TOPICAL REVIEW — 8th IUPAP International Conference on Biological Physics

Uncovering the underlying physical mechanisms of biological systems via quantification of landscape and flux*

Li Xu(徐丽)[1,†], Xiakun Chu(楚夏昆)[1], Zhiqiang Yan(晏致强)[1], Xiliang Zheng(郑喜亮)[1], Kun Zhang(张坤)[1], Feng Zhang(张锋)[1], Han Yan(闫晗)[1], Wei Wu(吴畏)[1], and Jin Wang(汪劲)[1,2,‡]

[1] *State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun* 130022, *China*
[2] *Department of Chemistry and Physics, Stony Brook University, Stony Brook, NY* 11794-3400, *USA*

(Received 8 May 2015; revised manuscript received 8 October 2015; published online 30 November 2015)

In this review, we explore the physical mechanisms of biological processes such as protein folding and recognition, ligand binding, and systems biology, including cell cycle, stem cell, cancer, evolution, ecology, and neural networks. Our approach is based on the landscape and flux theory for nonequilibrium dynamical systems. This theory provides a unifying principle and foundation for investigating the underlying mechanisms and physical quantification of biological systems.

Keywords: landscape, flux

PACS: 64.60.aq, 87.18.Vf, 87.19.ll, 87.23.–n **DOI:** 10.1088/1674-1056/25/1/016401

1. Introduction

Biological systems are typically complex nonequilibrium systems across many scales, ranging from the function and recognition of proteins,[1–7] to the cellular networks of macromolecular complexes and regulatory pathways,[8–21] and even the populations of organisms.[22–24] A biological system usually consists of a complex network of components working together to perform specific functions shaped by evolution. The function of a biological system resides not only in the individual components, but also, more importantly, in their interactions and collective behaviors. The challenge is how to explore the complex dynamics of nonequilibrium biological systems and uncover the underlying physical mechanisms. Various experimental, theoretical, and computational approaches have been developed to meet the challenge.[9–21,23–39]

Physical and mathematical modeling of biological systems can give a quantitative description of the dynamical behavior of biological processes.[3–5,8–11] It complements intuitive observations and limited resolutions of experimental investigations. Ordinary differential equations can be used to model the deterministic dynamics of biological systems. For a system under the influence of stochastic fluctuations, which is typically the case for biological systems, Langevin and Fokker–Planck equations are usually able to reasonably describe the system's stochastic dynamics.[8–11,40–43] With appropriate tools of statistical mechanics, the energy landscape theory offers a quantitative global approach to investigate the structure, function, and behavior of biological systems. Quantitative results and predictions from the energy landscape theory have been confirmed by both laboratory experiments and detailed simulations.[1–8] The landscape and flux theory, as a further development and an important extension of the energy landscape theory, brings out the indispensable role of the curl flux in nonequilibrium systems complementary to the potential landscape, which facilitates the unveiling of fundamental principles and physical mechanisms governing nonequilibrium biological systems with complex behaviors and varied functions across multiple scales.

On the molecular scale, the puzzle of how a denatured protein refolds to its native state (the Levinthal's paradox[44]) has been resolved by the funnel-like energy landscape.[1–7] The concept of funneled energy landscapes goes beyond the folding of protein monomers. Binding processes of biomolecules also have funneled landscapes whose topographies determine the binding mechanisms.[2,45–47] Protein folding can be regarded as a diffusion process on the energy landscape which eventually reaches the native state. As a result, the folding thermodynamics and kinetics can be inferred from the knowledge of the energy landscape. The energy landscape theory has been been further extended and applied to protein binding and conformational dynamics.[26,48–50] Significant progresses have been made on interpreting the results from the energy landscape perspectives.[51–53] Hence, the energy landscape theory gives a quantitative description with physical insights into the dynamics at the molecular level and has enhanced our knowledge of the corresponding biological processes.

In order to have a concrete understanding of the en-

*Project supported by the Natural Science Foundation of China (Grant Nos. 21190040, 11174105, 91225114, 91430217, and 11305176) and Jilin Province Youth Foundation, China (Grant No. 20150520082JH).

†All authors have equal contributions.

‡Corresponding author. E-mail: jin.wang.1@stonybrook.edu

© 2016 Chinese Physical Society and IOP Publishing Ltd

ergy landscape concept, we have used molecular dynamics simulations and equilibrium statistical mechanics to quantify the energy landscape of protein folding and binding dynamics.[29,54,55] The connections between theories and experiments have been established with quantified energy landscapes. We have explored a wide variety of models based on the energy landscape theory to investigate various biomolecular dynamics, including multi-domain protein folding,[56] intrinsically disordered proteins' (IDPs') binding–folding,[57–62] and protein conformational changes.[63–68] The simulation results are consistent with the experiments in many ways, identifying the critical role of the energy landscape in biological processes.

Further, the energy landscape theory allows us to use bioinformatics input to learn the rules of energetic interactions that govern protein folding and ligand binding. The energy landscape theory has been used to quantify not only the affinity but also the specificity for molecular recognition critical for drug discovery.[26,28,69–71] Protein structures and binding complexes have been predicted successfully using optimized energy functions based on the energy landscape theory.[26,28,30–38,69–71]

At the scales encompassed by systems biology, research on the dynamics, function, and stability of nonequilibrium biological networks has progressed.[9–19,23,24,39–41,72–78] In particular, the landscape and flux theory, which we have developed and applied to explore a variety of biological systems, is our contribution to the advancement of this field.[9–19,23,24,39] The general theoretical framework of landscape and flux consists of several components inherently interlinked by the central idea of the driving force decomposition for nonequilibrium dynamical systems. We have discovered that, in contrast to the equilibrium dynamics of biological systems determined only by the potential landscape, there are two essential parts that constitute the driving force of nonequilibrium biological systems: the gradient potential force and the curl flux force. More specifically, the potential here means the nonequilibrium potential landscape quantifiable via the steady-state probability distribution, and the flux refers to the steady-state probability flux, a signature of detailed balance breaking that indicates the system's nonequilibrium feature in the steady state. Both parts are required to give a complete characterization of the nonequilibrium dynamics of biological systems. Based on the driving force decomposition, other components in the landscape and flux framework, including the nonequilibrium thermodynamics, the fluctuation–dissipation theorem, the fluctuation theorem, the gauge field, and the path integral with kinetic paths and rates, become integral parts of a coherent theoretical framework.

