PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Universal Statistical Fluctuations in Thermodynamics and Kinetics of Single Molecule Recognition†

Xiliang Zheng,[a] Jin Wang,[ab*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

We investigated the main universal statistical distributions of the single molecular recognition. The distributions of the single molecule binding free energy spectrum or density of states were characterized in the ligand-receptor binding energy landscape. The analytical results are consistent with the microscopic molecular simulations. The free energy distribution of different binding modes or states for a single molecule ligand receptor pair is approximately Gaussian near the mean and exponential at the tail. The equilibrium constant of a single molecule binding is log-normal distributed near the mean and power law distributed near the tail. Additionally, we found that the kinetics distribution of a single molecule ligand binding can be characterized by log-normal around the mean and power law distribution near the tail. This distribution is caused by exploration of the underlying inhomogeneous free energy landscape. Different ligand-receptor binding complexes have the same universal form of distribution but differ in parameters.

1 Introduction

Understanding how two biomolecules recognize each other is a fundamental issue of molecular biology[1,2]. Meanwhile, it is also a central topic in drug design and pharmaceutical industry[3]. The affinity and the specificity are two major issues related to biomolecular binding. The former represents the binding strength and stability of binding complexes. The latter is used to discriminate the specific binding complexes from the non-specific ones[4,5] even if they have roughly the same stability (affinity). From microscopic point of view, the binding complex such as a particular ligand receptor pair can be seen as a network of atoms or residues interacting with each other. The binding process involves many possible binding modes or conformational states with the corresponding binding free energy. Thus a free energy spectrum for those binding modes or states is emerged from a particular pair of ligand receptor binding1.00,1.00,1.00rgb]1.00,0.00,0.00, 1.00,1.00,1.00rgb]1.00,0.00,0.00the free energy of native binding mode gives the affinity of the binding. For different ligands, we naturally have different affinities as well as free energy binding spectrums for the same receptor. Although affinity of the binding is often studied, the statistical nature of the binding free energy spectrum has not been fully explored yet[6]for a ligand receptor pair. It is very interesting to see how the statistical distributions for the binding free energy spectrum change with different ligands targeting the same receptor.

In addition to the requirement of the affinity for binding, the intrinsic specificity discriminating different modes of binding should be taken into consideration[6,7].(See panel B of Figure1). The high intrinsic specificity leads to funneled binding energy landscape, which guarantees the stability and specificity during the binding process.

1.00,1.00,1.00rgb]1.00,0.00,0.00Our studies in this work are based on the folding and binding energy landscape theory[7–16]. The former had been proposed and proved to be successful for almost thirty years[8–11,17],which assumed that the protein folding progresses through multiple pathways rather than a single one towards the bottom of the funnel. The decrease in total energy of the system with reducing conformational space naturally leads to the underlying funneled energy landscape of protein folding. For molecular recognition, the direction of spontaneous association for the system of receptor-ligand requires lowering the free energy with the reduction in the conformational, rotational/translational entropy of the receptor, ligand and solvent complex. Thus the binding energy landscape of the receptor-ligand system should also be funnel-shaped, mathematically similar to folding. For folding, because the solvent molecules in the environment squeeze the single polypeptide chain to collapse into molten globule intermediates, the protein is subject to the fast hydrophobic collapse powered by the entropy of the system. Analogously, for binding, the initial collisions or interactions between the receptor and the ligand will inevitably strip water/solvent

[a] State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China.
[b] Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, NY 11794-3400, USA. Fax: +123 (0)123 4445557; Tel: +123 (0)123 4445556; E-mail: jin.wang.1@stonybrook.edu

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

molecules from the receptor-ligand interfaces, with the release of the solvent molecules into the environment and the increase of the solvent entropy. Therefore, this process is likewise driven by entropy of the system, which mostly occurs in the lock and key model/scenario.

1.00,1.00,1.00rgb]1.00,0.00,0.00However, due to the lack of the chain connectivity in the intermolecular binding, the number of conformations in binding is substantially larger than in folding. Binding funnels can therefore be more complicated. Of particular note is the configurational entropy. For binding, the slope of the entropy is subject to be reduced under the conditions of the geometrical constraints. The entropy has certain quantitative differences compared to folding in regard to chain connectivity, rotation and translation of the system of interest. In addition, the implications of folding funnel concept for protein function are inherently limited. This is because stability may not be the ultimate goal for molecular evolution. The goal should be to optimize the function. The function is realized at the molecular level by the binding or recognition among biomolecules. Although mathematically similar to folding funnel, the binding funnel emphasizes a different concept[6,18], the optimization of function. In fact, individual proteins may not be funneled as evidenced by the presence of intrinsic disordered proteins. However, the binding landscape should still be funneled. Therefore, biology may require the binding funnel for function but not necessarily the individual folding funnel in some cases. This is to say, biology may prefer the function rather than the stability. The molecular evolution may bias towards binding since function is realized by the recognition. While folding funnel may emerge from enzymatic and metabolic proteins, the binding funnel may apply to wider systems including signal transduction and gene regulatory proteins, which often do not have well defined structures alone and therefore no folding funnels.

1.00,1.00,1.00rgb]1.00,0.00,0.00Recently, there are growing evidences of the hydrophobic interactions being a main driving force during folding and binding processes even though it is not necessarily dominant for the binding[19–22]. The long-range electrostatic interactions are also expected to contribute to the binding process significantly while not at all in folding. For the recognition, there exists a balance between the function of biomolecules with more hydrophobic residues on the surface of biomolecules (binding) and the self-stability of the biomolecules with more hydrophobic residues located inside (folding).

1.00,1.00,1.00rgb]1.00,0.00,0.00For receptor-ligand binding, the hydrophobic and electrostatic interactions drive the process and lead to the energy reduction, while searching for the native destiny requires the entropy reduction. These two trends compensate to each other, leading to a possibility of the emergence of two low free energy states: non-native unbinding state/native binding one. This is evidenced by thermodynamic heat capacity (peak) and kinetic measurements of many ligand-receptor (protein) binding complexes[23–26]. However, traps may occur and become significant. In those cases, the free energy is not self-averaging. The statistical fluctuations and whole distributions of the free energy become important in characterizing the traps. This is the focus of this study on the statistical nature of the thermodynamics and kinetics.

