

RESEARCH ARTICLE

# The Universal Statistical Distributions of the Affinity, Equilibrium Constants, Kinetics and Specificity in Biomolecular Recognition

Xiliang Zheng<sup>1</sup>, Jin Wang<sup>2\*</sup>

**1** State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, ChangChun, Jilin, P.R. China, **2** Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, New York, United States of America

\* [jin.wang.1@stonybrook.edu](mailto:jin.wang.1@stonybrook.edu)



## Abstract

We uncovered the universal statistical laws for the biomolecular recognition/binding process. We quantified the statistical energy landscapes for binding, from which we can characterize the distributions of the binding free energy (affinity), the equilibrium constants, the kinetics and the specificity by exploring the different ligands binding with a particular receptor. The results of the analytical studies are confirmed by the microscopic flexible docking simulations. The distribution of binding affinity is Gaussian around the mean and becomes exponential near the tail. The equilibrium constants of the binding follow a log-normal distribution around the mean and a power law distribution in the tail. The intrinsic specificity for biomolecular recognition measures the degree of discrimination of native versus non-native binding and the optimization of which becomes the maximization of the ratio of the free energy gap between the native state and the average of non-native states versus the roughness measured by the variance of the free energy landscape around its mean. The intrinsic specificity obeys a Gaussian distribution near the mean and an exponential distribution near the tail. Furthermore, the kinetics of binding follows a log-normal distribution near the mean and a power law distribution at the tail. Our study provides new insights into the statistical nature of thermodynamics, kinetics and function from different ligands binding with a specific receptor or equivalently specific ligand binding with different receptors. The elucidation of distributions of the kinetics and free energy has guiding roles in studying biomolecular recognition and function through small-molecule evolution and chemical genetics.

## OPEN ACCESS

**Citation:** Zheng X, Wang J (2015) The Universal Statistical Distributions of the Affinity, Equilibrium Constants, Kinetics and Specificity in Biomolecular Recognition. *PLoS Comput Biol* 11(4): e1004212. doi:10.1371/journal.pcbi.1004212

**Editor:** Guanghong Wei, Fudan University, CHINA

**Received:** May 29, 2014

**Accepted:** February 24, 2015

**Published:** April 17, 2015

**Copyright:** © 2015 Zheng, Wang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** JW thanks National Science Foundation for support. This work is supported by the National Natural Science Foundation of China (Grants 11174105, 91227114, 91430217 and 21190040). This work is also funded by Science and Technology Development Plan of Jilin Province (20130101179Jc) and Jilin province public computing platform. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Summary

Uncovering the principles and underlying mechanisms of biomolecular recognition and molecular binding process is crucial for understanding the function and evolution, yet challenging. We meet the challenge by quantifying the statistical natures of the relevant physical variables of biomolecular recognition using the analytical model combined with microscopic flexible docking simulation methods. We uncovered the universal statistical

**Competing Interests:** The authors have declared that no competing interests exist.

laws obeyed by the affinity, equilibrium constant, intrinsic specificity and kinetics for biomolecular recognition. The general statistical laws based on energy landscape theory can serve as a conceptual framework for molecular recognition in biological repertoires. They can be applied to molecular selection, in vitro evolution process, high throughput screening and virtual screening for drug discovery. The statistical laws in combinations with experiments provide quantitative signatures of a specific ligand binding to a specific receptor, these resultant laws as a guideline will contribute to drug design against a specific target. Our developed statistical methodology is general and applicable for all other biomolecular recognitions.

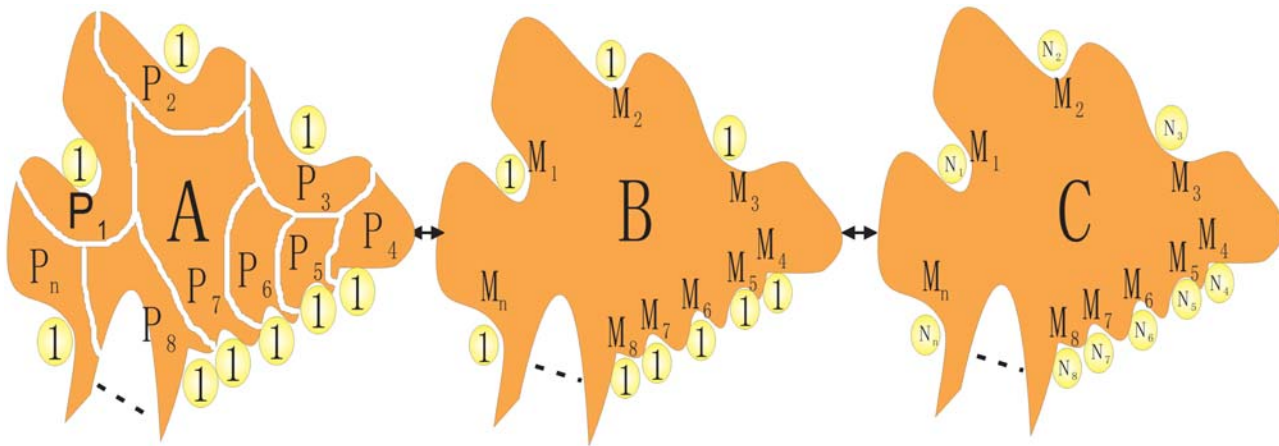
## Introduction

Molecular recognition has been a long standing issue of molecular biology [1, 2]. On the practical side, it is also under intensive investigations in drug discovery and pharmaceutical industry [3]. There are two major issues in biomolecular binding. One is the affinity which is responsible for the driving force and the stability of the binding complex. The other one is the specificity [4, 5] which is crucial for molecular recognition measuring discriminations of “good” binding against “bad” binding. Using microscopic atomistic descriptions to quantitatively study binding is rather difficult [6, 7]. Further more, the accurate estimates of the physical relevant quantities are limited by the delicate balance between hydrophobic interactions, electrostatic interactions, solvation effects and conformational entropy.

The technology advances in combinatorial synthesis of peptide [8] as well as virtual screening of small molecule databases provide a new route [9, 10] to address the above mentioned issue. Instead of studying a specific pair of biomolecular binding, for example, a receptor-ligand (or receptor-small molecule) complex, one can now search through the sequence space of ligands or different small molecules and perform the binding experiments for each specific sequence of ligand binding to a particular receptor. The sequences or segments with good binding properties with the receptor will be the ones to select. Therefore by exploring the sequences of ligands (typically  $10^6 \sim 10^9$  sequences, a large number which is perfect for the statistical description) or small molecule databases (the number of small molecules can be on the order of  $10^{60}$ ), one can explore the biological specificity and function by mimicking the natural evolution selection process with the fittest surviving from the ensemble of ligands. The combinatorial library or virtual screening of small molecule database therefore provides a natural laboratory for uncovering the fundamental principles in biomolecular binding and design through the study of the statistical features of the ensembles of the ligands or small molecules with different sequences binding to a specific receptor. In this way, one can characterize the receptor-ligand system and obtain the important statistical information and distributions on the relevant physical properties or observables of the system such as the binding free energy, equilibrium constants, and specificity. The distribution of the physical variables obtained can be universal revealing the common features among different biomolecular binding complexes. It helps the understanding of the evolution and function. The values of the parameters in the distribution characterizing the properties of the underlying energy landscape may be different for different binding complexes. They can be inferred from the experiments.

## Theory and Analytical Models

Experiments on random sequence protein folding [11] and protein design [12], have implied the statistical distributions of physical observable. Experimental and computational studies on



**Fig 1. The physical equivalences in probing interactions of molecular recognition.** (A) A specific ligand binding to different receptors,  $P_1 \sim P_n$  represent the different receptor proteins with the associated binding sites. (B) Different interactions through different atomic contacts of a specific pair of ligand-receptor complex,  $M_1 \sim M_n$  represent the different interactions with different set of contacts located at the different binding sites of a specific receptor. (C) A specific receptor binding to different ligands,  $N_1 \sim N_n$  represent the different ligands with different sequences.

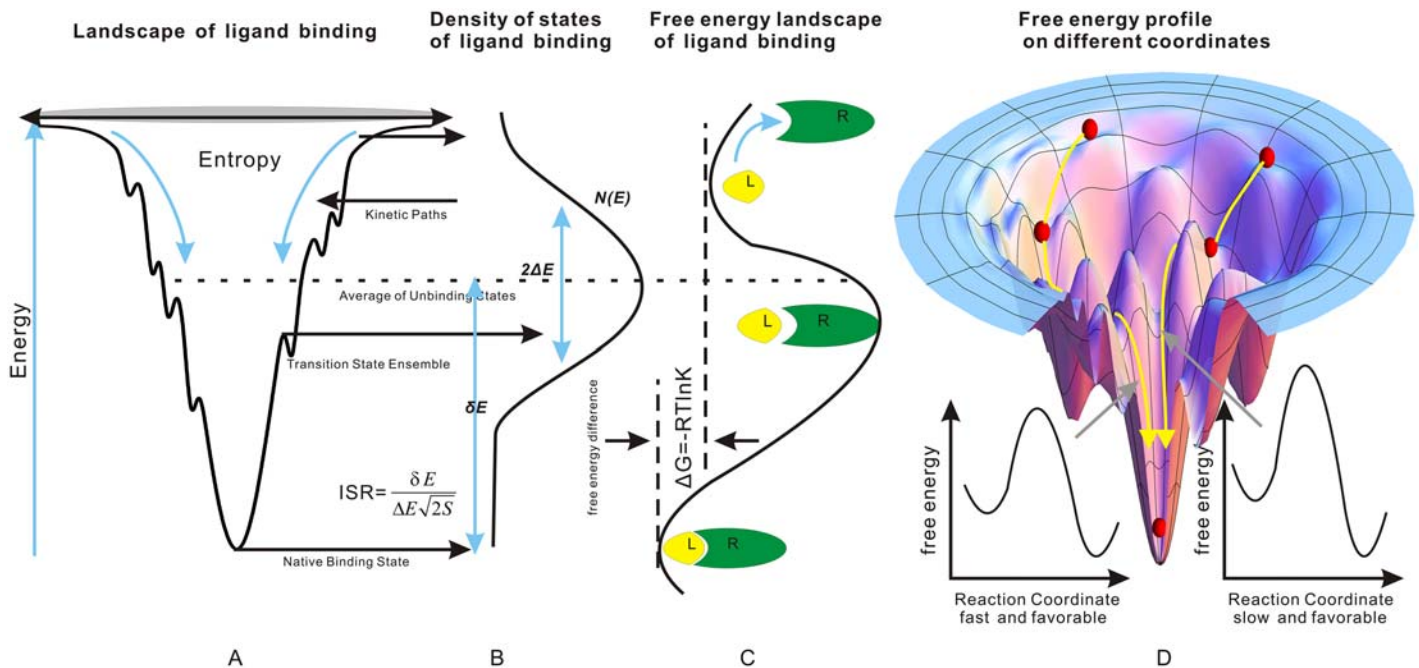
doi:10.1371/journal.pcbi.1004212.g001

biomolecular binding have also shown evidences of distribution of physical relevant quantities such as binding equilibrium constants  $K$  [13–18]. Since  $\text{Log } K$  is proportional to the free energy difference between the native and non-native states termed as the stability or affinity, this infers that the free energy also has a distribution. The experimental features on binding indicate that the appropriate physical variable is the free energy of binding of ligands to the receptor, not the energy. The similar situation also appears for random sequence protein folding and protein design. As a result, the free energy stability or affinity for each specific sequence of ligand can be obtained. Certain values of the affinity appear more often than others. The free energies therefore are distributed. When we discuss about the distribution of the free energies, we mean the sampling of different free binding energies from the different ligands with different sequences binding to the same receptor. The binding free energy of each individual ligand to a specific receptor can be calculated and measured directly from the experiments (through the equilibrium constant measurements). Collecting the free energies from different ligands binding to the same receptor, we can find the distributions of the free energy. The similar procedure applies to the statistics of specificity, equilibrium constants and kinetics for different ligands binding to the same receptor protein. We mainly focus on the statistics of the different ligands binding to the same receptor protein in this study. By exploring the sequence space of ligands, one also equivalently goes through the interactions and conformation space under the ergodicity condition. Here, we describe a physical hypothesis based on a thought experiment (see the Fig 1) [19–21]. Imagine we connect all the different receptors by linkers (for example, connecting N terminus of a protein and C terminus of another by glycines), then the whole universe of receptors now becomes one giant protein. When the receptor protein is large enough, probing the interactions of binding through different parts (binding sites/pockets) of the same receptor protein (panel B) and probing the interactions through the sequences (different receptors in panel A or different ligands in panel C) should be equivalent. The quantitative issue is how large the receptor protein should be in order to see the above mentioned approximate equivalence. Since the protein folds have been estimated to be on the order of a thousand [22], the actual number of the interactions via the atomic contacts of the ligands with the receptors is finite and enumerable. In other words, a large but finite size protein may already contain most

of the interactions encountered for ligand binding. Under the assumption of large receptor protein (large enough to effectively represent the sequences of the diverse receptor universe), obviously searching for all the binding sites or pockets of a particular finite size protein is equivalent to searching for the whole universe of the receptors. Therefore, in this case, probing interactions can be reached approximately equivalently by the following three approaches: (1) multiple ligands binding to the same receptor, or (2) multiple receptors binding to the same ligand, or (3) a ligand binding to a receptor exploring the different binding sites (modes). This hypothesis has been tested and validated for specificity of Cox-2/Cox-1 receptor-ligand complexes [21]. In other words, the statistical properties through the exploration of sequence space of different ligands with the same receptor is equivalent of searching through the conformational or structural space for a particular ligand receptor pair. Although the free energy is in general a complicated function of interactions and entropy, combinatorial library of ligands and database of small molecules provide us a great opportunity to study the interactions and underlying principles of binding. In the free energy distribution, the native (strongest) binding state(mode) should appear in the lowest end of the tail where the density of these binding states becomes discrete.