We have applied the general landscape and flux framework to investigate a wide range of more specific nonequilibrium systems with biological significance.[9–12,14,16,18,19,22–24,39] In the investigation of the cell cycle process, cell cycle phases and cell cycle checkpoints have been identified, respectively, as the local attractor basins on the landscape and the transition states between them, along the path of cell cycle oscillation.[9–11] The Waddington landscape and path for stem cell development, differentiation, and reprogramming have been quantified in the landscape and flux framework.[12,14,16,18,19] Cancer has been proposed to be a disease of cellular states associated with the underlying gene regulatory network, where the multiple cellular states are identified as the attractor basins on the landscape; we have studied the transitions between the normal state and the cancer state quantitatively.[39] From the perspective of the landscape and flux theory, evolution dynamics that goes beyond Wright and Fisher has been formulated and the red queen phenomennon has been explained.[23] Ecosystems have also been investigated in the landscape and flux framework, with the global stability of ecological systems quantified.[22] Neural networks with asymmetrical connections have been studied for learning and memory.[24] In addition, the non-adiabatic dynamics of gene expression,[17] self regulating genes,[17,79] chaotic strange attractor,[80] and spacial development with pattern formation[81–84] have also been explored.

The rest of this article is structured as follows. In Section 2, we discuss the folding, binding, and conformational dynamics of biomolecules in the framework of the energy landscape. In Section 3, we address the intrinsic specificity and kinetic specificity for ligand binding as well as their applications in drug discovery and design in the context of the energy landscape theory. In Section 4, we review the general theoretical framework of landscape and flux and explore its more specific applications in cell cycle, stem cell, cancer, evolution, ecology, and neural networks. The conclusion is given in Section 5.

2. Folding, binding, and conformational dynamics of biomolecules

2.1. Intrinsic energy landscape quantification of protein folding and binding

Although the concept of an energy landscape is widespread, researchers in the field of molecular biology often use the energy landscape theory in a qualitative way. This is mainly due to the fact that the energy landscape, reflecting the delicate relationship of energy, entropy, and structural information, cannot be explicitly measured by experiments. On the other hand, experimental measurements, such as the thermodynamic temperature and kinetic rate, cannot be well predicted by theories. Consequently, the relationship between the topography of the energy landscape and the thermodynamics

and kinetics of biomolecular dynamics is ambiguous. To meet the challenge, molecular simulation is proposed to serve as the meeting point, aiming to bridge the gap between theories and experiments.

Conventionally, molecular simulations are often performed under constant temperature, corresponding to the canonical ensemble. However, the energy landscape, which reflects the underlying intrinsic interactions of the system, is not supposed to be dependent on temperature explicitly. Therefore, one has to transform the temperature-dependent distribution of energies from the canonical ensemble ($n(E,T)$) into the micro-canonical ensemble ($n(E)$). This can be done using the Boltzmann relation $n(E) \sim n(E,T)\,\mathrm{e}^{E/k_{\mathrm{B}}T}$, as encoded in the weighted histogram analysis method (WHAM).[85] The distribution of the micro-canonical ensemble, referring to the density of states, corresponds to the intrinsic energy landscape.

In theories, there are three quantities describing the topography of the energy landscape:[2,4,5,45] energy gap δE between the native state and the average of non-native states, measuring the slope of the funnel;[86,87] energy roughness ΔE, quantifying the bumpiness of the funnel; entropy S, describing the size of the funnel. These three quantities result in a dimensionless ratio Λ, expressed as the energy gap versus energy roughness and entropy $\delta E/(\Delta E\sqrt{2S})$. As a characteristic of the energy landscape topography, Λ is a physical quantity measuring the funneled quantity of the energy landscape. All the topography quantities of the energy landscape can be calculated directly from the density of states.

We quantified the intrinsic energy landscapes of protein folding and presented the quantitative funnel diagrams.[54] The relationship of energy, entropy, and structure can be illustrated quantitatively from the energy landscapes (Fig. 1). On the thermodynamic aspect, there are two critical temperatures controlling protein folding: one is the folding temperature T_{f}, at which the folding and unfolding states are equal in probability; the other is the glass transition temperature T_{g}, at which the system becomes frozen with folding stagnation.[88] For a foldable chain, the folding temperature has to be higher than the glass transition temperature. On the kinetic aspect, the kinetic efficiency is often represented as the folding rate. In practice, both the thermodynamic temperature and the kinetic rate can be measured explicitly by experiments.

We found that the energy landscape topography Λ has a strong positive correlation with the thermodynamic stability (versus glass trapping) and kinetic rate. In other words, a more funneled energy landscape leads to more stable folding and a faster rate. Thus we have established the quantitative relationship between an energy landscape and folding thermodynamics and kinetics.

We then quantified the energy landscapes of protein binding.[29] For flexible biomolecular recognition, the central issue that needs to be addressed is the binding mechanism. We picked out 15 different homodimers and classified them into two groups based on different binding mechanisms found by experiments: 2-state "coupled binding–folding" homodimers and 3-state "folding prior to binding" homodimers. By quantifying the individual binding, folding, and global binding–

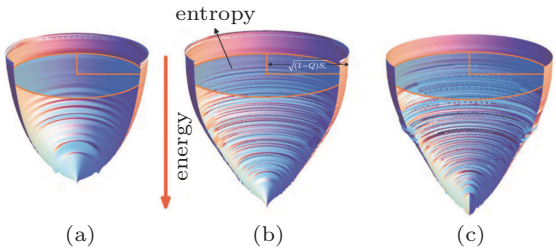


Fig. 1. Quantitative illustrations of protein folding funnels for (a) villin headpiece, (b) CI2, and (c) P13 from simulations. The Z axis is energy. The area of each ellipsoid stratum represents entropy. The semimajor axis of the ellipsoid at each stratum is $\sqrt{(1-Q)S_{\mathrm{r}}}$, where Q is similar to native structure and S_{r} is a constant. The quantitative relationship of energy, entropy, and structure can be inferred from the quantified energy landscapes.[54]