1.00,1.00,1.00rgb]1.00,0.00,0.00The binding can be characterized by the extent of the roughness and the slope as well as the size of the binding funnel. And, the binding funnel is more relevant to the binding specificity rather than the stability (or affinity) as in the folding. For example, a funneled binding landscape implies a differentiation of native binding state (mode) from the non-native binding states (modes). Here the relative affinity (specificity) rather than absolute affinity is closely tied up with the degree of binding funnel. Furthermore, the interplay of the ruggedness and the slope of the funneled energy landscape, which depicts the stability and dynamic behavior of the individual proteins, can also be used to characterize the intrinsic specificity of binding[16]. The tendency to maximize the ratio of the slope to the roughness of the energy landscape taking the entropy of the system into account, can be viewed as the criterion for the specificity of acting on the binding affinity and kinetics.

It is noteworthy to point out the effect of the intrinsic specificity on how binding process actually occurs. Since the energy landscape of ligand receptor recognition can be mapped out, the kinetics during the binding process can then be explored on the energy landscape through analytical and microscopic methods taking global and local connectivity into account[27,28]. The intrinsic specificity has been demonstrated to be the key factor determining the speed of molecular recognition[16].

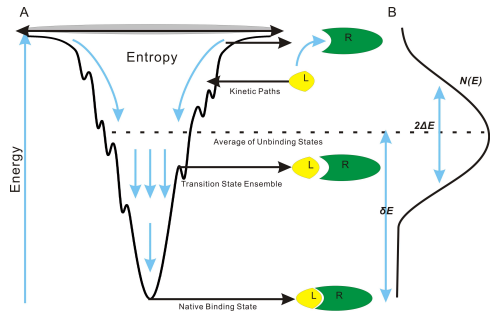

Fig. 1 The energy landscape of the biomolecular binding with a funnel. panel (A) shows the binding energy landscape with a funnel-like shape towards the native binding state. The multiple kinetic paths along the energy landscape are shown with the arrows. The cartoon showings of receptor(green)/ligand(yellow)complexes correspond to the different binding states of binding. panel (B) shows the density of states of the binding energy landscape. δE and ΔE are shown respectively.

Although the average-level description of the kinetics in the bulk is of importance, it can occasionally miss important features of the dynamic binding process that are critical for uncovering the fundamental mechanisms. The progression of conformational dynamics of biomolecules is expected to traverse the underlying energy landscape with different local barriers, from local to global. Naturally, there co-exists many possible time scales in the process, thus the kinetics can be the non-exponential. In addition, for bulk measurements on kinetics, it is often hard to determine whether the observed non-exponential is intrinsic or resulted from the inhomogeneity of single exponential processes. With recent advances on measurement techniques, single-molecule detections become possible and more mature[29–31]. Single molecules are the

probes with high sensitivity to the local micro-environments and therefore provide an ideal tool to understand the delicate structures of the energy landscape of the biomolecules[6,32–36]. And of late, numerous impressive single-molecule conformational dynamics experiments have been successfully undertaken[31,37–41]. Particularly, the power law decay kinetics has been increasingly observed in single molecule conformational dynamics[30,42–48], but how to reasonably interpret the underlying origin of this complicated kinetics is nontrivial. As we know, statistical fluctuations of single molecules are intrinsic. Fortunately, these fluctuations can now be directly measured rather than being weighted down by a great number of molecules in the bulk. Therefore, the average kinetic description is no longer valid and subject to be replaced by the probabilistic description to characterize the fluctuation natures of the single molecule dynamics. In this paper, we provide a probabilistic description of the single molecule binding kinetics of conformational dynamics through a diffusion approach along the free energy landscape from the results of the microscopic simulations of a specific receptor-ligand pair. We find that the single molecule binding kinetics is log-normal distributed around the mean and power law distributed near the tail. Such statistical behavior of the kinetics has been confirmed by single molecule experiments[30,42–47]. In addition, power law kinetics can also give clues of the density of states(DOS) of the free energy landscape being exponentially distributed[27,28,35,49–53].

It is worthwhile to note that classical biochemistry and chemistry bulk experiments can give statistical distributions of intermolecular recognition from collecting statistics of different pairs of molecules[54]. In this way, we explore the recognition from different sequences. On the other hand, we can probe the recognition also through the contact interactions for a given sequence. For example, for single molecule binding, the recognition process goes through different conformations for a single molecule binding pair. The distribution of the recognition collected from the different conformations in this way gives us the statistical information of the underlying intrinsic binding energy landscape. When the recognition complex is large enough, one expects probing the recognition from the ensemble of different sequences and from ensemble of different contacts or conformations for a given sequence would be equivalent (throwing many dices at the same time versus throwing one dice many times). In practice, for finite size complexes, these two distributions would not have perfect correlations. They give statistical information on the recognition with different perspectives. In this study, we focus on the statistics of conformational dynamics in a single molecule binding for given binding pairs without changing on sequences. We leave the discussion for its connection to the statistics of recognition for different sequences to later study.

2 Materials and Methods

2.1 Energy function of the binding

Firstly, we can obtain the free energy based on the contact variable as the order parameter, which can be directly delineated as the contact probability during the atomic contact space constituted by all atoms of the receptor and a small ligand molecule: σ_{ij} (σ_{ij}=1 if $d < \Delta$ and σ_{ij}=0 if $d > \Delta$) where Δ is a predefined cutoff distance about several angstrom while 1.00,1.00,1.00rgb]1.00,0.00,0.00d is the distance between two atoms which are from the receptor and the ligand, respectively. 1.00,1.00,1.00rgb]1.00,0.00,0.00In this work, we introduce the contact Hamiltonian to determine the energy of the system for simplicity. The determination of the cutoff distance will certainly have some effects on the interactions of interest. However, we do not expect the statistical distributions of the interactions are dependent of the predefined cutoff distance Δ unless given an unreasonable large value.