The conventional specificity refers to the discrimination of binding affinities between a specific ligand with different receptors. However, it is challenge to go through the whole universe of receptors to determine the binding specificity due to the lack of the full information. From the above equivalence discussions in terms of probing the interactions of molecular recognition, another way of quantify the specificity is to find the discrimination in binding affinities of a ligand binding with different binding sites of a receptor. This is referred to the intrinsic specificity. Under large protein assumption, the intrinsic specificity and conventional specificity should be equivalent. The supporting evidence was illustrated in the Cox-2/Cox-1 binding with ligands [19–21]. Both affinity and specificity are crucial for determining the molecular recognition. An optimal quantitative criterion for the intrinsic specificity in discriminating the native from the non-native states for the best binding sequences is found to be the maximization of the intrinsic specificity ratio (ISR) of the energy gap between the energies of the native state and the average of the other states versus roughness or fluctuation modularized by the entropy or size of underlying binding energy landscape [19, 20], see panel B of Fig 2 for the density of states of binding. This implies that the underlying binding landscape is funneled towards the native state. This is not surprising since the driving force of the binding is the same as folding, the hydrophobic interactions. The folding can be seen as self binding and binding can be viewed as folding with multiple domains without the linkages between the domains. Therefore, one expects that the resulting binding energy landscape should have a funneled shape towards the native state to guarantee the stability and intrinsic specificity as well as the kinetic speed of recognition against the bumps or wiggles along the binding paths. (see panel A,C of Fig 2). This is in parallel to the protein folding studies where similar optimal criterion for folding has been used to design fast folding protein sequences [23–25]. Different ligands or small molecules will have different specificity for binding with a specific receptor. Therefore, the intrinsic specificity(ISR) should also have a statistical distribution. This reflects different degrees of binding specificity. There is a small group of high specificity ligands among all the available ones. This is rare and lies in the high end tail of the distribution of specificity.

There is a topologically unique native binding state at the bottom of the funnel (See Fig 2). We can use RMSD as order parameter to describe the position of an ensemble of states in the landscape. Through the RMSDs, the binding landscape is stratified and simplified. We can define an average energy over many states with different energies and RMSDs within each stratum. Then the local kinetics between the binding states with the different average energies can be explored, the global kinetics between the native state and non-binding ones can also be



**Fig 2. Funneled energy landscape of the biomolecular binding.** Panel (A) shows the energy landscape of the receptor-ligand binding with a funneled shape towards the native state. Panel (B) shows the density of states of the ligand binding landscape. The intrinsic specificity ratio  $ISR = \frac{\delta E}{\Delta E \sqrt{2S}}$ , where energy gap  $\delta E$  and energy roughness  $\Delta E$  and size of the binding funnel measured by the entropy  $S$  are shown respectively. Panel (C) shows the free energy landscape of ligand binding, cartoon showing of receptor/ligand complex corresponds to the different binding states. The affinity measured by the free energy difference between the native binding state and unbound states are shown as  $\Delta G$ . It can also be measured by the equilibrium constant  $K$  where  $\Delta G = -RT \ln K$ . Panel (D) shows the energy landscape and free energy profiles on different reaction coordinates.

doi:10.1371/journal.pcbi.1004212.g002

described. In other words, the kinetics of ligand receptor recognition on the binding energy landscape can also be studied. We have previously performed analytical studies of kinetics with both global and local connectivity [26, 27] as well as microscopic flexible docking studies of ligand receptor binding [21]. We also found that intrinsic specificity quantified by ISR not only guarantees the thermodynamic stability and specificity but also guarantees the kinetic accessibility and the speed of the recognition. While the average kinetics has been investigated, there are only limited studies on the statistics of the kinetics, mostly at the analytical level [27–32].

The theoretical and modeling studies on the statistical distributions of the physical variables such as the affinity, equilibrium constants, specificity, and kinetics to uncover the underlying processes and the associated energy landscape of binding and conformational dynamics are currently far less explored than their counter part, the mean. Thus it is the purpose of this paper to fill the gap. Furthermore, the advances of combinatorial chemistry with the capability of generating large number of small molecules at once and high throughput screening provide us a great opportunity of obtaining the statistical information on thermodynamics and kinetics for molecular recognition through the exploration of the ensemble of sequence space of small molecules (combinatorial ligand binding). The statistical information on thermodynamics can be probed from the experiments by measuring the thermodynamic stability through equilibrium constant measurements for different ligands binding with a receptor. On the other hand, the statistical information on the kinetics (on/off) can be probed from the experiments by measuring the kinetic rates through dissociation constant measurements for different ligand binding with a receptor. These statistical information will be crucial to uncover the global and

universal features of the thermodynamics and kinetics of the associated molecular recognition. Characterizing the shapes of the statistical distributions of these physical variables relevant to the thermodynamics and kinetics not only allows a better statistical description of the experimental and theoretical results but also permits the testing of hypothesis about the underlying molecular recognition processes. It is also important for the practical application of drug discovery.

## Results and Discussion

We will first present the analytical results and then the simulation results. For analytical studies, since we are mostly interested in the general laws governing the biomolecular binding, the task here is to obtain the universal features. We can use the coarse grained level description instead of the microscopic detailed one to characterize the system.

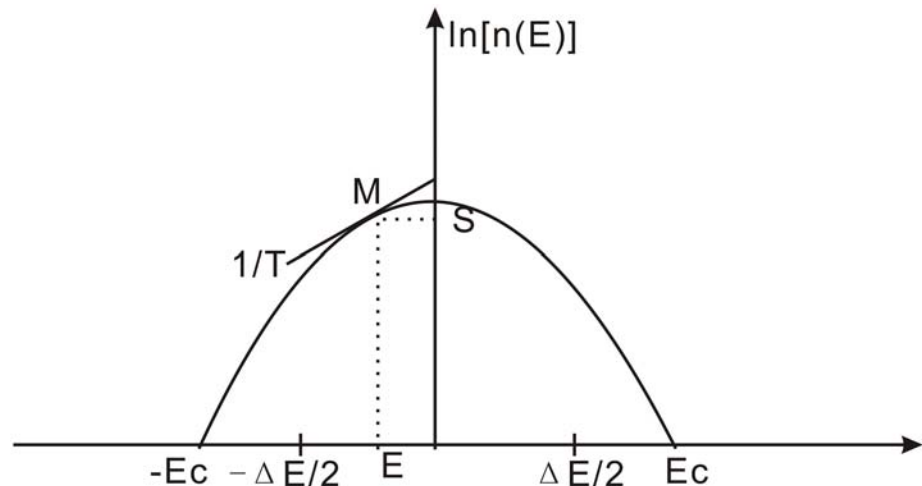
The Hamiltonian or energy function for the interactions between ligands and receptors can be described by the collections of contact interactions between the atom pairs  $E = \sum_{ij} J_{ij} \sigma_{ij}$ , where  $\sigma_{ij}$  is the contact variable between atoms  $i$  and  $j$  with certain distance cutoff and the  $J_{ij}$  is the coupling strength for specific contact pair between  $i$  and  $j$ . Since there are many different types of atoms and also many different cutoff distances for the interactions, different  $J_{ij}$  can have different values. This forms a distribution for the coupling strength  $J$ . Since the number of different  $J$  couplings are large, the statistical distribution should have a Gaussian form from the large number theorem. Since the energy is linearly related to the coupling. This leads to a random energy model with the interaction energy follows a Gaussian distribution [19, 33, 34]. This reflects the complexity of the underlying interactions in contrast to the conventional deterministic models for simple systems where the coupling strengths are fixed and not distributed [35]. Furthermore, we can study the energy landscape of ligand binding using the random energy model with the certain bias towards the native state [19]. This assumes that the energies of non-native states and their interactions are random variables with given probability distributions as discussed above. The binding process can be illustrated as random walks on a rough binding energy landscape with bias towards the native state [19]. Based on the more detailed description in S1 Text [26–30, 32, 36, 37] and the simulation method of this study, we can obtain the free energy, equilibrium constant, intrinsic specificity and the kinetics distribution from the random energy model analytically and those from the direct simulations.

### Analytical Models of Distribution of Affinity, Equilibrium Constants, Specificity and Kinetics

**Universal distribution of free energy of biomolecular binding.** Because the underlying interaction energy is Gaussian distributed, the functional form of the distribution of the resulting free energy can be obtained by carefully studying the moments of the partition function of the resulting random energy model [36, 37]. The distribution function  $f$  has the form of:

$$f(F) \sim \exp\left[-\frac{(F - \bar{F})^2}{2\Delta E^2}\right] \quad (1)$$

where  $f(F)$  is a Gaussian distribution of the free energy  $F$  around its mean  $\bar{F}$  (See details in S1 Text) with the variance of the distribution  $\Delta F^2 = \Delta E^2$  (The width of the energy distribution  $\Delta E$  here has the meaning of the roughness of the underlying energy landscape above the trapping transition temperature  $T_c$ .  $T_c = \sqrt{\frac{\Delta E^2}{2S}}$  is the trapping transition temperature representing the onset of the local trapping of the underlying energy landscape,  $S$  represents the configurational entropy measuring the size of the configurational space which scales with the size of the system



**Fig 3. The entropy  $S(E)$  as a function of energy for the random energy model.** The curve is entropy  $S(E) = \ln[n(E)]$ , where  $n(E)$  is the density of states with energy  $E$  from the spectrum,  $E_c$  is the corresponding energy of the trapping transition temperature  $T_c$ . As we see, the point  $M$  moves on the curve with the changes in the slope ( $1/T$ ).

doi:10.1371/journal.pcbi.1004212.g003

(number of atoms), herein, using the ligand-receptor binding complex as a system.  $S = K_B \log \Omega$  where  $\Omega$  represents the number of states in the configurational space.) Furthermore,  $f(F)$  can be shown to have the exponential distribution in the tails near or below  $T_c$  [32, 36, 37]:

$$f(F) \sim \exp\left[\frac{\rho(F - F_c)}{T}\right] \theta(F_c - F) \quad (2)$$

where  $\rho = \frac{T}{T_c}$  and  $F_c$  is the cut-off free energy ( $\theta$  function is defined as  $\theta(x) = 1$  when  $x > 0$ ;  $\theta(x) = 0$  when  $x \leq 0$ ). Notice that the mean free energy is close to  $|F_c - T_c|$  (here,  $K_B = 1$ ) and the width of the distribution is on the order of  $T_c$ . Notice that the distribution of free energy develops an exponential tail which decays slower than the Gaussian distribution. Therefore, the rare (low probability) events with low free energies in the spectrum of the specific sequences are more prominent in the exponential distribution.

It should be noted that in the aforementioned analytical random energy model [33, 38, 39], from the thermodynamic Maxwell relation of  $\delta S(E)/\delta E = 1/T$  (Fig 3) where  $S(E)$  is the entropy of the system. Since entropy is related to the distribution of the states in energy or density of states ( $S = \ln[n(E)]$ ), we can see the relationship between the energy (left side of the equation) and temperature (right side of the equation). Physically it is clear that the slope of distribution is inversely proportion to temperature (Fig 3). In other words when the temperature is high (the slope of the distribution  $1/T$  is low), then the energy is at the center of the distribution and when the temperature is low (slope of the distribution  $1/T$  is high), then the energy is at the tail of the distribution (Fig 3). This can be seen clearly in Fig 3. Since free energy is related to energy, high energy leads to high free energy. Therefore, when temperature is high, then the free energy is at the center of the distribution and when the temperature is low, then the free energy is at the tail of the distribution (Fig 3). This is also true for the affinity, equilibrium constant, ISR and kinetic rate distributions where all of these variables are directly related to the free energy. Therefore, these give the correspondences of the analytical results for distribution of the variables in different temperature ranges with the simulation results for distribution of the

variables in different variable ranges. On the other hand, since partition function  $Z$  has an asymptotic form and a power law distribution at the tail, the associated free energy is therefore exponentially distributed at the tail ( $x \rightarrow \infty$ ) (See details in [S1 Text](#)). In other words, for the distribution of free energy, we can obtain two distinct regions (such as low and high free energy) of the corresponding physical variable.