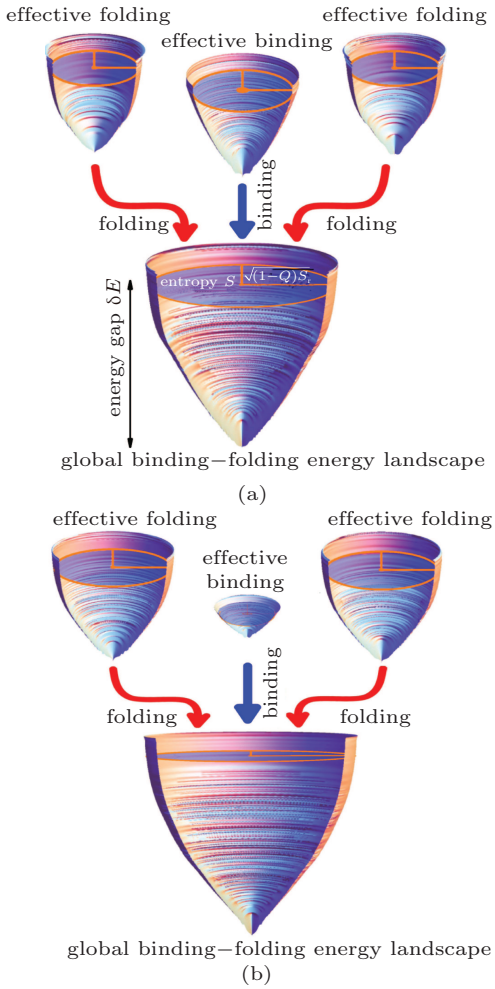


Fig. 2. Energy landscape quantifications during flexible recognition for (a) 2-state and (b) 3-state homodimers. Each homodimer has two individual monomeric folding energy landscapes, one individual interfacial binding energy landscape, and a combined global binding–folding energy landscape. The funnel plot details are the same as Fig. 1.[29]

folding energy landscapes (Fig. 2), we found that the 3-state homodimer has a more significant folding-versus-binding energy landscape funnel than the 2-state homodimer. We proposed a ratio expressed as the energy landscape topography of binding versus folding, $\Lambda^{\rm bind}/\Lambda^{\rm fold}$, to measure the binding mechanism quantitatively. Thus we classified the binding mechanisms successfully through the quantified energy landscape approach.

By calculating the global binding–folding energy landscape topography $\Lambda^{\rm global}$, we found that $\Lambda^{\rm global}$ is strongly correlated with the thermodynamics and kinetics of protein binding, similar to the situation in protein folding. Our findings show that the energy landscape controls the pathway, thermodynamic feasibility, and kinetic efficiency of protein to realize its function.

Using molecular simulations, we demonstrated the crucial role of the energy landscape in biomolecular dynamics. The thermodynamics and kinetics of protein folding and binding can be predicted well with the quantified energy landscape due to the strong correlation between them. On the other hand, with experimental measurements on thermodynamics and kinetics, the energy landscape can be portrayed accurately. These results bridge the gap between theories and experiments, providing a novel way to understand the biomolecular dynamics by combining the results from both theories and experiments.

2.2. Binding–folding free energy landscape quantification of intrinsically disordered proteins

The long-standing paradigm that a biomolecular function is determined by a unique folded structure is challenged by intrinsically disordered proteins (IDPs).[89–91] An IDP lacks a unique tertiary structure in an isolated state, either entirely or in parts, but often folds upon binding to its partners.[92,93] This "binding induced folding" scenario gives IDPs many advantages,[94,95] including fast association/disssociation rate, multiple binding targets, low affinity but high specificity, etc, resulting in IDPs' presence in many critical cellular activities, such as transcription, signaling, and regulation.[96–98] Understanding IDPs' binding–folding will refresh our view on protein folding and contributes to our knowledge of molecular biology.

The structure-based model (SBM) is a simplified molecular simulation model, derived from the energy landscape theory.[99] The SBM only takes into account native interactions, leading to a perfect funneled energy landscape. The validity of SBM is supported by the "principle of minimal frustration" indicating that the native structural topography controls the folding mechanism, and also the increasing reports in which the simulation results are consistent with experiments in many ways.[27,99,100] IDPs' binding–folding processes were investigated extensively by a coarse-grained SBM,[58,101–103] built on the basis of the final binding complex structure located at the bottom of the energy landscape, which reproduced the binding–folding mechanism very well in agreement with experiments. Huang *et al*. applied a plain coarse-grained SBM, which modulates the strengths of the native contacts, to examine the kinetic advantage of IDPs' binding.[104] The thermodynamic free energy landscapes and kinetic rates were analyzed comprehensively from the simulation results. The analysis led to the conclusion that, compared to ordered chains, the flexibility in IDPs endows them with a stronger tendency to form the bound complex from an encounter complex with the partners, thus providing a practical explanation for the "fly-casting" mechanism.[95]

Recently, we developed a two-bead SBM with explicit incorporation of electrostatic interactions to investigate IDP histone chaperone Chz1 binding–folding to histone H2A.Z-H2B[60] (Fig. 3). By analyzing the interactions in the transition state ensemble, we found that the interactions are charge-oriented, indicating that the electrostatic interactions serve as the driving force at the early stage of the recognition, consistent with experiments.[105] We also found that Chz1 collapses to a non-native conformation at low salt concentration due to the strong intra-chain electrostatic interactions. Although the collapsed structures in IDPs are widely found in experiments,[106–108] their effects on IDPs' function remain unclear. We found that it consumes time to unravel the collapsed structure in Chz1 during binding. Therefore, we proposed opposite roles of electrostatic interactions in IDPs' binding–folding: the inter-chain interactions facilitate binding through the fly-casting mechanism, while the intra-chain interactions impede the binding via local collapsing or trapping. We argued that IDP itself is able to optimize the two effects to ensure its functional performance.

Although the coarse-grained SBM has made significant progress on the description of IDPs' binding–folding processes, molecular simulation at the atomic level is in urgent need. However, due to the limit of computational resources, atomistic simulations are often focused on a class of small single-domain proteins on specifically-designed computers.[109,110] To overcome the difficulty, a variety of multi-scale molecular simulations were developed by combining the coarse-grained and atomistic simulations to explore biological processes at a wide range of spatial and temporal scales. We used a multi-scale molecular simulation approach combining coarse-grained SBM and atomistic optimal path calculation method to explore the binding–folding process of IDP inhibitor IA3 to its target enzyme (Fig. 4).[59] With the analysis of the kinetic data from laser temperature-jump fluorescence spectroscopy and fluorescence resonance energy transfer (FRET), experimentalists concluded that binding is prior to folding for IA3.[111] Our coarse-grained model