Thus the energy function of a ligand-receptor system according to the variables can be expressed as:

$$H = \sum_{ij} J_{ij}\sigma_{ij} \tag{1}$$

whereas J_{ij} is the interatomic coupling strength occurring between one atom of a receptor and another one of a ligand. Here, although the form of interactions is relatively simple, we can still obtain the general features of the system of interest via this concise form.

This interaction form aforementioned has been widely applied to many studies. Several models were derived from this form, such as analytical models[55–57], lattice simulation models[58,59], off-lattice models[60,61] of protein folding and protein-structure predictions[62–64].

Naturally, the single molecule binding between the pair of ligand and receptor explores a vast number of different values of the coupling strength J_{ij} for various distances between the atoms through contact interactions. The coupling strength J_{ij} due to the multiplicity will be expected to have a statistical distribution. 1.00,1.00,1.00rgb]1.00,0.00,0.00Here, the coupling strengths between the atomic pairs J_{ij} is assumed to be Gaussianly distributed.

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

where $\bar{J}$ represents the average of coupling strengths and ΔJ^2 is the corresponding variance. Then we can derive the energy distribution of the system by calculating $< \delta(E-H) >$, wherein the average is over interaction coupling strengths J_{ij}. It is also Gaussian distributed:

$$f(E) \sim exp[-\frac{(E-\bar{E})^2}{2\Delta E^2}] \tag{3}$$

where $\bar{E}$ means the average energy of the system and $\Delta E^2 = N\Delta J^2$ is the variance of the energy. 1.00,1.00,1.00rgb]1.00,0.00,0.00N means total number of contacts. 1.00,1.00,1.00rgb]1.00,0.00,0.00The central limit theorem can be applied to the total energy which is the sum of the J_{ij} as defined in eq. 1. Since the coupling J_{ij} can have different values for different types of atomic pairs and different distances for the same type of atomic interactions, the coupling J_{ij} can be taken as random variables. Due to the large number of the possible couplings, from central limit theorems, the energy from the large sum of random coupling variables is expected to have a Gaussian distribution. For simplicity of the analytical

derivations here, we assume the J_{ij} is Gaussianly distributed. However, we must notice that the final conclusion of this work on the distribution of energy, free energy, equilibrium constant and kinetics do not dependent on this assumption of coupling strength being Gaussinaly distributed. This is because the energy follows the Gaussian distribution by the central limit theorem irrespect whether the individual component J_{ij} in the sum is Gaussianly distributed or not. The distributions of other quantities mentioned are derived from the Gaussian distribution of the energy. Here, the independent random energy model is applied[49,50] without regard to the potential correlations between the diverse energy states. Note that in the distribution of energy, the native (strongest) binding state is supposed to appear in the low end of the tail (the extreme left value) where the density of binding states has become discrete. The association between the receptor and the ligand occurs along the surface of the receptor, and this binding is spontaneous. We call it the binding complex. There are many binding modes with different ligand binding conformations. The free energy of binding for each binding mode can be gained. The binding mode of the lowest free energy is defined as the native state or native binding mode. 1.00,1.00,1.00rgb]1.00,0.00,0.00All other binding modes will be viewed as the non-native states.

2.2 Kinetics of the binding

In the kinetic study, RMSD was used as an order parameter that portrayed the process of binding by approaching the global minimum or native state. We can start from a general kinetic master equation, considering the local connectivity in reaction coordinates or order parameters, derive a diffusion equation[8,27,28,35,51–53]:

$$\frac{\partial}{\partial t}P(RMSD,t)=\frac{\partial}{\partial RMSD}[D(RMSD)\frac{\partial P(RMSD,t)}{\partial RMSD}+P\frac{\partial(F(RMSD)/\kappa_B T)}{\partial RMSD}] \quad (4)$$

where 1.00,1.00,1.00rgb]1.00,0.00,0.00$P(RMSD,t)$ represents the probability of the binding complex with a specific RMSD at time scale t, 1.00,1.00,1.00rgb]1.00,0.00,0.00$D(RMSD)$ corresponds to the diffusion coefficient and 1.00,1.00,1.00rgb]1.00,0.00,0.00$F(RMSD)$ represents the free energy of the system of interest at the aforementioned RMSD. 1.00,1.00,1.00rgb]1.00,0.00,0.00Here, K_B is set to 1 for simplicity. In essence, the diffusion coefficient can be viewed as the mean time leaving the local minimum binding state or energy site. According to the order parameter RMSD, the problem is converted into one dimensional diffusion. The diffusion equation is integrated to gain the mean first passage time:

$$\bar{\tau}=\int_{RMSD_i}^{RMSD_f}dRMSD\int_{RMSD_i}^{RMSD}dRMSD'\frac{exp[\frac{F(RMSD)-F(RMSD')}{\kappa_B T}]}{D(RMSD)} \quad (5)$$

$RMSD_i \sim 1A$ means where the global minimum or native binding state is reached. The boundary conditions of the Fokker-Planck equation are here set as a reflecting one for the specific $RMSD_i$, where the system of interest is in native state:$[P(RMSD,t)\frac{\partial}{\partial RMSD}F(RMSD)+\frac{\partial}{\partial RMSD}P(RMSD,t)]|_{RMSD=RMSD_i}=0$, and an absorbing one while $RMSD=RMSD_f$, where the system of interest is in the non-native states: $P(RMSD_f,t)=0$. This boundary condition is for obtaining the off time kinetics while the reverse order of which is for obtaining the on time kinetics. Here, we choose an absorbing boundary condition during the calculations to facilitate to access the first passage time and its corresponding 1.00,1.00,1.00rgb]1.00,0.00,0.00FPT distribution for a specific ligand binding with a receptor.