As one can see, it is essential to not only know the information about the mean, but also at least the width of the distribution, in order to characterize the whole free energy spectrum. Failure to do so results incomplete information of the system. It is interesting to notice that experimental evidence for the distribution of free energy seems clear. However, the average is usually obtained and used to study the thermodynamics and the kinetics in the current microscopic atomistic simulations on the binding, where the free energy perturbation and other approximation methods are often used. The higher order statistical information such as variance and the whole distribution are not noticed with much attention and hardly explored. As we point out, with the computer hardware speed rapidly improving, the more comprehensive information can be obtained in silico. Careful analysis based on microscopic simulation studies, the determination of the parameter values appearing in the current analytical approach and comparisons with the experiments are crucial for the more complete description of the binding process.

**The universal distribution of the equilibrium constant of biomolecular binding.** The equilibrium constant  $K$  for the biomolecular binding complex measured often in the experiments is defined as the difference between the free energy of the native  $F_n$  and the completely non-native states  $F_{un}$ :  $\log K = \frac{F_n - F_{un}}{RT}$ . Note that as the differences of the free energies of the native state relative to the non-native states  $F_n - F_{un}$  decreases to more negative values, the stability increases. The  $\log K$  decreases too. The biomolecular binding complex is more stable. The affinity also increases. Therefore from the definition, the  $\log K$  measures the ability of the two biomolecules associate with each other or the affinity. The equilibrium constant can be measured directly from the experiments. In the combinatorial experiments, one can further measure the statistical information of the equilibrium constant by exploring different ligand binding with a receptor one at a time. Based on the probability theory and statistics, if the random variable  $X$  is log-normally distributed, then  $Y = \log(X)$  has a normal distribution. Likewise, if  $Y$  has a normal distribution, then  $X = \exp(Y)$  has a log-normal distribution. In other words, by knowing the distribution of free energy, one can derive the probability distribution of the equilibrium constant  $K$  ( $\log K = \frac{F_n - F_{un}}{RT}$ ) as well since the logarithm of the equilibrium constant  $K$  is directly related to the free energy difference between the native and non-native states (the non-native states are Gaussian distributed while the native state is very narrowly distributed). Since the aforementioned distribution of free energy is the Gaussian in this study, one can easily obtain the distribution function of the equilibrium constant  $K$  as log-normal. So, the distribution function  $f$  of the equilibrium constants  $K$  or the affinity can be shown to be:

$$f(K) \sim \frac{1}{K} \exp \left[ -\frac{T^2}{2\Delta E^2} (\log K - \log \bar{K})^2 \right] \quad (3)$$

which is a log-normal distribution near the mean above  $T_c$  while

$$f(K) \sim K^{-1 - \frac{T}{T_c}} \quad (4)$$

which shows a power law decay near the low  $K$  value tail of the distribution (near or below  $T_c$ ).

This power law behavior is familiar in the physics and chemistry community as a signature of the long range order and self similarity often appeared in the critical phenomena of phase transition, fractals, turbulence and earth-quakes etc. The slow power law decay (as compared with the Gaussian like distribution) indicates a long tail in the equilibrium constant



distribution. In this situation, similarly, the average is no longer a good representative of the distribution. The whole distribution function is needed to characterize the system. This implies that the rare events at the tail may play a dominant role. The phenomenon is often termed as intermittency. The power law behavior of the equilibrium constant in the tail therefore indicates that most of the binding molecules show small binding affinities, only occasionally a specific pattern will lead to high affinity. Therefore, characterizing the global shape of the statistical distribution of equilibrium constant is necessary for the molecular binding. Finding high affinity or the associativity is rare but important to study the evolution and function for binding. By obtaining the experimental measurements of the distributions one can estimate the relevant thermodynamic quantities that fit the experiments from more microscopic approaches based on molecular interactions and structures. Comparing model distributions here with the experiments may also guide the approaches to uncover the underlying correlations between native binding free energy and affinity (equilibrium constant).

**The intrinsic specificity and its universal distribution for biomolecular binding.** From the distribution of the free energy studied in this paper  $f(F) \sim \exp\left[-\frac{(F-\bar{F})^2}{2\Delta E^2}\right]$ , we see that the width or the fluctuations of the free energy spectrum  $\Delta F$  is equal to  $\Delta E$ , and the average of the free energy is  $\bar{F}$ . As discussed earlier, affinity giving the ability of association of two molecules in general can not guarantee the intrinsic specificity that gives the discrimination of the native state from non-native ones. To select a good sequence for binding and realize the specificity, one should have the free energy gap between the ground state or the cluster (group) of states with the lowest free energy of binding relative to the average free energy to be large compared with the width of the distribution of the binding free energy. Only in this way one can discriminate the native state from others and guarantee the intrinsic specificity. That is:

$$\text{Maximize } \frac{|F_n - \bar{F}|}{\Delta F} \tag{5}$$

where  $F_n$  is the free energy of the native state located in the low energy tail of the distribution. This gives the optimal criterion of binding specificity. If we consider further the detail of the exponential distribution near the tails:  $f(F) \sim \exp\left[\frac{T}{T_c} \frac{(F-F_c)}{T}\right]$  where  $F_c$  as previously mentioned is the cutoff free energy, the mean is equal to  $F_c - T_c$  and the width is equal to  $T_c$ , the optimal criterion for specificity now becomes the maximization of  $\frac{|F_n - (F_c - T_c)|}{T_c}$ .

From the expression of the free energy distribution, one can easily obtain the distribution of the native free energy to be:  $f(F_n|N) = f(F_n)f(F_n < N_t - 1)$  where  $F_n$  is the native free energy,  $f(F_n) \sim \exp\left[-\frac{(F_n - \bar{F})^2}{2\Delta F^2}\right]$  is the probability of finding the free energy  $F_n$  and  $f(F_n < N_t - 1) = \left[\int_{F_n}^{\infty} f(F) dF\right]^{N_t - 1}$  is the conditional probability of free energy of native state less than the rest of the spectrum assuming there are total  $N_t$  number of states [40, 41]. Based on this, one can easily convert the above distribution of the native state free energy distribution to the one for the specificity  $ISR$ , ( $ISR = \frac{|F_n - \bar{F}|}{\Delta F}$ ). That is:

$$f(ISR) \sim \exp\left[-\frac{ISR^2}{2}\right] \left[\frac{1}{2} \left(1 + \text{Erf}\left(\frac{ISR}{\sqrt{2}}\right)\right)\right]^{N_t - 1} \tag{6}$$

near the mean and above the trapping temperature  $T_c$ . The distribution becomes narrower and the center of the distribution shifts towards larger values of  $ISR$  when the number of states is larger. In addition, the distribution becomes more skewed under this situation. This implies that as the number of states become more and more, averagely speaking, there are more and more chances of finding the more stable complexes with larger values of  $ISR$ . The

corresponding distribution  $ISR(ISR = \frac{|F_n - (F_c - T_c)|}{T_c})$  near the tail around or below  $T_c$  can be shown to be:

$$f(ISR) \sim \exp[-ISR - 1](1 - \exp[-ISR - 1])^{N_i - 1} \quad (7)$$

Therefore there is exponentially small probability of finding large values of  $ISR$  although the actual absolute number of ligand sequences with large  $ISR$  can be large. If one makes a histogram for the values of the specificity  $ISR$  for the native states and random sequences, one would expect to see the center of the two distributions are well separated by a gap which guarantees the intrinsic specificity. Therefore one can use the optimal criterion to select the stable and specific complexes with good and discriminative binding properties from random sequences.

**Kinetics and its universal distribution for biomolecular binding.** Under the quasi-equilibrium condition, the time scale  $\tau$  of binding of a particular sequence of ligand with a receptor can be approximated as:  $\log(\frac{\tau}{\tau_0}) = \frac{F^\ddagger - F_{un}}{T}$  [28, 29] where  $F^\ddagger$  is the free energy of the transition state ensemble,  $F_{un}$  is the free energy of the completely non-native states,  $\tau_0$  is prefactor for the time scale for the binding process between binding states. Therefore, the time scale of binding from one binding state to a neighboring state is determined by the free energy difference of these two states. In this study, the first passage time(FPT) to reach the native binding state(the time required for the random walker to visit order parameter  $RMSD \sim 0$  for the first time) is used as a typical or representative time scale for binding. The mean first passage time( $\tau$ ) representing the binding time has a U(or V) shape dependence on temperature [27]. Similar to the behavior of K, assuming a common  $F^\ddagger$ , the distribution of  $\tau$  can be shown to have the form as:

$$f(\tau) \sim \frac{1}{\tau} \exp \left[ -\frac{T^2}{2\Delta E^2} \left( \log \frac{\tau}{\tau_0} - \log \frac{\bar{\tau}}{\tau_0} \right)^2 \right] \quad (8)$$

giving log-normal distributions near the mean above  $T_c$  while

$$f(\tau) \sim \tau^{-1 - \frac{T}{T_c}} \quad (9)$$

gives much slower power law decay distributions near the tails of the distribution near or below  $T_c$ ) [31, 32].

Previously we analytically explored the origin of power law distribution of kinetics observed in single molecule conformational dynamics experiments [28, 29, 32]. We established a diffusion and a kinetic master equation approach to study statistically the microscopic state dynamics. The exponential density of states emerges when the system becomes glassy and landscape becomes rough with significant trapping under low temperatures as shown earlier from the analytical studies [32, 36, 36, 37]. We show that the underlying landscape with exponentially distributed density of states leads to power law distribution of kinetics( $f(x) \sim \tau^{-1 - T/T_c}$ ). We predicted that the power law decay coefficient is monotonically dependent on temperature which can be tested from ongoing experiments. This may bridge statistics from single molecule kinetic experiments and topography of conformational energy landscape [42–48].

Therefore, there is a physical explanation of the difference in distributions of binding kinetics: Above the characteristic transition temperature  $T_c$ , there are multiple parallel kinetic paths, each experiencing certain barriers (barrier has a normal distribution), resulting in log-normal kinetics (seen also in protein conformational dynamics simulations [31]). When the temperature drops below  $T_c$ , the distribution of FPT deviates from log-normal significantly. The traps become more important. Discrete pathways emerge. The distribution at long times approaches a power law. This indicates that the fluctuations in FPT are significant and deviate from the mean [27]. The fluctuations start to diverge. This means that the actual conformation dynamic

process may happen on multiple time scales, and the nonself-averaging behavior emerges. The distribution of FPT then has fatty tails. This indicates that the intermittent phenomenon, where rare events can give a significant contribution to the binding. Due to the ruggedness of the binding energy landscape at low temperatures, specific different discrete paths give distinct contributions to the kinetics. The shape of distribution of FPT is then required to characterize the whole system dynamics rather than only the mean.

It is worthwhile to emphasize that our analytical studies in conformational dynamics, protein folding and ligand binding all indicate that the kinetics should follow log-normal distribution at higher temperatures and power law distribution at the low temperatures [27–29, 31, 32]. Our analytical studies formed a mathematical foundation and motivation for us to explore the distribution in kinetics in real ligand binding.

Notice that due to the equivalence of the ensemble exploring different ligands and ensemble exploring different interactions through different atomic contacts of specific ligands to receptor, the statistical distribution of the physical relevant quantity such as binding affinity from different ligands binding to the same receptor is equivalent to the distribution of the one obtained from different interactions of the specific ligand-receptor binding interactions through spatial atomic contacts (See Fig 1). Therefore in analytical model, for example, we can obtain the distribution of binding affinity for the ensemble of different atomic contact interactions (with different non-native unbinding states). Due to the equivalence, this implies that the affinity for receptor binding to different ligands or ligand binding to different receptors is expected to follow the same statistical distribution.