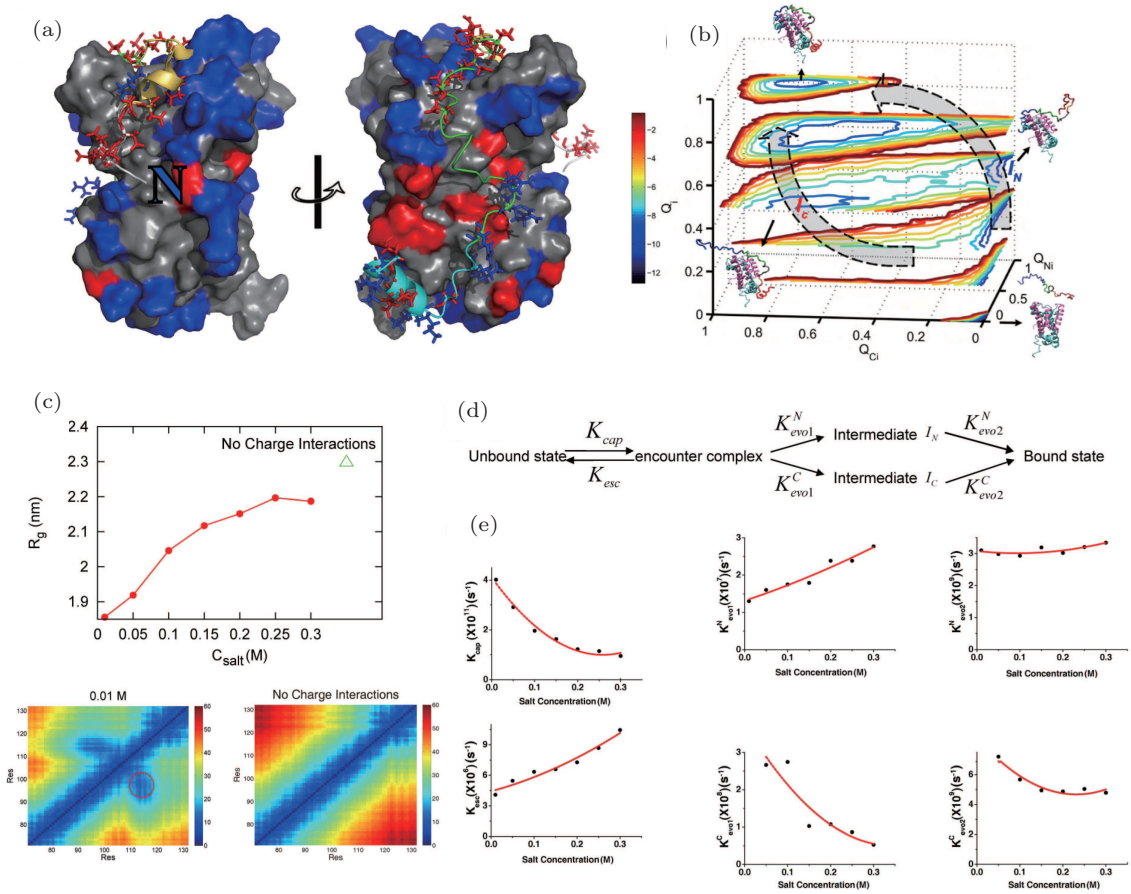


Fig. 3. Free energy landscape quantifications and kinetic analyses of Chz1 binding–folding to histone H2A.Z-H2B. (a) The charged property of the Chz1-H2A.Z-H2B system. Positively and negatively charged residues are colored by blue and red, respectively. (b) Two parallel binding pathways from free energy landscapes. (c) Collapsed structure in Chz1 at low salt concentrations. The radius of gyration R_g and the inter-residue distance describe the degree of collapsed structure at different salt concentrations. (d) Schematic illustration of two parallel binding pathways. (e) Binding rates with different salt concentrations.[60]

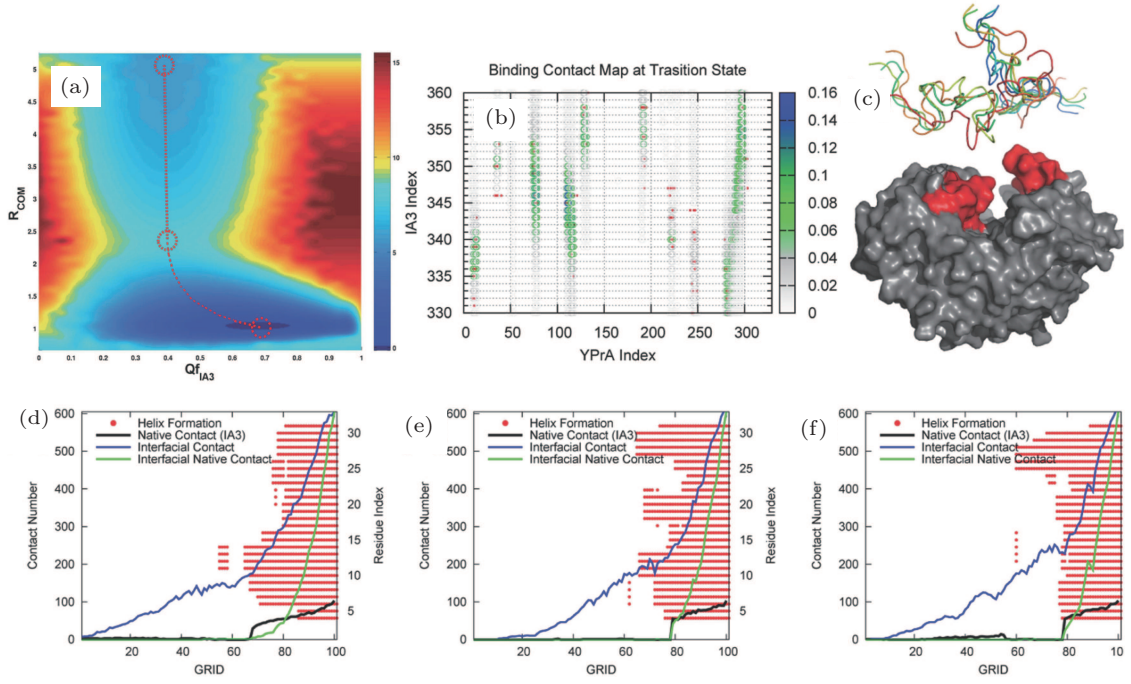


Fig. 4. Free energy landscape, contact map, and kinetic pathway analyses of IA3 binding–folding to its target enzyme. (a) Free energy landscape projected onto the reaction coordinates of binding (R_{com}, distance between the center of mass of IA3 and the enzyme) and folding (Q_{fIA3}, fraction of the native contacts). (b) Contact map at the transition state ensemble. Colored circles represent contacts formed in the transition states and red points indicate native contacts in the final complex. (c) Starting structures in the atomistic path simulation. (d)–(f) Three typical binding–folding pathways for atomistic simulations.[59]

generated thermodynamic free energy landscapes, which confirmed the binding mechanism. Further analysis on the contact map in the transition state underlined the important role of non-native interactions in the binding–folding process. With the quantified optimal pathways at the atomic level, we analyzed in detail the binding prior to folding mechanism by the formation of a helix in IA3 after going through significant non-native interactions. We argued that the transient non-native interactions can shrink the entropy of the energy landscape, leading to an efficient binding–folding process.

