functions.

We find that the predictive ability of the existing machine-learning models is better than that of the linear models developed in this study. It is known that the existing machine-learning-based scoring functions usually exhibit better performances with the data sets adopted in the original studies, and they may not perform as well with new data sets (22). In real-world applications, the universality or generalization ability of a predictive model in various new data sets is crucial. Comparatively speaking, the predictive ability (performances) of the linear models with simple, explicit mathematical expressions tends to be less affected by data sets, as these models are based more on the underlying principles/mechanisms and thus may have an inherent potential to perform in a data-set-independent fashion. Based on the insights gained from this study, it is hoped that in the future, the performances of the linear models (with explicit mathematical expressions) may be further improved when higher-quality training data sets become available and when more exhaustive explorations are adopted.

When a model is more complex, its generalization ability on new targets often is reduced (22). The artificial intelligence models are very complex and can predict the protein-ligand binding affinity accurately based on the ligands only (26). Different protein-ligand complexes have different ligand binding conformations and poses. The shapes of the ligand conformations and the ligand binding poses in the binding pocket can partially reflect the surrounding environment and the interactions between a protein and ligand. In the pre-linear models of this work, there are 738 or 415 distance-based or contact number-based variables. In the final linear models, there are thousands of variables. A larger number of variables make the linear models more complicated. How to determine the formations (appropriate power exponents) and coefficients for each atom-type pair in the whole background and environment of protein-ligand interactions is a critical problem. Developing effective methods to solve this problem may represent a research direction in

the field of developing more accurate computational methods to predict protein-ligand interactions. Efficient integration of the interacting energies of all the atom-type pairs with effective formations and coefficients may present a promising approach to improving the performances in protein-ligand binding affinity prediction.

The relatively strong performances of the linear models DS3 and NS2 developed in this study likely are attributed to three aspects: 1) the extension of the power exponents beyond those used in the traditional energy terms; 2) the different interatomic interactions in the binding pocket; and 3) concomitant determinations of the coefficients (parameters) in the scoring functions, which may help take into joint consideration the multibody interactions between atoms in the affinity models. Additionally, we noted that the differences between the performances of the best linear models are actually rather small. The influences of solvation (water) and metal ions may be further explored in the future when a sufficient quantity of high-quality data becomes readily available.

Lastly, it is of note that for prediction of the protein-ligand binding affinity based on one-snapshot structures, the computational speed is generally very fast and the cost is usually not very expensive for both machine-learning-based and non-machine-learning-based scoring functions. After simple testing, we noted that prediction of the binding affinity for a protein-ligand complex only takes ∼1.5 s for X-Score (6), ∼2 s for PRODIGY-LIG (9), ∼14 s for $\Delta_{vina}RF_{20}$ (a machine-learning-based method) (27), and ∼4 s (on average) for each linear scoring function developed in this study (Data S8).

Relationship between protein-ligand binding affinity and the binding process

Protein-ligand binding is a biochemical process, during which the protein-ligand interactions may have many different conformations. Molecular dynamics (MD) simulation can sample numerous conformations and simulate the protein-ligand binding process. MD-based binding energy calculation methods include molecular mechanics-Poisson Boltzmann/generalized Born surface area (MM-PB/GBSA) (28,29), free energy perturbation (FEP) (30,31), thermodynamic integration (TI) (32–34), and amino-acid-specific interaction entropy (ASIE) (35–37). While these methods can calculate the binding energy with relatively high accuracy, they usually consume a large amount of computational resources. When virtual screening in a library with a large number of ligands, these methods are not practical compared to the one-snapshot-based scoring functions. The snapshot presented in the experimental structure is likely the most stable conformation during the protein-ligand binding process. Additionally, one-snapshot-based scoring functions can also be employed to calculate the binding energies of the MD-generated conformations.

There are three widely used conceptual models for describing protein-ligand binding interactions (38). These are the key-lock theory (39), induced-fit theory (40), and conformational selection theory (41,42). The key-lock theory underlines the geometric shape complementarity between the protein and ligand. The induced-fit theory emphasizes the co-adaptation and co-adjustment between the protein and ligand. The conformational selection is that the ligand selects the more appropriate conformation(s) to bind from the existing multiple conformations of the receptor. These conceptual models play different roles in the structural prediction of the protein-ligand complexes. Their roles in affinity prediction are still not adequately considered in the present study and may need to be explored in the future.