1.00,1.00,1.00rgb]1.00,0.00,0.00The calculations of MFPTs can be carried out by setting the D(RMSD)=1,K_B=1 of each (individual MFPT calculation for a molecular recognition complex) for comparing with each other (MFPTs from different molecular recognition pair complexes). The conformations or poses are generated by molecular simulations (here via molecular docking). The RMSDs relative to the native binding state are collected from the docking simulations registering the coordinates of each conformation state. The free energy profiles in terms of RMSD are given by the docking scoring function at each conformation. It is worthwhile to note that, although the generated conformations have not been relaxed such as molecular dynamics or other methods and have not been evaluated by other computationally expensive force fields, due to the large number theorem and the central limit theorem in statistics, the statistical distributions of the large number of interactions occurring in the single molecular recognition should have the same form in principal and be independent of sampling methods and score functions. We then use the free energy profile generated by Autodock (for the details of the Autodock score function, See ESI†) in RMSD to calculate the MFPT according to the equation 5.

1.00,1.00,1.00rgb]1.00,0.00,0.00We can also get the following relation for the mean first passage time distribution $P_{FPT}(\tau)$ 1.00,1.00,1.00rgb]1.00,0.00,0.00

$$P_{FPT}(\tau)=\frac{d}{d\tau}(1-\Sigma\tau)=-\frac{d\Sigma\tau}{d\tau} \quad (6)$$

1.00,1.00,1.00rgb]1.00,0.00,0.00where $\Sigma\tau\equiv\int_0^{\rho_f}d\rho G(\rho,\tau)$, the $G(\rho,\tau)$ represents the function of probability distribution for the ligand-receptor pair binding complex at time τ. We can obtain the $P_{FPT}(\tau)$ by solving the $\tilde{P}_{FPT}(s)$, which is Laplace transforms of $P_{FPT}(\tau)$. 1.00,1.00,1.00rgb]1.00,0.00,0.00The results show that for $T<T_c$ (near or below T_c) $\tilde{P}_{FPT}(s)$ is approximately described as $\tilde{P}_{FPT}(s)\approx e^{-cs\alpha}$. Then by the transform, we can obtain $P_{FPT}(\tau)\sim\tau^{-(1+\alpha)}$ $(0<\alpha<1)$, where $\alpha=T/T_c$[51,53](For details, See ESI†).

2.3 Simulations

After the universal statistical features were predicted using the analytical models, we then initiated a microscopic investigation of the implications of specific ligands binding to COX-2 as a system of interest to validate the analytical predictions. Initially a diverse set of 720 small molecules was chosen out from the NCI Diversity set having similar molecular weights to the reference compound SC-558, for which the three-dimensional structure of the COX-2 complex is accessible (PDB ID: 1CX2)[65,66]. Each of these 720 selected molecules was docked into COX-2 using the AutoDock package[67], all resulting conformers of each ligand por-

tray a thermodynamic binding energy landscape for the receptor-ligand pair. Next, the intrinsic specificity quantified by intrinsic specificity ratio (ISR) representing the ratio of free energy gap between the native binding mode and the average of all other binding modes versus the roughness (characterized by the spread or variance) of the underlying free energy landscape for each a specific ligand binding to COX-2 was gained as mentioned above and in Figure 1B.

3 Results and Discussion

3.1 The Distribution of free energy of Single Molecule Binding

We can obtain the distribution functional form for the single molecule binding free energy based on the random energy model[27,28,35,49–53]. 1.00,1.00,1.00rgb]1.00,0.00,0.00And we have verified the model by exploring the density of states(DOS) of protein folding and protein binding in our previous studies[14,15]. Here,1.00,1.00,1.00rgb]1.00,0.00,0.00as mentioned above, the energy function of the ligand-receptor system can be expressed as $H = \sum_{ij} J_{ij}\sigma_{ij}$. Because various atom types and cutoff distances will be involved in the contact interactions, different J_{ij}s with the different values form a distribution 1.00,1.00,1.00rgb]1.00,0.00,0.00The coupling strengths between the atomic pairs J_{ij} is assumed to be Gaussianly distributed. Since the energy of system is linearly associated with the 1.00,1.00,1.00rgb]1.00,0.00,0.00J, this easily leads to a random energy model with the corresponding interaction energy following a Gaussian distribution[6]. Then, the distribution function of the free energy is as follows (For details, See ESI†):

$$f(F) \sim exp[-\frac{(F-\bar{F})^2}{2\Delta E^2}] \tag{7}$$

This is a gaussian distribution for the single molecule binding free energy 1.00,1.00,1.00rgb]1.00,0.00,0.00F near its mean $\bar{F}$ with the variance of the distribution $\Delta F^2 = \Delta E^2$ (The width of the distribution of energy ΔE here means the roughness of fluctuations of the energy landscape above the glassy trapping transition temperature T_c. $T_c = \sqrt{\frac{\Delta E^2}{2S}}$ represents the glass transition temperature implying the onset of the local trapping along the global energy landscape, 1.00,1.00,1.00rgb]1.00,0.00,0.00S, as the configurational entropy, is used to measure the size of the whole configurational space and scales with the size of the receptor-ligand complex system. $S = K_B log\Omega$ where Ω is the number of multifarious states in the configurational space). Additionally, the single molecule binding free energy distribution 1.00,1.00,1.00rgb]1.00,0.00,0.00$f(F)$ has the exponential distribution near the tails (near or below T_c)[27,28,35,49–53] as follows:

$$f(F) \sim exp[\frac{\rho(F-F_c)}{T}]\theta(F_c-F) \tag{8}$$

where $\rho = \frac{T}{T_c}$ and F_c represents the cut-off (upper limit) free energy (θ function is described as follows: $\theta(x) = 1$ while $x > 0; \theta(x) = 0$ while $x <= 0$). Furthermore, the mean free energy is almost equivalent to $|F_c - T_c|$ 1.00,1.00,1.00rgb]1.00,0.00,0.00(here,$K_B = 1$)and the distribution width is close to T_c. Obviously, the resulting distribution of the single molecule binding free energy has simple exponential tails implying slower decays than gaussian distribution. Thus, the events with low probability and binding free energies in the spectrum may play more important role at the exponential tail than the Gaussian center.