With our recently developed hybrid all-atom model and multi-scale simulation approach, we explored the binding–folding process of an IDP in measles virus nucleoprotein (α-MoRE) to its target.[61] α-MoRE, regarded as an IDP, forms a perfect α helix during its molecular recognition.[112–114] Experiments revealed that α-MoRE in solution is not an ideal random coil but in rapid interconversion between the unfolded state and multiple partially helix formed states.[115] We first set up an REMD simulation with a Charmm22* force field[116] to explore the inter-chain dynamics of α-MoRE in solution. Through clustering analysis,[117] we found that α-MoRE dominates at a series of partially helix formed clusters (Fig. 5), consistent with the experimental findings. Additionally, the simulations and experiments are strongly correlated by fitting chemical shift data, confirming the existence of the dynamic conformational equilibrium of α-MoRE in solution. Then we combined the empirical force field and the SBM potential together, constructing a novel hybrid SBM. The empirical force field describes the precise local-in-sequence interactions, aiming to capture the secondary formed structure in α-MoRE, while the SBM potential describes the non-local or long-range interactions, aiming to accelerate the recognition process in simulations. With quantified free energy landscapes, we proposed a "conformational selection followed by induced-fit" binding scenario for α-MoRE. We argued that this binding scenario will facilitate IDPs' binding via a "kinetic divide-and-conquer" mechanism to ensure a fast rate of association and dissociation, in favor of IDPs' biological function.

The quantified multi-dimensional free energy landscapes along various reaction coordinates provide clear pictures on how the binding–folding process in IDPs occurs. Analyses from free energy landscapes, such as interactions and clustering, give comprehensive explanations on the role of non-native interactions, collapsed structure, downhill folding, intrinsic disorder, and molten globule in IDPs' biological function. Based on the energy landscape theory, SBM is sufficient to capture the characteristic of IDPs' binding–folding, indicating that the theory can be applied to proteins with even weakly-funneled or anti-funneled energy landscapes.

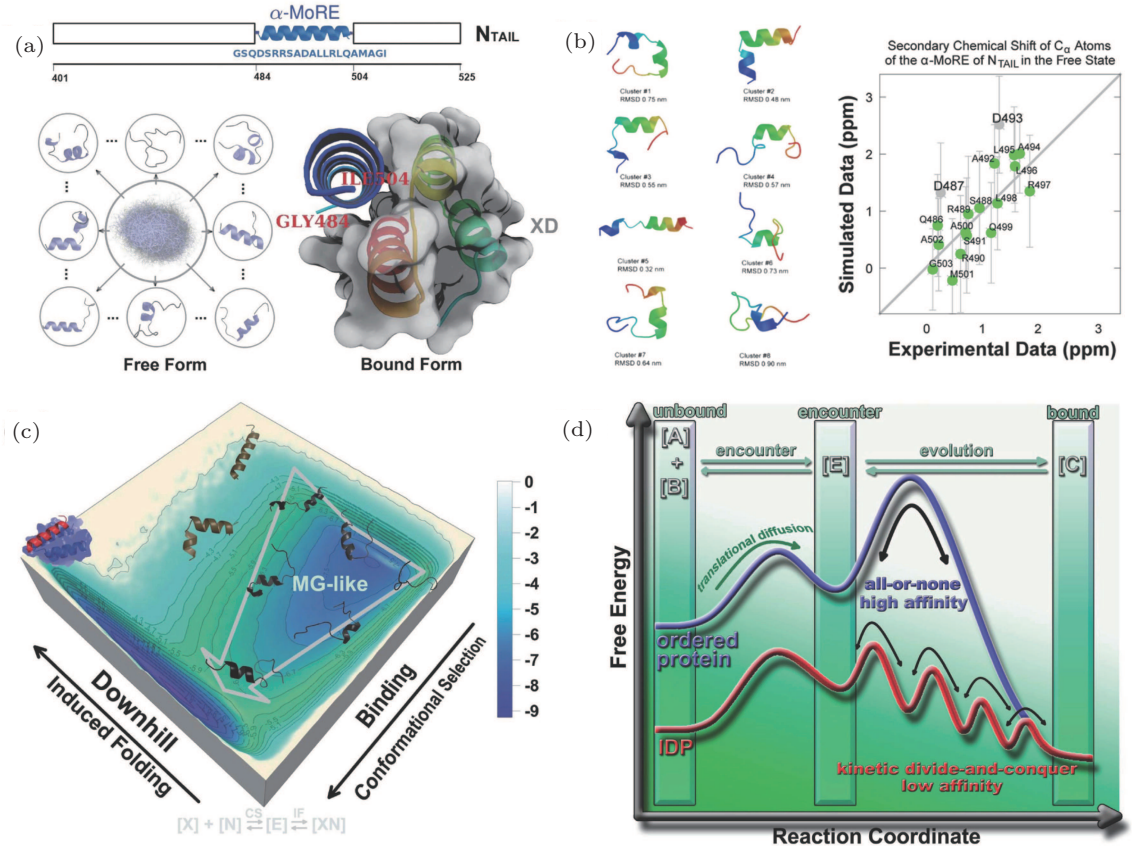


Fig. 5. Free energy landscape quantifications of α-MoRE binding–folding to its partner. (a) Schematic representations of α-MoRE in solution and binding complex with its partner. (b) Cluster analysis of α-MoRE derived from REMD simulations and chemical shift compared with experiments. The top eight clusters are shown to occupy 70.5% of the population of the trajectory at 298 K. (c) Binding–folding free energy landscape from hybrid SBM. (d) Schematic illustration of kinetic divide-and-conquer binding for IDPs.[61]

2.3. Functional free energy landscape quantification of large-scale conformational dynamics

Protein realizes its function by binding to its partners. During recognition, protein often undergoes large-scale conformational changes to achieve its specific biological function.[118] Even in isolated states, protein does not always possess a single well-defined structure. Instead, protein often fluctuates among multiple states to facilitate its biological function, implying there are multiple basins on the energy landscapes.[48,49,119,120] To understand how protein realizes its function, it is essential to understand how functional structural rearrangements occur at the molecular level. Recent developments in experimental techniques, such as single-molecule and nuclear magnetic resonance (NMR) measurements, have improved our knowledge of biomolecular dynamics significantly. However, due to limited spatial and temporal scales, it is still challenging to characterize functional conformational dynamics experimentally. Hence, various molecular simulations have been developed to quantify free energy landscapes.