Data quality and completeness

It is clear that the quality and completeness of the available data sets have a significant influence on the development of scoring functions with practical values and also on improving the theory on protein-ligand interactions. According to the descriptions provided for PDBbind (10,11) and CASF-2016 (4), the order of the data quality in the three main sets are CASF-2016 > PDBbind refined set > PDBbind general set. In this work, the data in *SET-1* and *SET-2* are from the PDBbind general set and refined set, respectively. *SET-3* is the CASF-2016. Overall, the best performances of the "negative sum" ($-\sum\sum d_{ij}^{-1}$ or $-\sum n_i$), the pre-linear models, and the linear models are $R \approx 0.39 \rightarrow 0.60 \rightarrow 0.61$ in *SET-1*, $R \approx 0.51 \rightarrow 0.66 \rightarrow 0.67$ in *SET-2*, and $R \approx 0.63 \rightarrow 0.73 \rightarrow 0.73$ in *SET-3*. As expected, the data quality (i.e., the structural quality and the reliability of the experimental binding affinity data) affects the predictive power, and higher-quality data correspond to better performances of the developed models. Additionally, no single data set (*SET-1*, *SET-2*, or *SET-3*) contains all the atom-type pairs. Based on the available experimental structures and binding affinity data, it is recommended that the data sets should be as complete as possible when training the effective models. In this study, the training data sets for the pre-linear and linear models were stochastically selected from the three sets, and the process was repeated many times to improve the completeness of the training data sets in terms of the atom-type pairs, the structural diversity, and the range of the affinity values.

Even though the data in CASF-2016 (4) have higher quality, there are also several protein-ligand complexes with potentially confusing information. In some complexes, there are multiple ligands with the same kind in the neighboring regions of the binding pocket in the experimental structures (Fig. S1, *A*–*J*). The positions of the ligands defined in CASF-2016 may be not reasonable for three protein-ligand complexes (Fig. S1, *J*–*L*). In several cases, there are multiple binding pockets with the same ligands (Fig. S1, *I*, *M*, and *N*) because of the high degree of symmetry of the structures

or the promiscuous specificity of the protein-ligand interactions (Fig. S1 *N*). In another two cases, there is an additional ligand bound to the surface of the protein (distant from the binding pocket) (Fig. S1, *O* and *P*). Accordingly, the existence of multiple ligands may affect the binding affinities. The lack of adequate consideration and treatments of the multiple ligands would be a confounding factor that would make the training or prediction less accurate.

Although a huge amount of experimental data on the protein structures and protein-ligand binding affinities have been accumulated in the past several decades, the availability of large sets of high-quality data with complete and necessary information remains an unrealistic demand at present. The availability of these high-quality data sets would be essential to developing computational methods that can accurately predict protein-ligand binding affinities.

CONCLUSIONS

In this study, the quantitative distance-energy and contact number-energy relationships ($E \propto -d^{-k}$ and $E \propto -n^{k}$) were further explored and extended toward improving the ability to predict the binding affinity in protein-ligand complexes. The new linear models are superior to the traditional physics-based or experience-based scoring functions. It is hoped that in the future when the quality of the structure and binding affinity data of the protein-ligand complexes improves and the explicit mathematical expressions of the linear models are exhaustively explored, the performances of affinity prediction will further improve accordingly through the use of strategies and protocols developed in this work. The interpretation of each term in the mathematical expressions can be derived against the background of the protein-ligand interactions at the atomic level. While the potential effectiveness of the linear models has been validated in the present preliminary explorations, further exploration and extension of the quantitative distance-energy and contact number-energy relationships may offer new insights for future investigations of the molecular interactions. It is expected that the affinity predictive powers of the distance or contact number variables will also be explored in the future by employing powerful artificial neural networks.

DATA AND CODE AVAILABILITY

All data of the study are described and presented in the results section.

ACKNOWLEDGMENTS

This work is supported by research grants from Shenzhen Peacock Plan (no. KQTD2016053117035204), Shenzhen Key Laboratory of Steroid Drug Discovery and Development Research (no. ZDSYS20190902093417963), Shenzhen Bay Laboratory (no. SZBL2019062801007), 2022 Stable Funding Support Program for Shenzhen Institutions of High Learning, and Longgang District Science and Technology Bureau's Key Laboratory Program.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.X.Y. and B.T.Z.; methodology, Y.X.Y.; formal analysis, Y.X.Y. and B.T.Z.; investigation, Y.X.Y. and B.T.Z.; resources, B.T.Z.; data curation, Y.X.Y.; writing –original draft, Y.X.Y. and B.T.Z.; writing – review & editing, Y.X.Y. and B.T.Z.; visualization, Y.X.Y.; supervision, B.T.Z.; project administration, B.T.Z.; funding acquisition, B.T.Z.