In order to quantify the global single molecule binding free energy spectrum of a particular ligand-receptor binding pair, we need not only the simple information near the mean, but also the width of the distribution. Interestingly, the free energy perturbation and some other approximation methods are often applied in the current microscopic atomistic simulations on the bio-molecule binding to estimate affinity. We have illustrated the distribution of free energy beyond the mean and average of binding from the analytical model, the average is often used to study the thermodynamics and kinetics. The variance and the whole distribution are less explored. The average binding energy of the different ligand-receptor pair is self-averaging if the distribution is nearly Gaussian. In other words, the average of binding energy is representative of the typical binding. However, when the underlying distribution is non-Gaussian with long tails, there are significant fluctuations around the mean. Therefore, the average of binding energy is not sufficient in characterizing the typical binding. Statistical distribution is needed. With the development and continuous efforts in computational capability, more complete information can be achieved and more comprehensive analysis and calculations based on microscopic studies can be also realized. The determination of the values of the parameters in terms of the current analytical approach as well as the comparisons and confirmations with the ones from the experiments are vital to the comprehensive description of the underlying free energy landscape and the associated single molecule binding process.

3.2 The Distribution of equilibrium constant K of Single Molecule Binding

The equilibrium constant K or dissociation constant often needs be measured in the experiments such as chemical reactions.1.00,1.00,1.00rgb]1.00,0.00,0.00For the single molecule binding complex, it can be considered to correlate with the difference between F_n and F_{un}:$LogK = \frac{F_n - F_{un}}{T}$ where F_n is the free energy of the native state and F_{un} is the free energy of the non-native states. Actually, the $LogK$ indicates the ability of two molecules associating with each other. From the free energy distribution function, the 1.00,1.00,1.00rgb]1.00,0.00,0.00$f(K)$ for the distribution of the equilibrium constants K will be written as (assuming native binding has little spread):

$$f(K) \sim \frac{1}{K} exp[-\frac{T^2}{2\Delta E^2}(LogK - Log\bar{K})^2] \tag{9}$$

which shows a log-normal distribution about the mean (above T_c) while

$$f(K) \sim K^{-1-\frac{T}{T_c}} \tag{10}$$

which presents the distribution with a power-law decay near the tail of the distribution with the low K value (near or below T_c).

The slow power law decay has a long tail in the distribu-

tion of single molecule binding equilibrium constant for the ligand-receptor pair. In this case, the average is not suitable to 1.00,1.00,1.00rgb]1.00,0.00,0.00characterize the typical equilibrium constant.. This implies that the low probability events at the long tail of the distribution may play an important role in characterizing the whole system. The power law distribution of single molecule binding equilibrium constant therefore implies that most of the bindings occurred in the system show weak binding affinities, just occasionally, a specific binding pattern or mode will be subject to high stability and affinity. Thus, achieving high affinity or the associativity will be very vital although it often is rare to explore the evolution and function for molecular recognition. The results from the theoretical and experimental studies can support each other to further validate the theoretical model and experimental methods. For a few examples, a physical model was applied to reproduce the degree distribution of the experimentally determined protein-protein interactions (PPI) networks[68]; the universal statistical distributions for biomolecular recognition with different sequence pairs were recently uncovered[54].

3.3 The Distribution of Time Scale of Single Molecule Binding

Under the quasi-equilibrium condition, the corresponding time scale τ of single molecule binding for a particular receptor-ligand pair can be expressed as: $log(\frac{\tau}{\tau_0}) = \frac{F^\ddagger - F_{un}}{T}$ where $F^\ddagger$ represents the free energy of the transition state ensemble of system, F_{un} indicates the free energy of the complete non-native states, τ_0 is the time scale. The distribution of single molecule binding kinetic time τ for the ligand-receptor pair can be shown to have the form as (assuming a common transition state free energy):

$$f(\tau) \sim \frac{1}{\tau} exp[-\frac{T^2}{2\Delta E^2}(log\frac{\tau}{\tau_0} - log\frac{\bar{\tau}}{\tau_0})^2] \quad (11)$$

giving a log-normal distribution around the mean of the distribution (above T_c). while

$$f(\tau) \sim \tau^{-1-\frac{T}{T_c}} \quad (12)$$

giving a distribution with much slower power-law decay near the tail (near or below T_c).

Again, we see power law statistics which implies intermittent kinetics of association of binding. 1.00,1.00,1.00rgb]1.00,0.00,0.00Some good examples are from Xie's group[45,46]. The authors had performed a series of single molecule protein conformational dynamics experiments by exploring the photo-induced electron transfer in the flavin reductase system. By the exploration of the fluorescence lifetime of the single molecule system on a photon-by-photon basis, the authors find that the distance of flavin-tyrosine varies over time. And the results further reveal that the conformations fluctuate at multiple time scales covering from hundreds of microseconds to seconds. These demonstrate the presence of various conformations of the single molecule system as well as the different time scales of the interconversion among these conformations at room temperature. Using the similar methods, the authors further experimentally determine the memory kernel $K(t)$ which is a reflection of the time scale fluctuations. In addition,1.00,1.00,1.00rgb]1.00,0.00,0.00the extreme kinetics are rare, but can give important contributions to the whole single molecule binding event and function. Furthermore, the kinetics in the tail can give us hints on the structure of underlying energy landscape through density of states[8,27,28,35,51–53]. This is particularly important in practice as the residence time (off kinetics) characterizes the duration of drugs staying with the target. Both the average and distribution of off kinetics (residence time) is crucial for the effectiveness of drugs.