The basic strategy in simulation is to build a multi-basin SBM potential function. There are three classes of models at different levels: macroscopic model,[121,122] where multiple potentials, each of which has a single energy minimum unambiguously defined by a prior structure, are connected smoothly at the joined point, resulting in a multi-basin energy function; microscopic model,[63–65] where the elementary potential between two atoms in proteins is modeled to have multiple minima, corresponding to multiple structures; mesoscopic model or microscopic mixing contact map,[66,67,123] where individual energy terms (usually native contacts) from multiple structures are mixed separately with different mixing protocols.

We first focused on an excellent allosteric model, adenylate kinase (ADK), which is composed of a core domain (CORE), an ATP-binding domain (LID), and a nucleoside monophosphate binding domain (NMP). During its catalytic function, ADK undergoes large-scale domain arrangement from open to closed state.[124,125] To explore the conformational dynamics of ADK, both microscopic and mesoscopic models were investigated in our group. The microscopic model was developed to explore the intrinsic dynamics of ADK.[63] Using free energy landscapes and contact maps, we addressed the critical dynamical interplay between LID and NMP domain and characterized the hot residues controlling the intrinsic dynamics. Furthermore, the mesoscopic model developed recently with explicit consideration of substrates (AMP and ATP) has made a great contribution to the global understanding of intrinsic and functional dynamics in ADK (Fig. 6).[67]

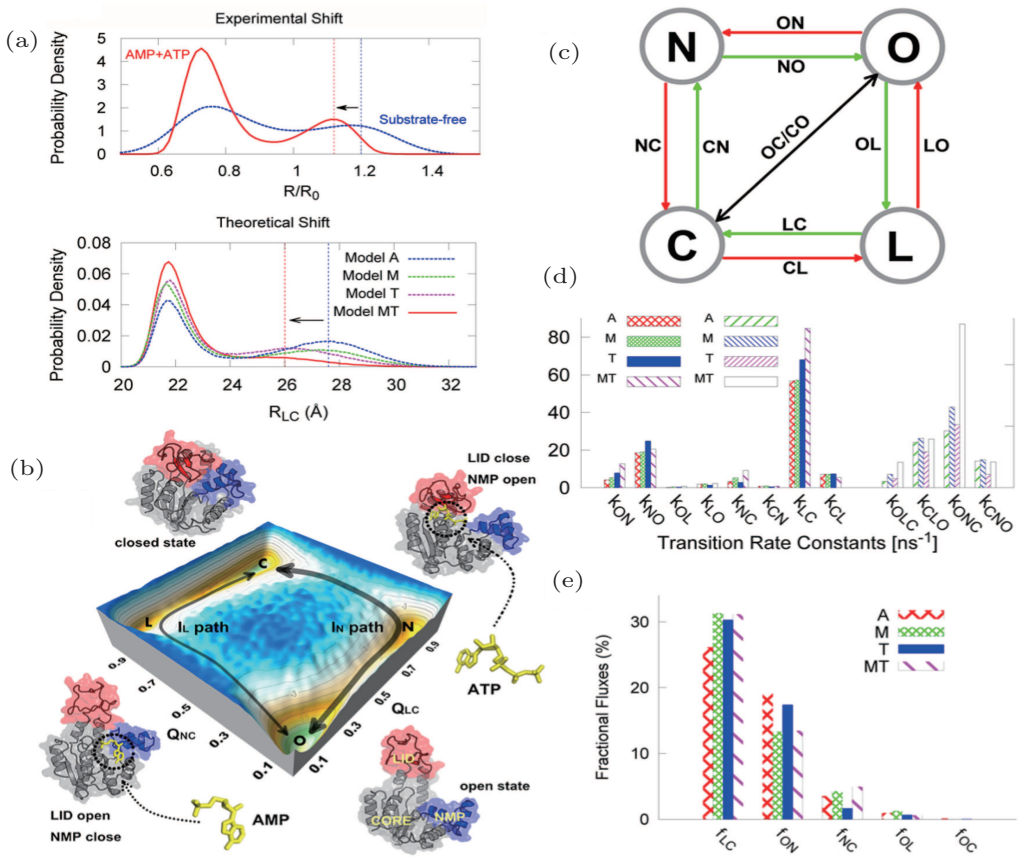


Fig. 6. Functional free energy landscape and kinetic analyses of ADK with substrates binding. (a) Comparison of the simulation results with the experimental measurements on the LID-CORE distance distributions, which is used to distinguish the open and closed states of ADK. (b) Free energy landscapes of ADK induced by substrates binding. (c) Illustrations of the four-state kinetic model. (d), (e) Kinetic analyses of rates and fluxes during the process of ADK realizing its function with substrates binding.[67]

Combining both models, we proposed that only with binding can ADK be induced to a dominate closed state, confirming the pivotal role of substrates on the functional dynamics of ADK. In addition, we found that there are two parallel binding pathways with two intermediates from the quantified energy landscapes. Based on the thermodynamic and kinetic results, we concluded that the motion of the NMP domain is the rate-limiting step for ADK's conformational change. The remarkable degree of fluctuating intrinsic dynamics in ADK is supposed to favor the functional conformational changes during substrates binding.