## Microscopic Atomic Binding Model and Simulation Results

We have performed the investigation of the significance and implications of flexible docking of ligands with the receptor target COX-2. Initially a diverse set of 720 small molecules were selected from the NCI-Diversity database [49] having molecular weights similar to that of the reference compound SC-558, for which the crystal structure of the COX-2 complex is available (PDB code 1CX2) [50, 51]. All conformers of each of the 720 selected molecules were docked with COX-2 using AutoDock [52] to generate a binding energy spectrum for each. Furthermore, the correlation coefficient of 0.65 between experimental and predicted affinities (Fig 4 and S1 Table) demonstrate reasonable reliability of Autodock scoring for the Cox-2 target.

From this data, the statistical distribution of the affinity (defined as the free energy difference between the native state and the average of the free energy) of 720 small molecules with COX-2 was described. Fig 5 shows the distribution of affinity. As easily seen, near the center or the mean, the distribution can be fitted well with a Gaussian. Near the tail, the distribution of the affinity can be fitted well with exponential. This confirms the analytical results discussed above. Most of the ligands bind to a receptor with relatively small affinities. A small number of the ligands has high affinity to a receptor, these ligands are crucial for exploring the biological function of the receptor.

In Fig 6, we show the logarithm of the equilibrium constant distribution for the 720 small molecule binding with Cox2. We see that the logarithm of equilibrium constant can be fitted well with a normal distribution near the mean while near the tail can be fitted well with a exponential distribution. Thus, we can extrapolate the statistical properties of the equilibrium constant distribution. That is, the equilibrium constant can be fitted well with a log-normal distribution near the mean while near the tail can be fitted well with a power law distribution. In contrast to the gaussian distribution, the log-normal distribution has a longer tail representing the higher frequencies of equilibrium constant when plotted in the original scale without taking the logarithm. The power law distribution of the equilibrium constant suggests most of

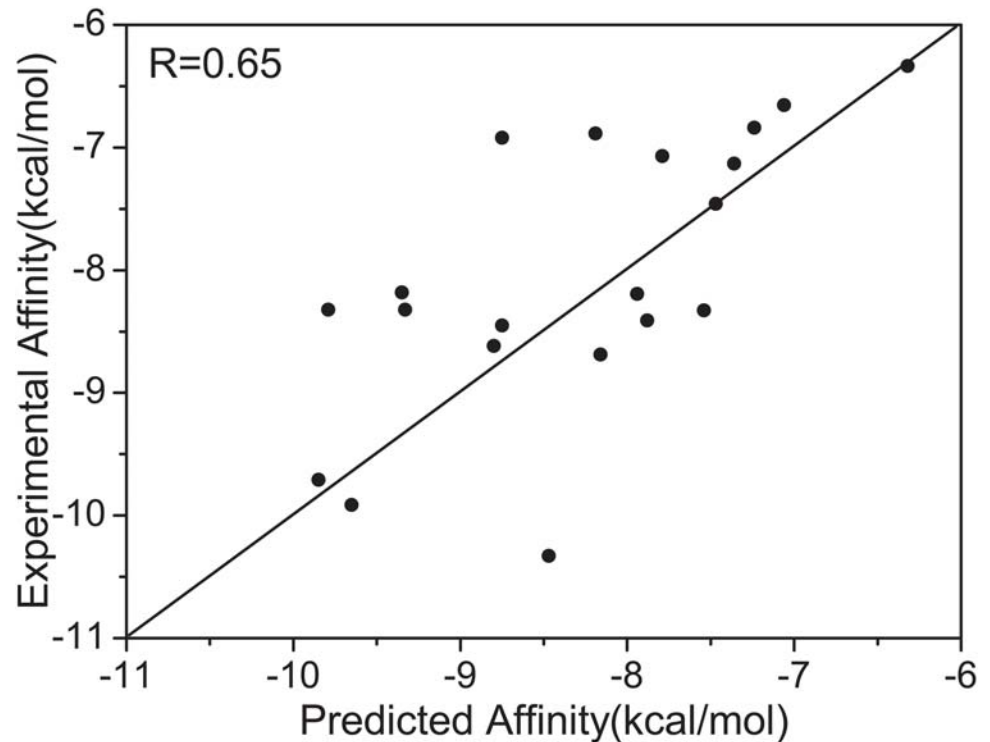


Fig 4. Validation of Autodock scoring to predict the binding affinities for 20 drugs against the Cox-2.

doi:10.1371/journal.pcbi.1004212.g004

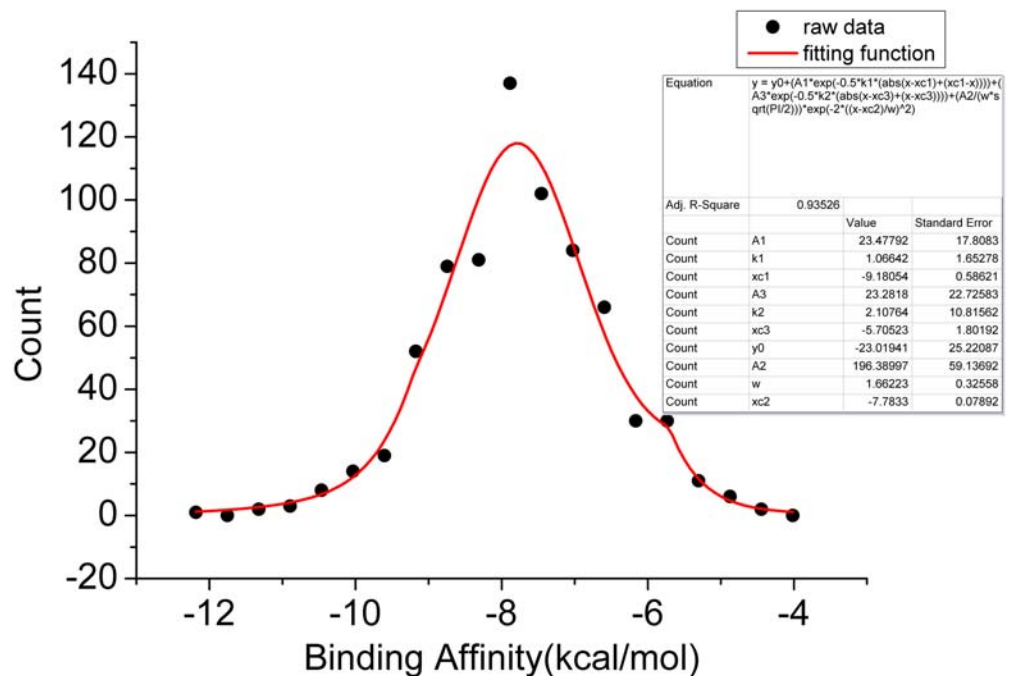
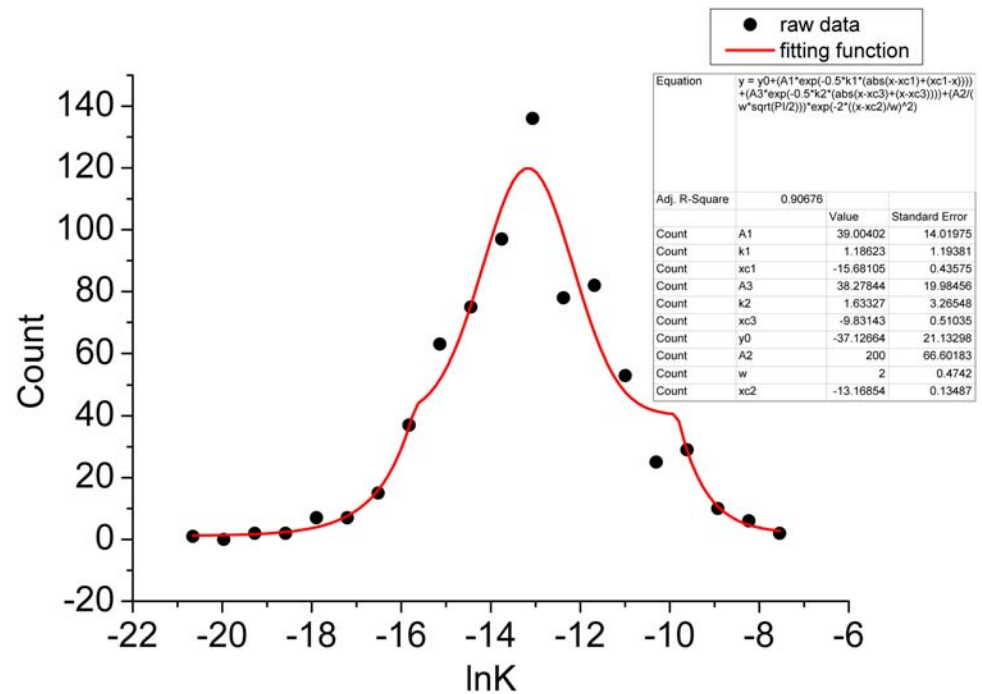


Fig 5. The distribution of the free energy with 720 compounds binding to the Cox2. the vertical axis represents the number of states or probability for the free energy, gaussian curve: in the center; exponential curve: near the tail. The parameter values for these distributions are presented. The analytical function consisting of a Gaussian and two exponential functions is highlighted with red line.

doi:10.1371/journal.pcbi.1004212.g005



**Fig 6. The distribution of the logarithm of equilibrium constant with 720 compounds binding to the Cox2.** the vertical axis represents the number of states or probability for the equilibrium constant, gaussian curve: in the center; exponential curve: near the tail. The parameter values for these distributions are presented. The analytical function consisting of a Gaussian and two exponential functions is highlighted with red line.

doi:10.1371/journal.pcbi.1004212.g006

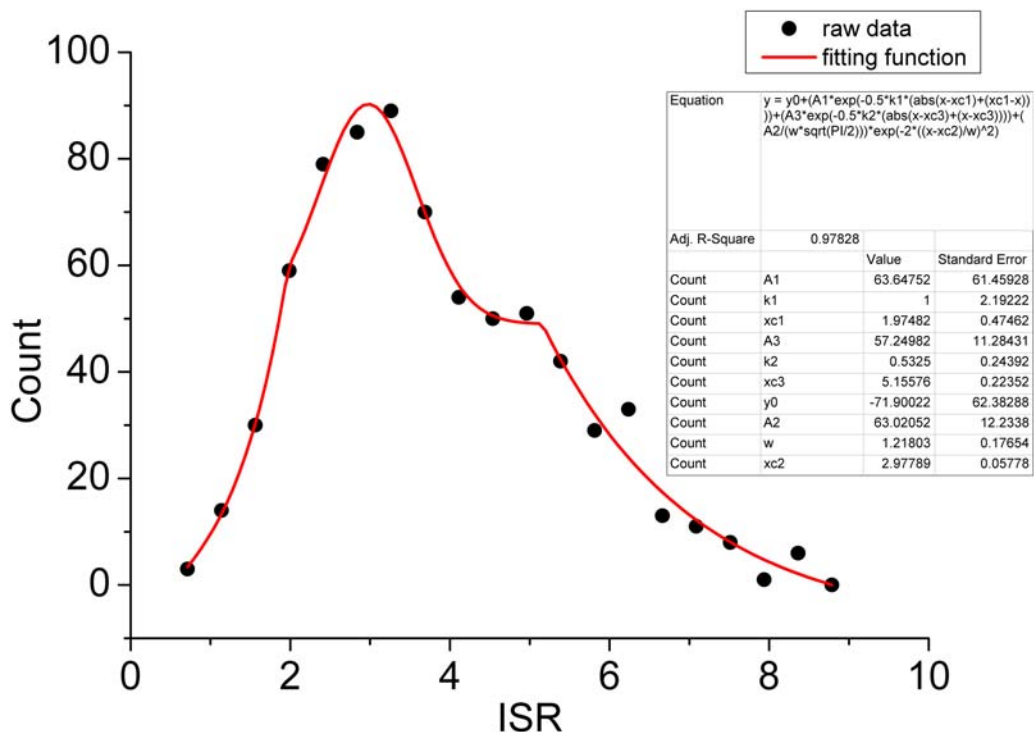
the equilibrium constants of the ligands with random sequence binding to the receptor are small and binding complex are less stable. The large equilibrium constants although rare can contribute greatly in determining the biological stable function. They are the critical ones in molecular recognition.

In Fig 7, we show the statistical distribution of the intrinsic specificity characterized by ISR. We also see that the distribution of the intrinsic specificity can be fitted well with the normal distribution near the mean and also well fitted with the exponential distribution at the tail. This means most of the ligands with random sequence non-specifically bind to a receptor with low specificity. A few ligands specifically bind with a receptor with high specificity. They are at the high end tail of the intrinsic specificity distribution.