3.4 Microscopic simulations

We have performed the flexible docking studies on different ligands associating with the receptor COX-2. By exploring specific small molecules binding to the receptor, we can quantify and characterize the different ligand-receptor free energy landscapes. We characterize the mapped landscapes and gain the important statistical information on these single molecule binding systems constituted by numerous specific receptor-ligand pairs. The statistical distributions of the relevant physical properties characterizing the system are obtained. In principle, according to the aforementioned funneled energy landscape theory, the extreme left value of distributions corresponds to the global minimum, which is generally considered as the folded or bound native state. However, in practice, it is possible the lowest free energy state may not correspond to the most geometrically matched binding structure between the ligand and receptor. In those cases, one may need to cluster the low free energy state ensemble around the native and lowest free energy states and create a small native binding ensemble (distribution) against the non-native binding ensemble.

The distributions of the single molecule binding free energy for the different ligands binding with the same receptor in combinations of high, medium and low affinities with high, medium and low intrinsic specificities are shown in Figure 2 (See the fitting details and the parameters of the fitting in ESI†) and Figure S1-S8(ESI†). They are all Gaussian distributed at the center and exponentially distributed near the tail. This is quite consistent with the results from analytical studies mentioned above. We can see that the width of the distribution however is different for each case. For high specific binding (high ISR), the width of the distribution is small relative to the gap, while for low binding specificity (low ISR), the width is more spread and comparable to the energy gap. The width of the distribution is viewed as a measure of the roughness of the binding free energy landscape[6,7]. Rougher energy landscape has a larger width or variance in binding free energy.

In Figure 3 (See the parameters of the fitting in ESI†) and Figure S9-S16(ESI†), we show the single molecule binding equilibrium constant K distribution for the different ligands binding with the same receptor. Consistent with the analytical results, we find that the single molecule binding equilibrium constant or dissociation constant is log normal distributed around the mean and power law distributed near the tail. The distribution coefficients are different for each case with faster decay for high specific bind-

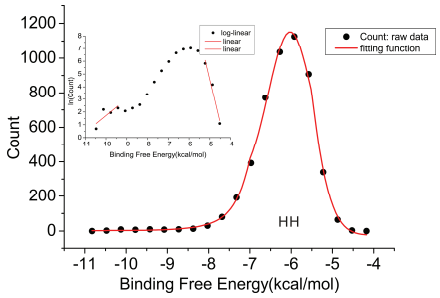

Fig. 2 The distribution of free energy with the compound with high affinity and specificity associating with the Cox-2, the vertical axis indicates the number or probability of states for free energy, the gaussian curve (in the center) and exponential curve (near the tail) are shown. The abbreviation for High is H (High→ H). 1.00,1.00,1.00rgb]1.00,0.00,0.00The insert is the corresponding log-linear plot with the straight line fit near the tail.

ing or smoother landscape and slower decay for low specific binding or rougher landscape of binding.

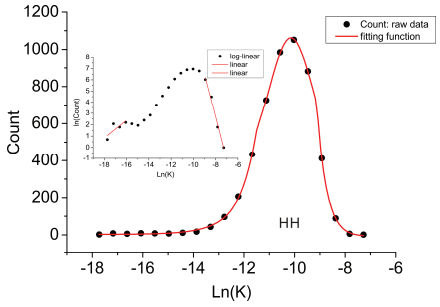

Fig. 3 The distribution of the logarithm of equilibrium constant K with the compound with high affinity and specificity binding with the Cox-2, the vertical axis indicates the number or probability of states for the logarithm of equilibrium constant K, the gaussian curve (in the center) and exponential curve (near the tail) are shown. The abbreviation for High is H (High→ H). 1.00,1.00,1.00rgb]1.00,0.00,0.00The insert is the corresponding log-linear plot with the straight line fit near the tail.

As mentioned, although the statistical nature of the energy landscape of single molecule binding discussed here can be explored directly by single molecule experiments. The interpretation for the origin of kinetics remains challengeable. In this study, we provide a probabilistic description of the kinetics of conformational dynamics through a diffusion approach along the free energy landscape from the result of the microscopic simulations of a specific receptor-ligand pair. In Figure 4 (See the parameters of the fitting in ESI†) and FigureS17-S24(ESI†), for the different receptor-ligand pairs constituted by different ligands binding with a receptor, The results show that the single molecule binding kinetics is log-normal distributed around the mean of the distributions and power law distributed near the tail of the distributions. This is quite consistent with our analytical expectations mentioned above.

It is worthwhile to point out the power law distribution of the single molecule kinetics has been 1.00,1.00,1.00rgb]1.00,0.00,0.00explicitly observed in single molecule experiments. The distribution of energy barriers

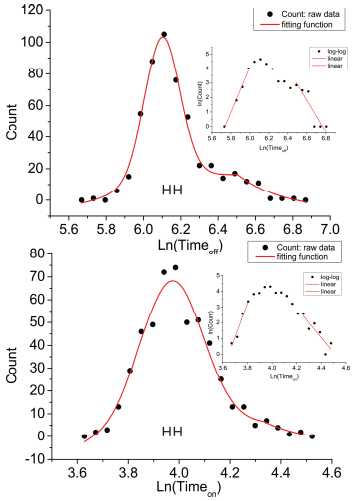

Fig. 4 The distribution of the kinetics with the compound with high affinity and specificity binding with the Cox-2, the vertical axis indicates the number or probability of states for the kinetics, the log-normal (in the center) and power law curve (near the tail) are shown. The abbreviation for High is H (High→ H). 1.00,1.00,1.00rgb]1.00,0.00,0.00The inserts are the corresponding log-log plots with the straight line fit near the tail.

and the observed fluctuations or relaxations of single molecule such as the protein have been extensively applied to interpret the kinetic behaviors at low temperatures and characterize some features observed at high temperatures[30,42–47,69–71]. 1.00,1.00,1.00rgb]1.00,0.00,0.00Our theoretical results of kinetics such as the MFPT and the FPT statistical behavior may provide a possible basis and new insight into the statistics and mechanisms of single molecular recognition dynamics. In addition, except for these experiments on single molecule protein conformational dynamics performed in room temperature, the theoretical predictions for the temperature dependence on the kinetics statistics can be also validated from the serial single molecule experiments by means of controlling temperatures 1.00,1.00,1.00rgb]1.00,0.00,0.00[48,51,72]. These predicted fluctuations or relaxations of the protein can not only be used to interpret the kinetic behavior at low temperatures but also give us a valuable insight into some features observed at high temperatures. This will contribute to our detailed understanding of microscopic mechanism which can not be obtained from the ensemble-averaged experiments. Our models and simulations on single molecule binding, by completely exploring the underlying conformational energy landscape and explicitly considering the kinetic fluctuations, can figure out fairly convincing explanations for the qualitative trends of the single molecule kinetic data observed. Furthermore, we can use the experimental statistical data on kinetics to portray the structure of the underlying energy landscape such as density of states and the topography.