DECLARATION OF INTERESTS

The first author has a patent application and/or registration related to this work.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2025.02.021.

REFERENCES

1. DiMasi, J. A., H. G. Grabowski, and R. W. Hansen. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* 47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012.
2. Blaney, J. 2012. A very short history of structure-based design: how did we get here and where do we need to go? *J. Comput. Aided Mol. Des.* 26:13–14. https://doi.org/10.1007/s10822-011-9518-x.
3. Beddell, C. R., P. J. Goodford, …, R. Wootton. 1976. Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.* 57:201–209. https://doi.org/10.1111/j.1476-5381.1976.tb07468.x.
4. Su, M., Q. Yang, …, R. Wang. 2019. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* 59:895–913. https://doi.org/10.1021/acs.jcim.8b00545.
5. Liu, J., and R. Wang. 2015. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* 55:475–482. https://doi.org/10.1021/ci500731a.
6. Wang, R., L. Lai, and S. Wang. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* 16:11–26. https://doi.org/10.1023/A:1016357811882.
7. Bao, J., X. He, and J. Z. H. Zhang. 2020. Development of a New Scoring Function for Virtual Screening: APBScore. *J. Chem. Inf. Model.* 60:6355–6365. https://doi.org/10.1021/acs.jcim.0c00474.
8. Pan, X., H. Wang, …, J. Z. H. Zhang. 2022. AA-Score: a New Scoring Function Based on Amino Acid-Specific Interaction for Molecular Docking. *J. Chem. Inf. Model.* 62:2499–2509. https://doi.org/10.1021/acs.jcim.1c01537.
9. Vangone, A., J. Schaarschmidt, …, A. M. J. J. Bonvin. 2019. Large-scale prediction of binding affinity in protein–small ligand complexes: the PRODIGY-LIG web server. *Bioinformatics.* 35:1585–1587. https://doi.org/10.1093/bioinformatics/bty816.
10. Wang, R., X. Fang, …, S. Wang. 2004. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* 47:2977–2980. https://doi.org/10.1021/jm030580l.
11. Liu, Z., M. Su, …, R. Wang. 2017. Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions. *Acc. Chem. Res.* 50:302–309. https://doi.org/10.1021/acs.accounts.6b00491.
12. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242. https://doi.org/10.1093/nar/28.1.235.
13. O'Boyle, N. M., M. Banck, …, G. R. Hutchison. 2011. Open Babel: An open chemical toolbox. *J. Cheminf.* 3:33. https://doi.org/10.1186/1758-2946-3-33.