Herein, we have computationally determined the kinetic specificity of binding using kinetic off time of binding  $Time_{off}$  to represent the time scale  $\tau$ . To demonstrate the reliability of the  $Time_{off}$  we have defined, the analysis regarding the correlation between the predicted and experimental kinetics have been performed for the 22 drugs against the Cox-2 target, a reasonable correlation with the coefficient 0.62 (Fig 8 and S1 Table) was obtained.

As described in Fig 9, we found that the distribution of kinetics can be fitted quite well with log-normal distribution near the mean and power law distribution at the tail. This is consistent with our analytical models above.

The affinity and equilibrium constants can be measured by experiments directly. From combinatorial point of view, each ligand binds to a receptor with certain affinity and equilibrium constant. By collecting the measurements of different ligands (combinatorial chemistry of peptides or screening with a library of small molecules) each binding with a receptor, average



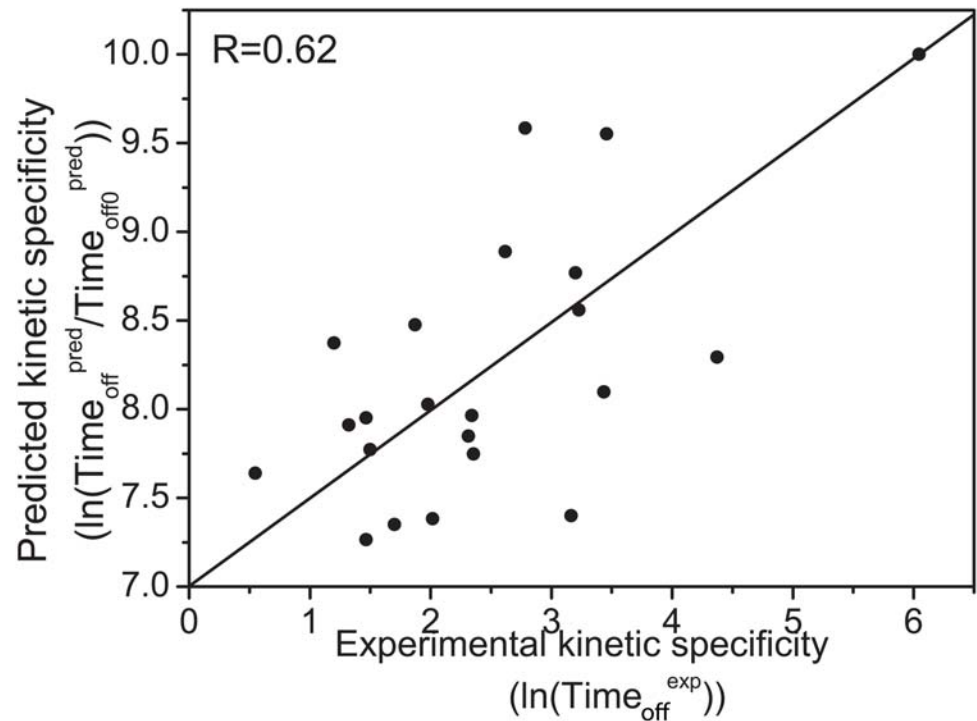
**Fig 7. The distribution of the specificity (ISR) with 720 compounds binding to the Cox2.** the vertical axis represents the number of states or probability for the specificity, gaussian curve: in the center; exponential curve: near the tail. The parameter values for these distributions are presented. The analytical function consisting of a Gaussian and two exponential functions is highlighted with red line.

doi:10.1371/journal.pcbi.1004212.g007

of these quantities can be obtained. Furthermore the statistical information such as high order moments and distribution (histograms) or probability can be also obtained. Similarly, the kinetics can be also explored easily for the statistical information. Intrinsic specificity, on the other hand, is quite difficult to directly measure. However one can indirectly infer the statistical information for the intrinsic specificity from the statistical measurements of both the affinity and kinetics. The ligands with the same affinity might still have different specificity from kinetic point of view.

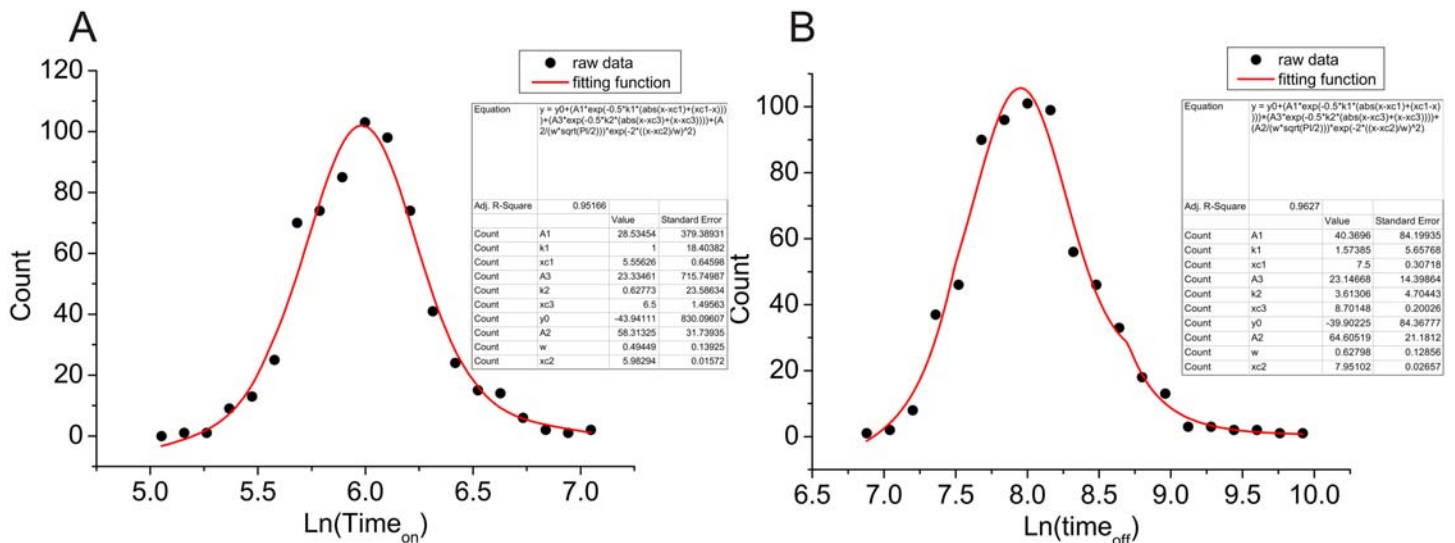
It is important to realize that although the functional form of the distribution is universal from one ligand to another, the fitting coefficients or parameters in the distribution are different for each ligand. This gives a quantitative signature or characterization of a specific ligand binding to a specific receptor. Importantly, we have found analytical functional expressions for the distributions of free energy, specificity, equilibrium constants and kinetics covering the whole range of these variables including near the mean and at the tail as well as the regime in between (See details in [S2 Text](#)) which can give Gaussian statistics at the center and exponential statistics at the tail, respectively.

In [Fig 10A–10C](#), we also show the one dimensional projection of binding free energy landscape to RMSD with different ligands with different intrinsic specificity characterized by ISR. Different free energy landscapes give different free energy barrier heights between the non-native and native states. This leads to different kinetics for binding ([Figs 2 and 10D](#)). As we can see the large intrinsic specificity characterized by high ISR gives smaller free energy barrier and lower ISR gives higher barrier. There is also a structural correspondence shown in the figure for different ligands. The ligands with the fork structure have high structural specificity and



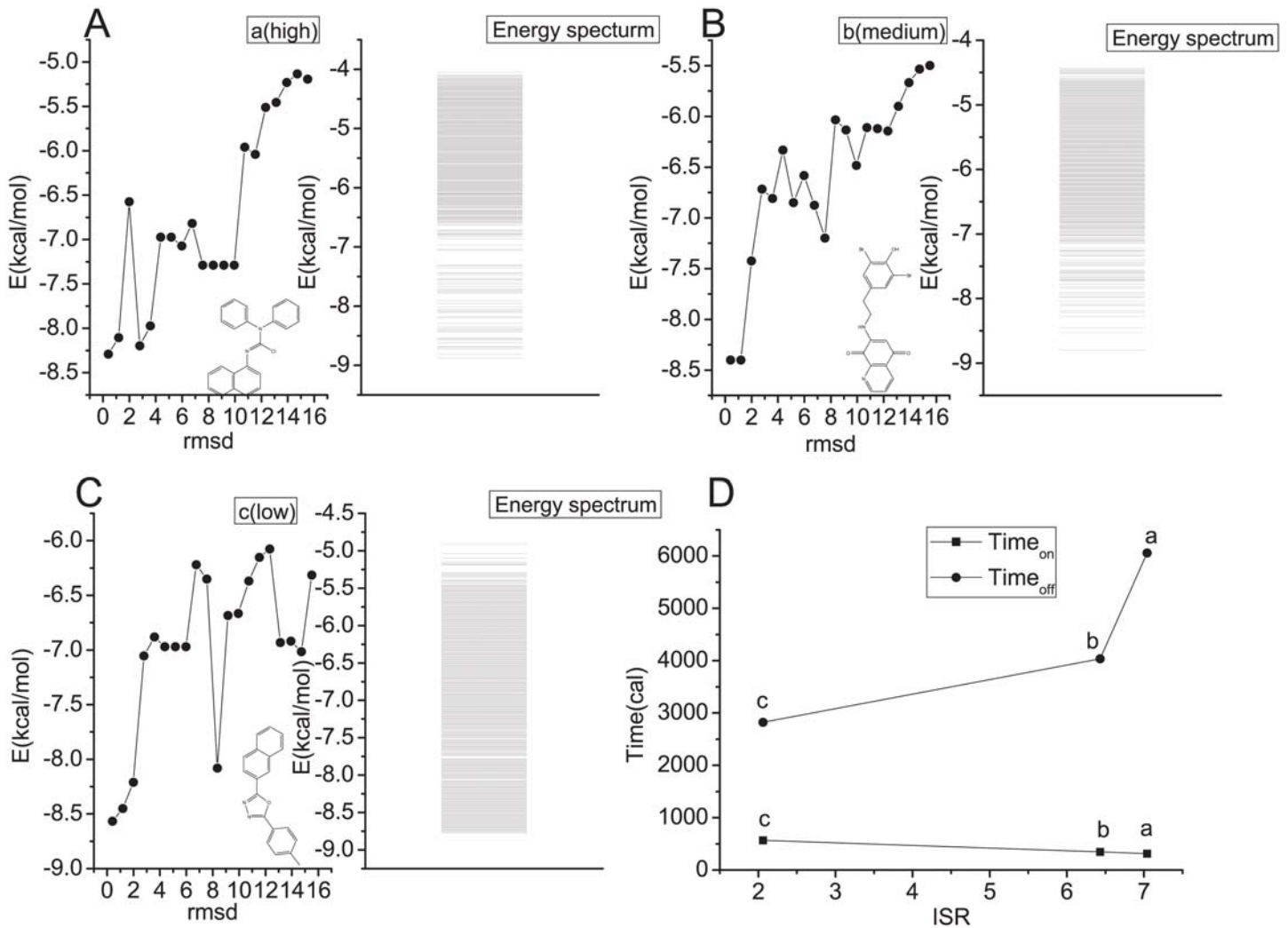
**Fig 8. The relationship between the predicted and experimental kinetic specificities for 22 drugs against the Cox-2.**  $Time_{off}^{pred}$  represents the relative predicted residence time and  $Time_{off_0}^{pred}$  is the constant weighting factor,  $Time_{off}^{exp}$  is the experimental residence time.

doi:10.1371/journal.pcbi.1004212.g008



**Fig 9. The distribution of the predicted kinetic specificity.** (A) The distribution of the logarithm of the predicted  $Time_{on}$  with 720 compounds binding to the Cox2; (B) The distribution of the logarithm of the predicted  $Time_{off}$  with 720 compounds binding to the Cox2. gaussian curve: in the center; exponential curve: near the tail. The parameter values for these distributions are presented. The analytical function consisting of a Gaussian and two exponential functions is highlighted with red line.

doi:10.1371/journal.pcbi.1004212.g009



**Fig 10. Three representative ligands analyzed in the docking simulations.** The left panels of (A), (B) and (C) show the one dimensional projection of binding free energy landscape to RMSD with high, medium and low ISR with similar affinity as well as the corresponding structures of the different ligands, where the docked pose with the most stable affinity or the strongest binding state is chosen as the reference structure to calculate the RMSDs; the Autodock score function is used to evaluate the interaction energies between the ligand and the receptor. (high:fork structure, medium:near-linear structure, low:linear structure);the right panels of them show the corresponding binding energy spectrum for each, the sparse part of the spectrum implies there are fewer states, and dense part of the spectrum implies there are more states. (D) show the kinetic time for binding for the above three different ISRs. The upper black line represents the predicted kinetic  $Time_{off}$  and the under black line represents the predicted kinetic  $Time_{on}$ . The vertical axis represents the calculated kinetic time.

doi:10.1371/journal.pcbi.1004212.g010

faster kinetics, and ligands without the fork structure have low specificity and slower kinetics. We plot the free energy barrier height and the kinetic time for binding of these three cases as shown in Fig 10D. By exploring the different off rates  $Time_{off}$  and on rates  $Time_{on}$  of binding for three ligands, we found that the ligand with higher ISR has faster on rate and slower off rate. In other words, the ligand has longer residence time for binding. In contrast, the ligand with lower ISR has slower on rate and faster off rate with shorter residence time. In a word, we found indeed that the large ISR leads to lower barrier height and faster kinetics. This implies that the kinetics is determined by the intrinsic specificity which is determined by the competition of the slope of the landscape towards the native state and the roughness of the landscape,



not just by the affinity alone as conventionally people would have thought. There are recently growing experimental evidences of the similar affinity but completely different kinetic accessibility [53, 54].