Then we explored the functional dynamics of DNA Y-family polymerase IV (DPO4) binding to its target DNA.[68] Structural analysis reveals that DPO4 exhibits distinct conformations in the apo and DNA bound state, implying that DPO4 undergoes a large-scale conformational transition during its binding to DNA.[126] Accordingly, we built a two-basin mesoscopic SBM, aiming to capture the details of flexible protein-DNA recognition. From thermodynamic free energy landscapes, we found that the DPO4-DNA recognition process undergoes four different kinetically connected stages: unbinding state (US), encounter complex (EC), intermediate state (IS), and binding state (BS) (Fig. 7). Furthermore, we found that at different stages, DPO4 itself has different conformational distributions among the apo state (A-state), DNA-bound state (B-state), and intermediate state (I-state). Therefore, we drew a dynamic picture of DPO4-DNA binding accompanied with the distributions of the DPO4 conformations varying at different binding stages. With the calculated kinetic rates at each step, we found that the conformational dynamics or flexibility in DPO4 contributes to the efficiency of DNA recognition. Intriguingly, DPO4 is still in the conformational equilibrium between I- and B-state at the final protein-DNA complex, leading to fluctuating interactions between protein and DNA. DPO4 is a typical Y-family polymerase, responsible for low-fidelity of DNA synthesis.[127] The flexible transient interactions between DPO4 and DNA may have implications on DPO4 accommodating the bypass of various DNA lesions. The theoretical investigations on DPO4 binding to DNA offer a unique insight on the conformational dynamics affecting the protein's catalytic function.

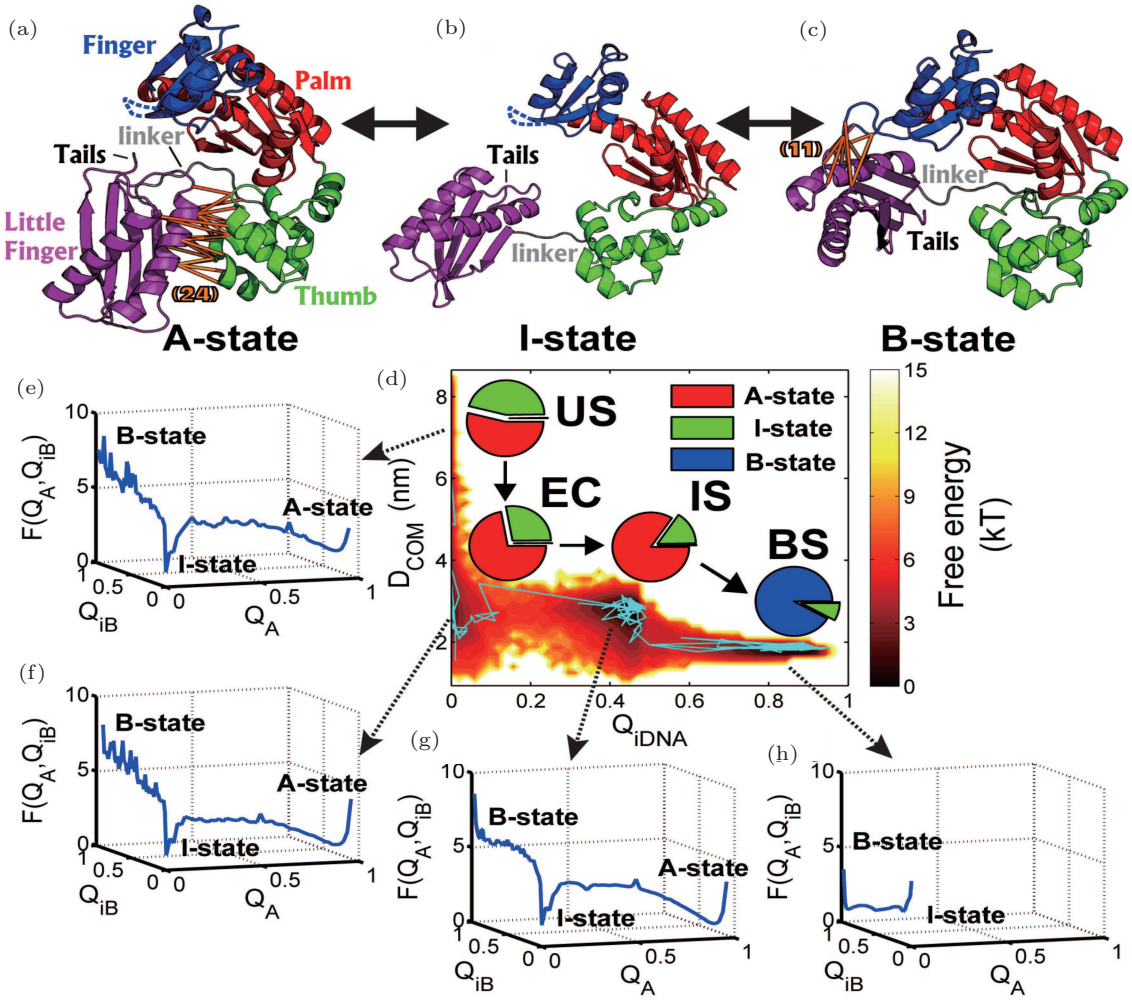


Fig. 7. Functional free energy landscape of DPO4 binding to its target DNA. Structural illustrations of DPO4 in (a) apo-state, (b) intermediate state, and (c) DNA-bound state. (d) Free energy landscape of DPO4 binding to DNA, with pie-charts representing the population distributions of DPO4 in the three states at each binding stage. (e)–(g) Free energy landscapes of DPO4 in US, EC, IS, and BS binding stages.[68]

Built on the energy landscape theory, the multi-basin SBM can explore the functional conformational dynamics in protein during its binding. The multi-basin model aims to capture the intrinsic conformational fluctuation in protein and its role in protein's function. With the quantified free energy landscapes and kinetic rate/flux analysis, we can gain a global picture of the functional conformational dynamics, giving a deeper understanding of the protein dynamics.

3. Ligand binding specificity and drug discovery

3.1. The intrinsic specificity and its applications in drug discovery

Specificity refers to the discrimination of the same ligand binding to its partner against other possible receptors (Fig. 8(a)). In thermodynamics, it is the relative difference in the affinity between the specific binding complex and other competitive complexes.[128–131] In terms of Boltzmann distribution ($P \sim \exp[-F/k_{\rm B}T]$), the populations of different conformations depend on the binding free energy exponentially. As a result, the binding free energy gap between the specific complex and the competitive ones can result in a large population distinction in equilibrium. In practice, it is challenging to quantify this conventional concept of specificity since the receptor universe is too large and the information about them is usually limited.