Furthermore, the equilibrium constant or the dissociation constant is usually measured as the average in the bulk. In single molecules, the equilibrium constant depends on each individual measurements. Since the equilibrium constant depends on the underlying free energy landscape, the inherent distribution of the free energy leads to the distribution of equilibrium constant or

dissociation constant. This can be analyzed directly from the trajectories or binding states. We can collect and combine the statistics of the forward and backward rate process to obtain the statistical information about the equilibrium constant. The statistical fluctuations of the equilibrium constant in single molecules will not be averaged out among the large number of the molecules in bulk. Thus single molecules provides a direct route to approach for the statistical fluctuations of the equilibrium properties or the equilibrium constant which can not be probed by the bulk measurements.

4 Conclusions

We have investigated the statistical features of the free energy, equilibrium constant and kinetics of single molecule binding by analytical model as well as microscopic simulations. We gained the definite analytical form of the distribution of single molecule binding free energy being gaussian distributed around the mean and exponentially distributed at the tail. The single molecule binding equilibrium constant is log normal distributed near the mean and power law distributed at the tail. We uncovered the statistical feature of the single molecule binding kinetics to be log-normal distributed around the mean and power law distributed at the tail.

The theoretical predictions can be tested in the single molecule experiments. The experiments can give information on the underlying construction of the energy landscape of single molecule binding. Furthermore, since each ligand or receptor can be different, although the distribution has universal form, the actual coefficient or parameter for the distribution is different for different ligand receptor pair. The specific parameters give quantitative characterizations for specific ligand-receptor binding systems.

The statistical methodology and approach herein rooted on energy landscape theory is quite general. They can be generally applied not only to protein-ligand binding, but also to protein-protein, protein-RNA and protein-DNA binding in the future. In this study, to the first order approximation, we have ignored correlations between different binding states. The correlations may to a certain degree influence the properties of tail of the distribution of the relevant physical variables. We will incorporate this potential effect[55,56] in the future studies.

5 Acknowledgements

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). XLZ would like to thank Professor Erkang Wang for the enlightening discussion. We are grateful to Computing Center of Jilin Province for essential support. This work is also funded by Science and Technology Development Plan of Jilin Province (20130101179Jc) and Jilin province public computing platform.

References

1 D. Koshland Jr, Proc. Natl. Acad. Sci. U. S. A., 1958, **44**, 98.
2 J. A. McCammon, Curr. Opin. Struct. Biol., 1998, **8**, 245–249.
3 N. C. Cohen, N. C. Cohen, Academic Press, 1996.
4 A. Wlodawer and J. W. Erickson, Annu. Rev. Biochem., 1993, **62**, 543–585.
5 T. Clackson and J. A. Wells, Science, 1995, **267**, 383–386.
6 J. Wang and G. M. Verkhivker, Phys. Rev. Lett., 2003, **90**, 188101.
7 J. Wang, X. Zheng, Y. Yang, D. Drueckhammer, W. Yang, G. Verkhivker and E. Wang, Phys. Rev. Lett., 2007, **99**, 198101.
8 J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem., 1989, **93**, 6902–6915.
9 J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, Proc. Natl. Acad. Sci. U.S.A., 1995, **92**, 3626–3630.
10 J. N. Onuchic and P. G. Wolynes, Science, 1995, **267**, 1619–1620.
11 M. Gruebele and P. Wolynes, Nat. struct. biol., 1998, **5**, 662–665.
12 Y. Levy, P. G. Wolynes and J. N. Onuchic, Proc. Natl. Acad. Sci. U. S. A., 2004, **101**, 511–516.
13 J. Wang, L. Xu and E. Wang, Biophys. J.Lett., 2007, **92**, L109–L111.
14 J. Wang, R. J. Oliveira, X. Chu, P. C. Whitford, J. Chahine, W. Han, E. Wang, J. N. Onuchic and V. B. Leite, Proc. Natl. Acad. Sci. U. S. A., 2012, **109**, 15763–15768.
15 X. Chu, L. Gan, E. Wang and J. Wang, Proc. Natl. Acad. Sci. U. S. A., 2013, **110**, E2342–E2351.
16 Z. Yan, X. Zheng, E. Wang and J. Wang, Chem. Sci., 2013, **4**, 2387–2395.
17 K. A. Dill and H. S. Chan, Nat. Struct. Biol., 1997, **4**, 10–19.
18 C.-J. Tsai, S. Kumar, B. Ma and R. Nussinov, Protein Sci., 1999, **8**, 1181–1190.
19 W. KAUZMANN, Adv. Protein Chem., 1959, **14**, 1–63.
20 C. N. Pace, B. A. Shirley, M. McNutt and K. Gajiwala, FASEB J., 1996, **10**, 75–83.
21 C.-J. Tsai, S. L. Lin, H. J. Wolfson and R. Nussinov, Protein sci.: a publication of the Protein Society, 1997, **6**, 53.
22 H. J. Dyson, P. E. Wright and H. A. Scheraga, Proc. Natl. Acad. Sci. U. S. A., 2006, **103**, 13057–13061.
23 R. Perozzo, G. Folkers and L. Scapozza, J Recept Signal Transduct Res., 2004, **24**, 1–52.
24 S. Peuker, A. Cukkemane, M. Held, F. Noé, U. B. Kaupp and R. Seifert, Biophys. J., 2013, **104**, 63–74.
25 R. Sridharan, J. Zuber, S. M. Connelly, E. Mathew and M. E. Dumont, BBA-BIOMEMBRANES, 2014, **1838**, 15–33.
26 X. Chu and J. Wang, PLoS Comput. Biol., 2014, **10**, e1003782.
27 J. Wang, W. Huang, H. Lu and E. Wang, Biophys. J., 2004, **87**, 2187–2194.
28 J. Wang, Chem. Phys. Lett., 2006, **418**, 544–548.
29 W. E. Moerner, Acc. Chem. Res., 1996, **29**, 563–571.
30 H. P. Lu, L. Xun and X. S. Xie, Science, 1998, **282**, 1877–1882.
31 A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz and S. Weiss, Proc. Natl. Acad. Sci. U. S. A., 2000, **97**, 5179–5184.
32 Wang and Wolynes, Phys. Rev. Lett., 1995, **74**, 4317–4320.