In the experiments, we can have a quantitative indicator through kinetics for specificities from each ligand binding to the receptor. It is also interesting to note that large ISR tends to imply the underlying energy landscape is a smooth funnel towards the native state and binding kinetics is fast. This usually results exponential kinetics. The smaller values of ISR tend to imply bumpy energy landscape with the kinetic rates following a slow decay distribution, producing non-exponential kinetics.

The non-covalent interactions in biomolecular recognition have been extensively analysed. HTS(High Throughput Screening) and VS(Virtual Screening) technologies which are pertinent to receptor and ligand repertoires, are increasingly becoming significant for exploring the universal statistical behavior of the entire ensemble of binding. This may provide us an insight to the physical universal laws that underlie non-covalent interactions in biomolecular recognition. An intriguing result of the present distribution analysis is the possibility of extrapolation of the fitted curves from the center to tail region(for example, the low affinity and the high-affinity domain). Thus, based on the fitted parameter from relatively small ligand database (about 700 members in the current study), which statistically attain binding affinities in the range of -4kcal/mol to -12kcal/mol; it is possible to derive information regarding the much higher binding affinities for larger libraries. The probability and physical relevant quantity range is found to fit adequately the distribution functions from the analytical model, and the two kinds of curves are fitted well to the data in the center and near the tail.

Based on its ability to describe the physical relevant quantities and patterns over the different range, from in the center to near the tail, the general distribution laws based on the analytical model from the energy landscape theory can serve as a conceptual framework to explore molecular recognition in biological repertoires. It could be naturally applied to study human olfactory threshold variability [13, 55], autoimmune phenomena such as self vs. non-self recognition [56–58], and specificity of drugs for the Cytochromes P450 repertoire. Remarkably, the comprehensive knowledge of the statistics of affinity distributions in many studies have led to a better understanding of molecular selection [59], and in vitro evolution process such as combinatorial library panning, SELEX technology exploring the RNA/DNA-ligand interactions [60–64], and high throughput screening for drug discovery to study the receptors or enzymes binding with the lead compounds [65, 66].

## Summary and Conclusion

We have discussed the main statistical features of binding free energy landscapes. In particular we obtained the analytical functional form of the probability function of the binding free energy to be Gaussian distributed near the mean and exponential-like distributed in the tail. Furthermore we obtain the log-normal distribution near the mean and power law decay distribution near the tail for the equilibrium constants of binding. The optimal criterion of native ground state dominance and the intrinsic specificity should be the maximization of the ratio of free energy gap between native state and the average non-native states versus the width of the distribution of the free energy landscape. The quantitative form of intrinsic specificity criterion can be used to develop new algorithms to design ligands with high intrinsic specificity for a particular receptor. Alternatively, one can design receptors that are selective for a particular ligand. The statistical distribution of the intrinsic specificity ratio is also Gaussian distributed near the mean and exponentially distributed near the tail. We give an analytical form of the distribution of the kinetics. We have confirmed the analytical form of the distribution

functions with 720 diversified small molecules binding with a specific receptor: Cox-2 protein enzyme.

We have found there are universal statistical distributions of affinity, equilibrium constant, intrinsic specificity and kinetics of ligand binding. Furthermore, the coefficients and parameters characterizing the ligand binding are different for each ligand. It is worthwhile to point out that there is a weak positive relationship between affinity and intrinsic specificity. We also found a positive correlation between affinity and kinetics of ligand binding (See [S1](#) and [S2](#) Figs). Molecular recognition in protein-ligand, protein-protein and protein-DNA interactions often involves interplay between affinity and specificity or thermodynamics and kinetics. As known, high affinity does not always guarantee the high specificity. We can see that the correlation between affinity and intrinsic specificity is weak because there is a large spread in the correlation plot between the two. That is to say there is a large dispersion in intrinsic specificity at given affinity ([S1 Fig](#)). In general, we need both affinity and specificity to quantify the molecular recognition process. For the lead compound search in drug discovery, high specificity is often achieved by increasing the specific hydrophobic interactions between the ligand and the target [\[20\]](#), while high affinity often relies on some crucial interactions such as hydrogen-bonding. So the quantification of both affinity and specificity rather than affinity alone will contribute to the fundamental understanding of molecular recognition.

On the other hand, the correlation between on time of ligand binding to affinity is weak ([S2A Fig](#)). The residence time has certain degree of negative correlation with affinity ([S2B Fig](#)). This might be due to the increase of the free energy barrier for escaping the binding from the increase of the affinity. The correlation of residence time with respect to affinity is also with spread. But the dispersion is narrower than the correlation between affinity to intrinsic specificity or on time of ligand binding. Therefore, in general, since there is no perfect correlation between affinity and specificity at the thermodynamic or kinetic level, one should consider both of them in exploring the biomolecular recognition.

It should be noted that the interplay between affinity and specificity has also motivated the design of the new score functions of molecular docking [\[21, 67, 68\]](#) by optimizing the corresponding interplays in molecular binding. These further indicate that one needs a comprehensive and multi-dimensional statistical characterizations for molecular recognition.

From these findings, we can provide an integrated quantitative label or indicator for each ligand. So this statistically characterizes each ligand binding to the specific receptor protein with a few characteristic parameters or indicators such as affinity and specificity. When applies to drug screening, this suggests a multi-dimensional screening strategy using both affinity and specificity rather than affinity alone in the traditional drug discovery.

The statistical methodology and approach based on energy landscape theory is quite general, one expects to apply not only to protein-protein/ligand bindings, but also protein-RNA, protein-DNA and RNA-DNA bindings. For more general interactions such as protein-protein interactions and protein nuclei acid interactions, we have performed the investigations on their underlying natures [\[67, 68\]](#). These studies are on the affinity and specificity at the mean level. We plan to generalize our statistical study here for ligand protein interactions to protein-protein and protein-nuclei acid interactions. For rigid binding, we expect similar statistical behaviors. We can take into account flexible binding at the local level by sampling all the important conformations. However, for flexible binding at the global level, it is still a challenge. Some initial efforts have already been made towards this at the mean level that we can use to explore further the associated statistics [\[69\]](#).

In this paper, to the first order approximation, we have ignored correlations between different states. It is expected that the correlations will influence the tail properties of the statistical

distribution of the physical relevant variables quantitatively. It will be interesting to extend the current study to incorporate this effect [70, 71].

## Materials and Methods

### Analytical Models for Binding

Let us now start to work out the functional form of the distribution of the free energy of the ligand-receptor binding complex by exploring the sequence space of different ligands. In order to obtain the free energy, one has to study the thermodynamics of the system. One can do so by starting with an order parameter characterizing the system. A natural choice of the order parameter is the contact variable defined as the contact probability between an atom of the receptor protein and an atom in a small molecule ligand:  $\sigma_{ij}$  ( $\sigma_{ij} = 1$  when  $d < \Delta$  and  $\sigma_{ij} = 0$  when  $d > \Delta$ ) where  $\Delta$  is a fixed cutoff distance usually on the order of several angstrom and  $d$  is the distance between two residues (or two atoms in the case of small molecule binding with a receptor) which measures how close the two residues are to each other.

The interaction hamiltonian or the energy function of the system in terms the contact variables can be written as

$$H = \sum_{ij} J_{ij} \sigma_{ij} \quad (10)$$

whereas  $J_{ij}$  is the coupling strength between the one atom on a receptor and another one on a ligand. Of course, one can include more complicated form of interactions such as multi-body interactions which mimic the hydrophobic interactions and non-additive properties of the interactions. But this will not influence the general features of the system. We will therefore postpone the discussion on more complicated forms of hamiltonian in the future.

It is worth noticing that this form of the interaction hamiltonian has been widely used in analytical models [70–72], in lattice simulation models [73, 74], off-lattice models [75, 76] of protein folding and protein-structure predictions [23–25].

The coupling strength  $J_{ij}$  between the two atoms can have variety of values. This is because there are  $n$  different kinds of atoms, so there are about  $n(n - 1)/2$  different kinds of interaction strengths at a particular distance between the two atoms. If we variate the distances, the number of interactions will be even more. On the other hand, the purpose of this paper is to study the ensemble of different ligands binding to the receptor. The ensemble of sequences is equivalent to the ensemble of interactions under the ergodicity hypothesis. In other words, changing the sequences of atoms is equivalent to changing the interactions. Therefore by shuffling sequences, one equivalently goes through different interactions. Based on this argument of the equivalence between many possibilities of interaction types due to the multiple types of atoms and ensemble of sequences of ligands, the coupling strengths  $J_{ij}$  can be seen to have a distribution. According to the large number theorem in statistics, this distribution is approximately Gaussian, that is:

$$f(J_{ij}) \sim \exp \left[ -\frac{(J_{ij} - \bar{J})^2}{2\Delta J^2} \right] \quad (11)$$

where  $\bar{J}$  is the average coupling strength and  $\Delta J^2$  is the variance. The distribution of the energy of the system can be obtained by calculating  $\langle \delta(E - H) \rangle$  where average is over interaction

coupling strengths  $J_{ij}$ . It is also Gaussian distributed:

$$f(E) \sim \exp \left[ -\frac{(E - \bar{E})^2}{2\Delta E^2} \right] \quad (12)$$

where  $\bar{E}$  is the average energy of the system and  $\Delta E^2 = N\Delta J^2$  is the variance.  $N$  is the total number of contacts. If we ignore the correlations between the different energy states, the result is an independent random energy model [36, 37].

## Simulations

While analytical model can give statistical features, we will further initiate a microscopic atomic detailed level investigation of the significance and implications of ligand binding with COX-2 as a model system. Initially a diversity set of 720 small molecules were selected from the NCI-Diversity database [49] having molecular weights similar to that of the reference compound SC-558, for which the crystal structure of the COX-2 complex is available (PDB code 1CX2) [50, 51]. All conformers of each of the 720 selected molecules were docked with COX-2 using AutoDock [52] to generate a thermodynamic binding energy landscape for each. From this data, the affinity and also intrinsic specificity quantified by ISR for each molecule with COX-2 were calculated as defined above and in Fig 2B.

To study the kinetics of ligand-receptor binding, we need to define an order parameter to describe the binding process on the statistical energy landscape. In this report, we use the RMSD as an order parameter (RMSD represents the root mean square distance relative to the native binding structure) that represents the progress of binding towards native state. For activation process, the order parameter or reaction coordinate RMSD is likely to be locally connected. It is straight-forward to see that the overall kinetic process involves diffusion in order parameter RMSD space. In other words, one can start with a general kinetic master equation, by assuming the local connectivity in RMSD, derive a diffusion equation [26–30, 32]:

$$\frac{\partial}{\partial t} P(\text{RMSD}, t) = \frac{\partial}{\partial \text{RMSD}} \left[ D(\text{RMSD}) \frac{\partial P(\text{RMSD}, t)}{\partial \text{RMSD}} + P \frac{\partial (F(\text{RMSD})/\kappa_B T)}{\partial \text{RMSD}} \right] \quad (13)$$

where  $P(\text{RMSD}, t)$  is the probability of the biomolecular binding complex with specific RMSD at time  $t$ ,  $D(\text{RMSD})$  is the diffusion coefficient and  $F(\text{RMSD})$  is the free energy of the system at RMSD. The diffusion coefficient is essentially the average time leaving the local minimum energy site. In terms of the order parameter RMSD, the problem becomes one dimensional diffusion. The diffusion equation can be integrated to give the mean first passage time:

$$\bar{\tau} = \int_{\text{RMSD}_i}^{\text{RMSD}_f} d\text{RMSD} \int_{\text{RMSD}_i}^{\text{RMSD}} d\text{RMSD}' \frac{\exp \left[ \frac{F(\text{RMSD}) - F(\text{RMSD}')}{\kappa_B T} \right]}{D(\text{RMSD})} \quad (14)$$

$\text{RMSD}_i \sim 1\text{\AA}$  is where the native binding state is. The boundary conditions for the Fokker-Planck equation are set as a reflecting one at  $\text{RMSD}_i$ , where system is in native state:  $[P(\text{RMSD}, t) \frac{\partial}{\partial \text{RMSD}} F(\text{RMSD}) + \frac{\partial}{\partial \text{RMSD}} P(\text{RMSD}, t)]|_{\text{RMSD}=\text{RMSD}_i} = 0$ , and an absorbing one at  $\text{RMSD} = \text{RMSD}_f$ , where the system is in the nonnative states:  $P(\text{RMSD}_f, t) = 0$ . This is for getting the off rate of binding. To get the on rate of binding, we just reverse the order of the reflecting (at non-native) and absorbing boundary condition (at native). The choice of an absorbing boundary condition facilitates the calculation for the first passage time and its distribution.