14. Nievergelt, Y. 2000. A tutorial history of least squares with applications to astronomy and geodesy. *J. Comput. Appl. Math.* 121:37–72. https://doi.org/10.1016/S0377-0427(00)00343-5.
15. Spearman, C. 1904. The Proof and Measurement of Association Between Two Things. *Am. J. Psychol.* 15:72–101. https://doi.org/10.2307/1412159.
16. Levitt, M. 2001. The birth of computational structural biology. *Nat. Struct. Biol.* 8:392–393. https://doi.org/10.1038/87545.
17. Trott, O., and A. J. Olson. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31:455–461. https://doi.org/10.1002/jcc.21334.
18. Krammer, A., P. D. Kirchhoff, …, M. Waldman. 2005. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.* 23:395–407. https://doi.org/10.1016/j.jmgm.2004.11.007.
19. Böhm, H.-J. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 8:243–256. https://doi.org/10.1007/BF00126743.
20. Böhm, H.-J. 1998. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* 12:309. https://doi.org/10.1023/A:1007999920146.
21. Meng, E. C., B. K. Shoichet, and I. D. Kuntz. 1992. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13:505–524. https://doi.org/10.1002/jcc.540130412.
22. Zhu, H., J. Yang, and N. Huang. 2022. Assessment of the Generalization Abilities of Machine-Learning Scoring Functions for Structure-Based Virtual Screening. *J. Chem. Inf. Model.* 62:5485–5502. https://doi.org/10.1021/acs.jcim.2c01149.
23. Yang, Y. X., P. Wang, and B. T. Zhu. 2022. Importance of interface and surface areas in protein-protein binding affinity prediction: A machine learning analysis based on linear regression and artificial neural network. *Biophys. Chem.* 283:106762. https://doi.org/10.1016/j.bpc.2022.106762.
24. Yang, Y. X., P. Wang, and B. T. Zhu. 2023. Binding affinity prediction for antibody–protein antigen complexes: A machine learning analysis based on interface and surface areas. *J. Mol. Graph. Model.* 118:108364. https://doi.org/10.1016/j.jmgm.2022.108364.
25. Yang, Y. X., J. Y. Huang, …, B. T. Zhu. 2023. AREA-AFFINITY: A Web Server for Machine Learning-Based Prediction of Protein-Protein and Antibody-Protein Antigen Binding Affinities. *J. Chem. Inf. Model.* 63:3230–3237. https://doi.org/10.1021/acs.jcim.2c01499.
26. Yang, J., C. Shen, and N. Huang. 2020. Predicting or Pretending: Artificial Intelligence for Protein-Ligand Interactions Lack of Sufficiently Large and Unbiased Datasets. *Front. Pharmacol.* 11:69. https://doi.org/10.3389/fphar.2020.00069.
27. Wang, C., and Y. Zhang. 2017. Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. *J. Comput. Chem.* 38:169–177. https://doi.org/10.1002/jcc.24667.
28. Sun, H., L. Duan, …, T. Hou. 2018. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* 20:14450–14460. https://doi.org/10.1039/c7cp07623a.
29. Wang, E., H. Sun, …, T. Hou. 2019. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* 119:9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055.
30. Abel, R., L. Wang, …, R. A. Friesner. 2017. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* 50:1625–1632. https://doi.org/10.1021/acs.accounts.7b00083.
31. Zhang, H., W. Jiang, …, Y. Luo. 2019. Ranking Reversible Covalent Drugs: From Free Energy Perturbation to Fragment Docking. *J. Chem. Inf. Model.* 59:2093–2102. https://doi.org/10.1021/acs.jcim.8b00959.
32. Bhati, A. P., S. Wan, …, P. V. Coveney. 2017. Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *J. Chem. Theor. Comput.* 13:210–222. https://doi.org/10.1021/acs.jctc.6b00979.
33. Lee, T.-S., Y. Hu, …, D. M. York. 2017. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theor. Comput.* 13:3077–3084. https://doi.org/10.1021/acs.jctc.7b00102.
34. Song, L. F., T.-S. Lee, …, K. M. Merz, Jr. 2019. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* 59:3128–3135. https://doi.org/10.1021/acs.jcim.9b00105.
35. Yan, Y., M. Yang, …, J. Z. H. Zhang. 2017. Interaction Entropy for Computational Alanine Scanning. *J. Chem. Inf. Model.* 57:1112–1122. https://doi.org/10.1021/acs.jcim.6b00734.
36. Liu, X., L. Peng, …, J. Z. H. Zhang. 2018. Computational Alanine Scanning with Interaction Entropy for Protein–Ligand Binding Free Energies. *J. Chem. Theor. Comput.* 14:1772–1780. https://doi.org/10.1021/acs.jctc.7b01295.
37. Yang, Y.-p., L.-p. He, …, J. Z. H. Zhang. 2019. Computational analysis for residue-specific CDK2-inhibitor bindings. *Chin. J. Chem. Phys.* 32:134–142. https://doi.org/10.1063/1674-0068/cjcp1901012.
38. Du, X., Y. Li, …, S. Q. Liu. 2016. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* 17:144. https://doi.org/10.3390/ijms17020144.
39. Fischer, E. 1894. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* 27:2985–2993. https://doi.org/10.1002/cber.18940270364.
40. Koshland, D. E. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–104. https://doi.org/10.1073/pnas.44.2.98.
41. Ma, B., S. Kumar, …, R. Nussinov. 1999. Folding funnels and binding mechanisms. *Protein Eng.* 12:713–720. https://doi.org/10.1093/protein/12.9.713.
42. Csermely, P., R. Palotai, and R. Nussinov. 2010. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* 35:539–546. https://doi.org/10.1016/j.tibs.2010.04.009.