33 J. Wang and P. Wolynes, J. Chem. Phys., 1999, **110**, 4812–4819.

34 J. N. Onuchic, J. Wang and P. G. Wolynes, Chem. Phys., 1999, **247**, 175 – 184.

35 J. Wang, J. Chem. Phys., 2003, **118**, 952–958.

36 G. K. Schenter, H. P. Lu and X. S. Xie, J. Phys. Chem. A, 1999, **103**, 10477–10488.

37 X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag and S. Chu, Science, 2000, **288**, 2048–2051.

38 X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter and S. Chu, Science, 2002, **296**, 1473–1476.

39 Y. Jia, D. S. Talaga, W. L. Lau, H. S. Lu, W. F. DeGrado and R. M. Hochstrasser, Chem. Phys., 1999, **247**, 69 – 83.

40 B. Schuler, E. A. Lipman and W. A. Eaton, Nature, 2002, **419**, 743–747.

41 E. A. Lipman, B. Schuler, O. Bakajin and W. A. Eaton, Science, 2003, **301**, 1233–1235.

42 X. Sunney Xie, J. Chem. Phys., 2002, **117**, 11024–11032.

43 H. Yang and X. S. Xie, J. Chem. Phys., 2002, **117**, 10965–10979.

44 H. Yang and X. Xie, Chem. Phys., 2002, **284**, 423–437.

45 H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun and X. S. Xie, Science, 2003, **302**, 262–266.

46 W. Min, G. Luo, B. J. Cherayil, S. C. Kou and X. S. Xie, Phys. Rev. Lett., 2005, **94**, 198302.

47 W. Min, B. P. English, G. Luo, B. J. Cherayil, S. C. Kou and X. S. Xie, Acc. Chem. Res., 2005, **38**, 923–931.

48 Y. Zhou, C. Zhang, G. Stell and J. Wang, J Am. Chem. Soc., 2003, **125**, 6300–6305.

49 E. Gardner and B. Derrida, J. Phys. A: Math. Gen., 1989, **22**, 1975.

50 M. Mézard, G. Parisi and M. Virasoro, J. Phys., Lett., 1985, **46**, 217–222.

51 C.-L. Lee, G. Stell and J. Wang, J. Chem. Phys., 2003, **118**, 959–968.

52 C.-L. Lee, C.-T. Lin, G. Stell and J. Wang, Phys. Rev. E, 2003, **67**, 41905–41909.

53 J. Wang, L. Xu, K. Xue and E. Wang, Chem. Phys. Lett., 2008, **463**, 405–409.

54 X. Zheng and J. Wang, PLoS Comput. Biol., 2015, **11**, e1004212.

55 S. S. Plotkin, J. Wang and P. G. Wolynes, Phys. Rev. E, 1996, **53**, 6271.

56 S. S. Plotkin, J. Wang and P. G. Wolynes, J. Chem. Phys., 1997, **106**, 2932–2948.

57 J. G. Saven and P. G. Wolynes, J. Phys. Chem. B, 1997, **101**, 8375–8389.

58 N. D. Socci and J. N. Onuchic, J. Chem. Phys., 1994, **101**, 1519–1528.

59 N. Socci, J. N. Onuchic and P. G. Wolynes, J. Chem. Phys., 1996, **104**, 5860–5868.

60 E. M. Boczko and C. Brooks, Science, 1995, **269**, 393–396.

61 Z. Guo, C. L. Brooks and E. M. Boczko, Proc. Natl. Acad. Sci. U. S. A., 1997, **94**, 10161–10166.

62 R. A. Goldstein, Z. A. Luthey-Schulten and P. G. Wolynes, Proc. Natl. Acad. Sci. U. S. A., 1992, **89**, 4918–4922.

63 E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U. S. A., 1993, **90**, 7195–7199.

64 D. Klimov and D. Thirumalai, Phys. Rev. Lett., 1996, **76**, 4070.

65 R. G. Kurumbail, A. M. Stevens, J. K. Gierse, J. J. McDonald, R. A. Stegeman, J. Y. Pak, D. Gildehaus, J. M. Miyashiro, T. D. Penning and Seibert, Nature, 1996, **384**, 644–648.

66 D. L. DeWitt, Mol. Pharmacol., 1999, **55**, 625–631.

67 D. S. Goodsell, G. M. Morris and A. J. Olson, J. Mol. Recognit., 1996, **9**, 1–5.

68 E. J. Deeds, O. Ashenberg and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U. S. A., 2006, **103**, 311–316.

69 H. Frauenfelder, S. G. Sligar and P. G. Wolynes, Science, 1991, **254**, 1598–1603.

70 N. Agmon and J. J. Hopfield, J. Chem. Phys., 1983, **79**, 2042–2053.

71 A. R. Panchenko, J. Wang, G. U. Nienhaus and P. G. Wolynes, J. Phys. Chem., 1995, **99**, 9278–9282.

72 J. Wang, L. Xu, K. Xue and E. Wang, Chem. Phys. Lett., 2008, **463**, 405–409.