## Supporting Information

**S1 Fig. the relationship between the predicted affinity and specificity (ISR) for 720 drugs against the Cox-2.**

(EPS)

**S2 Fig. the relationship between the predicted affinity and the predicted kinetic specificity for 720 drugs against the Cox-2.**

(EPS)

**S1 Table. Selective (bold) and non-selective nonsteroidal anti-inflammatory drugs (NSAIDs) of COX-2.** The predicted affinity ( $E^{pred}$  (uM)) and residence time ( $Time_{off}^{pred}$ ) are shown for the drugs. The  $IC_{50}$  (uM) and the corresponding affinities ( $E^{exp}$ ) for 20 drugs and experimentally determined half life ( $= 0.693 * \text{residence time}$ ,  $Time_{off}^{exp}$  (hr)) for 22 drugs are also listed.

(DOC)

**S1 Text. The derivation for the free energy distribution.**

(DOCX)

**S2 Text. The fitting procedure for the simulation results.**

(DOC)

## Acknowledgments

XLZ would like to thank Professor Erkang Wang for the enlightening discussion.

## Author Contributions

Conceived and designed the experiments: JW. Performed the experiments: XZ JW. Analyzed the data: XZ JW. Contributed reagents/materials/analysis tools: XZ JW. Wrote the paper: XZ JW.

## References

1. Koshland D Jr (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci U S A 44: 98. doi: [10.1073/pnas.44.2.98](https://doi.org/10.1073/pnas.44.2.98) PMID: [16590179](https://pubmed.ncbi.nlm.nih.gov/16590179/)
2. McCammon JA (1998) Theory of biomolecular recognition. Curr Opin Struct Biol 8: 245–249. doi: [10.1016/S0959-440X\(98\)80046-8](https://doi.org/10.1016/S0959-440X(98)80046-8) PMID: [9631300](https://pubmed.ncbi.nlm.nih.gov/9631300/)
3. Cohen NC (1996) Guidebook on molecular modeling in drug design. Gulf Professional Publishing.
4. Wlodawer A, Erickson JW (1993) Structure-based inhibitors of hiv-1 protease. Annu Rev Biochem 62: 543–585. doi: [10.1146/annurev.bi.62.070193.002551](https://doi.org/10.1146/annurev.bi.62.070193.002551) PMID: [8352596](https://pubmed.ncbi.nlm.nih.gov/8352596/)
5. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267: 383–386. doi: [10.1126/science.7529940](https://doi.org/10.1126/science.7529940) PMID: [7529940](https://pubmed.ncbi.nlm.nih.gov/7529940/)
6. Cherfils J, Janin J (1993) Protein docking algorithms: simulating molecular recognition. Curr Opin Struct Biol 3: 265–269. doi: [10.1016/S0959-440X\(05\)80162-9](https://doi.org/10.1016/S0959-440X(05)80162-9)
7. Oshiro CM, Kuntz ID, Dixon JS (1995) Flexible ligand docking using a genetic algorithm. J Comput Aided Mol Des 9: 113–130. doi: [10.1007/BF00124402](https://doi.org/10.1007/BF00124402) PMID: [7608743](https://pubmed.ncbi.nlm.nih.gov/7608743/)
8. Xiang X, Sun X, Briceno G, Lou Y, Wang KA, et al. (1995) A combinatorial approach to materials discovery. Science: 1738–1738. doi: [10.1126/science.268.5218.1738](https://doi.org/10.1126/science.268.5218.1738) PMID: [17834993](https://pubmed.ncbi.nlm.nih.gov/17834993/)
9. Gallop MA, Barrett RW, Dower WJ, Fodor SP, Gordon EM (1994) Applications of combinatorial technologies to drug discovery. 1. background and peptide combinatorial libraries. J Med Chem 37: 1233–1251. doi: [10.1021/jm00035a001](https://doi.org/10.1021/jm00035a001) PMID: [8176700](https://pubmed.ncbi.nlm.nih.gov/8176700/)
10. Gordon EM, Barrett RW, Dower WJ, Fodor SP, Gallop MA (1994) Applications of combinatorial technologies to drug discovery. 2. combinatorial organic synthesis, library screening strategies, and future directions. J Med Chem 37: 1385–1401. doi: [10.1021/jm00036a001](https://doi.org/10.1021/jm00036a001) PMID: [8182695](https://pubmed.ncbi.nlm.nih.gov/8182695/)

11. Davidson AR, Sauer RT (1994) Folded proteins occur frequently in libraries of random amino acid sequences. *Proc Natl Acad Sci U S A* 91: 2146–2150. doi: [10.1073/pnas.91.6.2146](https://doi.org/10.1073/pnas.91.6.2146) PMID: [8134363](https://pubmed.ncbi.nlm.nih.gov/8134363/)
12. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262: 1680–1685. doi: [10.1126/science.8259512](https://doi.org/10.1126/science.8259512) PMID: [8259512](https://pubmed.ncbi.nlm.nih.gov/8259512/)
13. Lancet D, Sadovskiy E, Seidemann E (1993) Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system. *Proc Natl Acad Sci U S A* 90: 3715–3719. doi: [10.1073/pnas.90.8.3715](https://doi.org/10.1073/pnas.90.8.3715) PMID: [8475121](https://pubmed.ncbi.nlm.nih.gov/8475121/)
14. Freier SM, Konings DA, Wyatt JR, Ecker DJ (1995) Deconvolution of combinatorial libraries for drug discovery: a model system. *J Med Chem* 38: 344–352. doi: [10.1021/jm00002a016](https://doi.org/10.1021/jm00002a016) PMID: [7830277](https://pubmed.ncbi.nlm.nih.gov/7830277/)
15. Jin L, Wells JA (1994) Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. *Protein Science* 3: 2351–2357. doi: [10.1002/pro.5560031219](https://doi.org/10.1002/pro.5560031219) PMID: [7538848](https://pubmed.ncbi.nlm.nih.gov/7538848/)
16. Shi YY, Miller GA, Qian H, Bomsztyk K (2006) Free-energy distribution of binary protein–protein binding suggests cross-species interactome differences. *Proc Natl Acad Sci U S A* 103: 11527–11532. doi: [10.1073/pnas.0604316103](https://doi.org/10.1073/pnas.0604316103) PMID: [16861299](https://pubmed.ncbi.nlm.nih.gov/16861299/)
17. Miller GA, Shi YY, Qian H, Bomsztyk K (2007) Clustering coefficients of protein-protein interaction networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 75: 051910. doi: [10.1103/PhysRevE.75.051910](https://doi.org/10.1103/PhysRevE.75.051910) PMID: [17677101](https://pubmed.ncbi.nlm.nih.gov/17677101/)
18. Yook SH, Oltvai ZN, Barabási AL (2004) Functional and topological characterization of protein interaction networks. *Proteomics* 4: 928–942. doi: [10.1002/pmic.200300636](https://doi.org/10.1002/pmic.200300636) PMID: [15048975](https://pubmed.ncbi.nlm.nih.gov/15048975/)
19. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys Rev Lett* 90: 188101. doi: [10.1103/PhysRevLett.90.188101](https://doi.org/10.1103/PhysRevLett.90.188101) PMID: [12786043](https://pubmed.ncbi.nlm.nih.gov/12786043/)
20. Wang J, Zheng X, Yang Y, Drucekhammer D, Yang W, et al. (2007) Quantifying intrinsic specificity: a potential complement to affinity in drug screening. *Phys Rev Lett* 99: 198101. doi: [10.1103/PhysRevLett.99.198101](https://doi.org/10.1103/PhysRevLett.99.198101) PMID: [18233118](https://pubmed.ncbi.nlm.nih.gov/18233118/)
21. Yan Z, Zheng X, Wang E, Wang J (2013) Thermodynamic and kinetic specificities of ligand binding. *Chem Sci* 4: 2387–2395. doi: [10.1039/c3sc50478f](https://doi.org/10.1039/c3sc50478f)
22. Thornton JM, Orengo CA, Todd AE, Pearl FM (1999) Protein folds, functions and evolution. *J Mol Biol* 293: 333–342. doi: [10.1006/jmbi.1999.3054](https://doi.org/10.1006/jmbi.1999.3054) PMID: [10529349](https://pubmed.ncbi.nlm.nih.gov/10529349/)
23. Goldstein RA, Luthey-Schulten ZA, Wolynes PG (1992) Optimal protein-folding codes from spinglass theory. *Proc Natl Acad Sci U S A* 89: 4918–4922. doi: [10.1073/pnas.89.11.4918](https://doi.org/10.1073/pnas.89.11.4918) PMID: [1594594](https://pubmed.ncbi.nlm.nih.gov/1594594/)
24. Shakhnovich EI, Gutin AM (1993) Engineering of stable and fast-folding sequences of model proteins. *Proc Natl Acad Sci U S A* 90: 7195–7199. doi: [10.1073/pnas.90.15.7195](https://doi.org/10.1073/pnas.90.15.7195) PMID: [8346235](https://pubmed.ncbi.nlm.nih.gov/8346235/)
25. Klimov D, Thirumalai D (1996) Criterion that determines the foldability of proteins. *Phys Rev Lett* 76: 4070. doi: [10.1103/PhysRevLett.76.4070](https://doi.org/10.1103/PhysRevLett.76.4070) PMID: [10061184](https://pubmed.ncbi.nlm.nih.gov/10061184/)
26. Wang J, Huang W, Lu H, Wang E (2004) Downhill kinetics of biomolecular interface binding: globally connected scenario. *Biophys J* 87: 2187–2194. doi: [10.1529/biophysj.104.042747](https://doi.org/10.1529/biophysj.104.042747) PMID: [15454421](https://pubmed.ncbi.nlm.nih.gov/15454421/)
27. Wang J (2006) Diffusion and single molecule dynamics on biomolecular interface binding energy landscape. *Chem Phys Lett* 418: 544–548. doi: [10.1016/j.cplett.2006.02.010](https://doi.org/10.1016/j.cplett.2006.02.010)
28. Lee CL, Stell G, Wang J (2002) First-passage time distribution and non-markovian diffusion dynamics of protein folding. *J Chem Phys* 118: 959–968. doi: [10.1063/1.1527672](https://doi.org/10.1063/1.1527672)
29. Lee CL, Lin CT, Stell G, Wang J (2003) Diffusion dynamics, moments, and distribution of firstpassage time on the protein-folding energy landscape, with applications to single molecules. *Phys Rev E* 67: 41905–41909. doi: [10.1103/PhysRevE.67.041905](https://doi.org/10.1103/PhysRevE.67.041905)
30. Wang J (2003) Statistics, pathways and dynamics of single molecule protein folding. *J Chem Phys* 118: 952–958. doi: [10.1063/1.1527633](https://doi.org/10.1063/1.1527633)
31. Zhou Y, Zhang C, Stell G, Wang J (2003) Temperature dependence of the distribution of the first passage time: results from discontinuous molecular dynamics simulations of an all-atom model of the second beta-hairpin fragment of protein g. *J Am Chem Soc* 125: 6300–6305. doi: [10.1021/ja029855x](https://doi.org/10.1021/ja029855x) PMID: [12785863](https://pubmed.ncbi.nlm.nih.gov/12785863/)
32. Wang J, Xu L, Xue K, Wang E (2008) Exploring the origin of power law distribution in singlemolecule conformation dynamics: energy landscape perspectives. *Chem Phys Lett* 463: 405–409. doi: [10.1016/j.cplett.2008.08.081](https://doi.org/10.1016/j.cplett.2008.08.081)
33. Derrida B (1981) Random-energy model: An exactly solvable model of disordered systems. *Phys Rev B* 24: 2613–2626. doi: [10.1103/PhysRevB.24.2613](https://doi.org/10.1103/PhysRevB.24.2613)
34. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* 84: 7524–7528. doi: [10.1073/pnas.84.21.7524](https://doi.org/10.1073/pnas.84.21.7524) PMID: [3478708](https://pubmed.ncbi.nlm.nih.gov/3478708/)

35. Reichl LE (1984) *A Modern Course in Statistical Physics*. University of Texas Press.
36. Gardner E, Derrida B (1989) The probability distribution of the partition function of the random energy model. *J Phys A Math Gen* 22: 1975. doi: [10.1088/0305-4470/22/12/003](https://doi.org/10.1088/0305-4470/22/12/003)
37. Mézard M, Parisi G, Virasoro M (1985) Random free energies in spin glasses. *Journal de Physique Lettres* 46: 217–222. doi: [10.1051/jphyslet:01985004606021700](https://doi.org/10.1051/jphyslet:01985004606021700)
38. Derrida B (1980) Random-energy model: Limit of a family of disordered models. *Phys Rev Lett* 45: 79–82. doi: [10.1103/PhysRevLett.45.79](https://doi.org/10.1103/PhysRevLett.45.79)
39. Wang J, Lee C, Stell G (2005) The cooperative nature of hydrophobic forces and protein folding kinetics. *Chem Phys* 316: 53–60. doi: [10.1016/j.chemphys.2005.04.045](https://doi.org/10.1016/j.chemphys.2005.04.045)
40. Gutin A, Abkevich V, Shakhnovich E (1995) Evolution-like selection of fast-folding model proteins. *Proc Natl Acad Sci U S A* 92: 1282–1286. doi: [10.1073/pnas.92.5.1282](https://doi.org/10.1073/pnas.92.5.1282) PMID: [7877968](https://pubmed.ncbi.nlm.nih.gov/7877968/)
41. Govindarajan S, Goldstein RA (1996) Why are some proteins structures so common? *Proc Natl Acad Sci U S A* 93: 3341–3345. doi: [10.1073/pnas.93.8.3341](https://doi.org/10.1073/pnas.93.8.3341) PMID: [8622938](https://pubmed.ncbi.nlm.nih.gov/8622938/)
42. Yang H, Luo G, Karnchanaphanurach P, Louie TM, Rech I, et al. (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science* 302: 262–266. doi: [10.1126/science.1086911](https://doi.org/10.1126/science.1086911) PMID: [14551431](https://pubmed.ncbi.nlm.nih.gov/14551431/)
43. Lu HP, Xun L, Xie XS (1998) Single-molecule enzymatic dynamics. *Science* 282: 1877–1882. doi: [10.1126/science.282.5395.1877](https://doi.org/10.1126/science.282.5395.1877) PMID: [9836635](https://pubmed.ncbi.nlm.nih.gov/9836635/)
44. Sunney Xie X (2002) Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics. *J Chem Phys* 117: 11024–11032. doi: [10.1063/1.1521159](https://doi.org/10.1063/1.1521159)
45. Yang H, Xie XS (2002) Probing single-molecule dynamics photon by photon. *J Chem Phys* 117: 10965–10979. doi: [10.1063/1.1521154](https://doi.org/10.1063/1.1521154)
46. Yang H, Xie X (2002) Statistical approaches for probing single-molecule dynamics photon-by-photon. *Chem Phys* 284: 423–437. doi: [10.1016/S0301-0104\(02\)00672-9](https://doi.org/10.1016/S0301-0104(02)00672-9)
47. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Observation of a power-law memory kernel for fluctuations within a single protein molecule. *Phys Rev Lett* 94: 198302. doi: [10.1103/PhysRevLett.94.198302](https://doi.org/10.1103/PhysRevLett.94.198302) PMID: [16090221](https://pubmed.ncbi.nlm.nih.gov/16090221/)
48. Min W, English BP, Luo G, Cherayil BJ, Kou SC, et al. (2005) Fluctuating enzymes: lessons from single-molecule studies. *Acc Chem Res* 38: 923–931. doi: [10.1021/ar040133f](https://doi.org/10.1021/ar040133f) PMID: [16359164](https://pubmed.ncbi.nlm.nih.gov/16359164/)
49. Nci. <http://dtpncinihgov/branches/dscb/diversityexplanationhtml>.
50. Kurumbail RG, Stevens AM, Gierse JK, McDonald JJ, Stegeman RA, et al. (1996) Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* 384: 644–648. doi: [10.1038/384644a0](https://doi.org/10.1038/384644a0) PMID: [8967954](https://pubmed.ncbi.nlm.nih.gov/8967954/)
51. DeWitt DL (1999) Cox-2-selective inhibitors: the new super aspirins. *Mol Pharmacol* 55: 625–631. PMID: [10101019](https://pubmed.ncbi.nlm.nih.gov/10101019/)
52. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of autodock. *J Mol Rec* 9: 1–5. doi: [10.1002/\(SICI\)1099-1352\(199601\)9:1%3C1::AID-JMR241%3E3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-1352(199601)9:1%3C1::AID-JMR241%3E3.0.CO;2-6)
53. Copeland RA, Pompliano DL, Meek TD (2006) Drug–target residence time and its implications for lead optimization. *Nat Rev Drug Discov* 5: 730–739. doi: [10.1038/nrd2082](https://doi.org/10.1038/nrd2082) PMID: [16888652](https://pubmed.ncbi.nlm.nih.gov/16888652/)
54. Lu H, England K, am Ende C, Truglio JJ, Luckner S, et al. (2009) Slow-onset inhibition of the fabi enoyl reductase from francisella tularensis: residence time and in vivo activity. *ACS Chem Biol* 4: 221–231. doi: [10.1021/cb800306y](https://doi.org/10.1021/cb800306y) PMID: [19206187](https://pubmed.ncbi.nlm.nih.gov/19206187/)
55. Lancet D, Horovitz A, Katchalski-Katzir E (1994) Molecular recognition in biology: models for analysis of protein-ligand interactions. In: *The lock-and-key principle*, John Wiley Sons, Ltd. pp. 25–71.
56. Mouthon L, Lacroix-Desmazes S, Nobrega A, Barreau C, Coutinho A, et al. (1996) The selfreactive antibody repertoire of normal human serum igm is acquired in early childhood and remains conserved throughout life. *Scand J Immunol* 44: 243–251. doi: [10.1046/j.1365-3083.1996.d01-306.x](https://doi.org/10.1046/j.1365-3083.1996.d01-306.x) PMID: [8795718](https://pubmed.ncbi.nlm.nih.gov/8795718/)
57. Nobrega A, Haury M, Grandien A, Malanchère E, Sundblad A, et al. (1993) Global analysis of antibody repertoires. ii. evidence for specificity, self-selection and the immunological “homunculus” of antibodies in normal serum. *Eur J Immunol* 23: 2851–2859. doi: [10.1002/eji.1830231119](https://doi.org/10.1002/eji.1830231119) PMID: [8223861](https://pubmed.ncbi.nlm.nih.gov/8223861/)
58. Tanaka MM, Sisson SA, King GC (2009) High affinity extremes in combinatorial libraries and repertoires. *J Theor Biol* 261: 260–265. doi: [10.1016/j.jtbi.2009.07.041](https://doi.org/10.1016/j.jtbi.2009.07.041) PMID: [19665466](https://pubmed.ncbi.nlm.nih.gov/19665466/)
59. Ji HF, Kong DX, Shen L, Chen LL, Ma BG, et al. (2007) Distribution patterns of small-molecule ligands in the protein universe and implications for origin of life and drug discovery. *Genome Biol* 8: R176. doi: [10.1186/gb-2007-8-8-r176](https://doi.org/10.1186/gb-2007-8-8-r176) PMID: [17727706](https://pubmed.ncbi.nlm.nih.gov/17727706/)

60. Levitan B (1997) Models and search strategies for applied molecular evolution. In: Annual Reports in Combinatorial Chemistry and Molecular Diversity, Springer. pp. 95–152.
61. Levitan B (1998) Stochastic modeling and optimization of phage display. *J Mol Biol* 277: 893–916. doi: [10.1006/jmbi.1997.1555](https://doi.org/10.1006/jmbi.1997.1555) PMID: [9545380](https://pubmed.ncbi.nlm.nih.gov/9545380/)
62. Mandecki W, Chen YC, Grihalde N (1995) A mathematical model for biopanning (affinity selection) using peptide libraries on filamentous phage. *J Theor Biol* 176: 523–530. doi: [10.1006/jtbi.1995.0218](https://doi.org/10.1006/jtbi.1995.0218) PMID: [8551746](https://pubmed.ncbi.nlm.nih.gov/8551746/)
63. Vant-Hull B, Payano-Baez A, Davis RH, Gold L (1998) The mathematics of selex against complex targets. *J Mol Biol* 278: 579–597. doi: [10.1006/jmbi.1998.1727](https://doi.org/10.1006/jmbi.1998.1727) PMID: [9600840](https://pubmed.ncbi.nlm.nih.gov/9600840/)
64. Chen CK, Kuo TC (2007) Simulations of selex against complex receptors with a condensed statistical model. *Comput Chem Eng* 31: 1007–1019. doi: [10.1016/j.compchemeng.2006.08.015](https://doi.org/10.1016/j.compchemeng.2006.08.015)
65. Kauvar LM, Higgins DL, Villar HO, Sportsman JR, Engqvist-Goldstein Å, et al. (1995) Predicting ligand binding to proteins by affinity fingerprinting. *Chem Biol* 2: 107–118. doi: [10.1016/1074-5521\(95\)90283-X](https://doi.org/10.1016/1074-5521(95)90283-X) PMID: [9383411](https://pubmed.ncbi.nlm.nih.gov/9383411/)
66. Young SS, Sheffield CF, Farmen M (1997) Optimum utilization of a compound collection or chemical library for drug discovery. *J Chem Inf Comput Sci* 37: 892–899. doi: [10.1021/ci970224+](https://doi.org/10.1021/ci970224+)
67. Yan Z, Wang J (2013) Optimizing scoring function of protein-nucleic acid interactions with both affinity and specificity. *PLoS One* 8: e74443. doi: [10.1371/journal.pone.0074443](https://doi.org/10.1371/journal.pone.0074443) PMID: [24098651](https://pubmed.ncbi.nlm.nih.gov/24098651/)
68. Yan Z, Guo L, Hu L, Wang J (2013) Specificity and affinity quantification of protein-protein interactions. *Bioinformatics* 29: 1127–1133. doi: [10.1093/bioinformatics/btt121](https://doi.org/10.1093/bioinformatics/btt121) PMID: [23476023](https://pubmed.ncbi.nlm.nih.gov/23476023/)
69. Chu X, Wang J (2014) Specificity and affinity quantification of flexible recognition from underlying energy landscape topography. *PLoS Comput Biol* 10: e1003782. doi: [10.1371/journal.pcbi.1003782](https://doi.org/10.1371/journal.pcbi.1003782) PMID: [25144525](https://pubmed.ncbi.nlm.nih.gov/25144525/)
70. Plotkin SS, Wang J, Wolynes PG (1996) Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E* 53: 6271. doi: [10.1103/PhysRevE.53.6271](https://doi.org/10.1103/PhysRevE.53.6271)
71. Plotkin SS, Wang J, Wolynes PG (1997) Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J Chem Phys* 106: 2932–2948. doi: [10.1063/1.473355](https://doi.org/10.1063/1.473355)
72. Saven JG, Wolynes PG (1997) Statistical mechanics of the combinatorial synthesis and analysis of folding macromolecules. *J Phys Chem B* 101: 8375–8389. doi: [10.1021/jp971707j](https://doi.org/10.1021/jp971707j)
73. Succi ND, Onuchic JN (1994) Folding kinetics of proteinlike heteropolymers. *J Chem Phys* 101: 1519–1528. doi: [10.1063/1.467775](https://doi.org/10.1063/1.467775)
74. Succi N, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104: 5860–5868. doi: [10.1063/1.471317](https://doi.org/10.1063/1.471317)
75. Boczek EM, Brooks C (1995) First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* 269: 393–396. doi: [10.1126/science.7618103](https://doi.org/10.1126/science.7618103) PMID: [7618103](https://pubmed.ncbi.nlm.nih.gov/7618103/)
76. Guo Z, Brooks CL, Boczek EM (1997) Exploring the folding free energy surface of a three-helix bundle protein. *Proc Natl Acad Sci U S A* 94: 10161–10166. doi: [10.1073/pnas.94.19.10161](https://doi.org/10.1073/pnas.94.19.10161) PMID: [9294180](https://pubmed.ncbi.nlm.nih.gov/9294180/)