To circumvent the challenge, an alternative way to determine quantitatively the ligand–receptor binding specificity was proposed.[26,28,36] In a thought experiment, one can assume that the N and C terminus of each protein receptor of the protein universe are connected by linkers. In this sense, multiple individual receptors connected by the linkers are equal to a single large composite receptor. The conventional specificity discriminating a ligand binding to a specific receptor against other receptors (Fig. 8(a)) becomes the discrimination of a ligand binding to a specific site (mode) against other binding sites (modes) of the large composite receptor (Fig. 8(b)). In other words, the large target receptor can be viewed as many small receptors linked together. In this way, the quantitative description of specificity refers to the gap of binding energy between the specific binding mode and other competitive binding modes, rather than the gap between the specific binding partner and other competitive binding complexes. Similarly, a receptor with diverse binding modes by a ligand (Fig. 8(b)) is equivalent to that of the ligand universe binding to the same receptor (Fig. 8(c)), under the assumption that the receptor is large enough. Thus, exploring the binding interactions via diverse sites of the receptor (Fig. 8(b)) and exploring the binding interactions via sequences (diverse receptors in Fig. 8(a) or diverse ligands in Fig. 8(c)) are equivalent. With the assumption of a large target receptor, one can term this criterion for specificity as the intrinsic specificity, in contrast with the conventional concept of specificity.

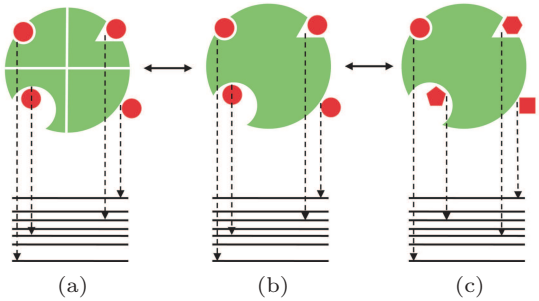


Fig. 8. Schematic view illustrating the specificity. (a) Conventional specificity: the same ligand (red) binding to diverse receptors (green) with the corresponding energy spectrum. (b) Intrinsic specificity: a particular ligand binding to diverse binding states (modes) of its receptor. (c) Conventional specificity: the same receptor with diverse ligands.[36]

The process of receptor–ligand binding can be described and physically quantified by the funnel-like energy landscape, with the native binding mode at the bottom and the binding roughness located in the binding paths.[25–29] From the funnel-like energy landscape, the native conformation has the lowest binding energy (Figs. 9(a) and 9(b)) and the energies of the non-native conformations are determined by the statistical

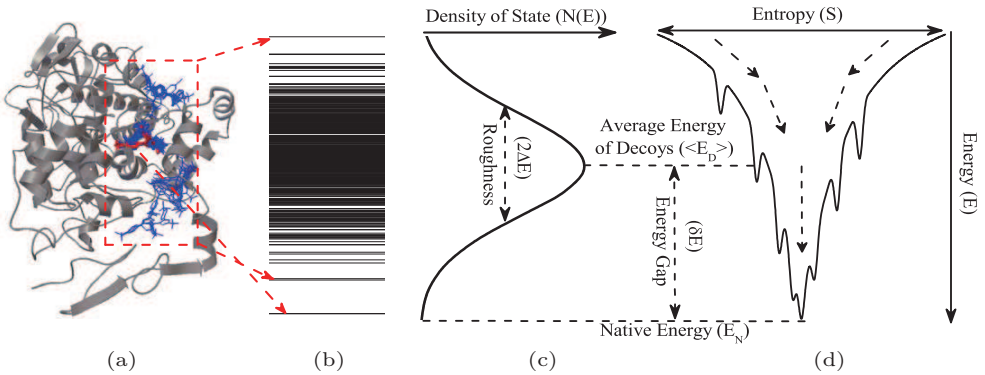


Fig. 9. Funneled energy landscape and energy spectrum of ligand binding. (a) Diverse docking conformations of the target COX-2 with a drug SC-558, the native binding pose (sticks in red); other decoys (lines in blue). (b) The energy spectrum for decoy ensemble. (c) The resultant binding energy distribution, where ΔE is the variance of energy, $\langle E_{\rm D}\rangle$ is the average of energy of decoy ensemble, and δE is the energy gap. (d) Binding energy landscape with a funnel shape pointing to the native state.[36]

Gaussian distribution (Fig. 9(c)). A dimensionless quantity ISR (intrinsic specificity ratio) is proposed to quantify the magnitude of intrinsic specificity (Figs. 9(c) and 9(d)), which is proportional to the ratio of the binding temperature and the trapping temperature (T_b/T_g).[26] The formula to calculate ISR is given by

$$\mathrm{ISR} = \alpha \frac{\delta E}{\Delta E}, \tag{1}$$

where α scales the entropy contribution ($1/\sqrt{2S}$) to the specificity; δE represents the gap between the native conformation E_N and the decoy ensemble $\langle E_D \rangle$; and ΔE represents the energy roughness of the decoys. Mathematically,

$$\delta E = \langle E_D \rangle - E_N, \tag{2}$$

$$\Delta E = \sqrt{\langle E_D^2 \rangle - \langle E_D \rangle^2}, \tag{3}$$

where $\langle \cdot \rangle$ is the average over decoy ensemble. A larger ISR suggests a higher discrimination for the native binding state against the decoys, therefore a higher intrinsic specificity. ISR gives a quantitative description of the intrinsic specificity without assessing the conventional specificity that requires the exploration of the ligand or receptor universe.

It is known that specificity and affinity are both required for efficient and specific biomolecular recognition. However, most quantitative descriptions of biomolecular recognition were focused on improving the computation of affinity while lacking the quantification of specificity. Based on the intrinsic specificity and the corresponding intrinsic specificity ratio (ISR),[28] Liu *et al*. explored and characterized the binding modes of levamlodipine and human serum albumin (HAS), validating a high ISR value for the binding site of levamlodipine.[132] Furthermore, Wang *et al*. developed two-dimensional computational drug screening considering both affinity and intrinsic specificity as the selection criteria.[28] By using the two-dimensional virtual screening with ISR and affinity, Wang *et al*. successfully identified three active hits against Ras protein.[133–137] Yang *et al*. further studied the target Ache using two-dimensional virtual screening for potential inhibitors.[70] Combined ligand screening of the shape and electrostatic similarity as well as local binding site similarity was also applied to search for known drugs against cognitive deficits. For the first time, we found that pazopanib, known as a tyrosine kinase inhibitor for cancer treatment, can inhibit acetylcholinesterase (AchE) activities at the sub-micromolar concentration.

To improve the accuracy of binding affinity prediction, Yan *et al*. developed the novel scoring functions of protein–ligand, protein–protein and protein–nucleic acid interactions, named SPA,[36] SPA-PP,[37] and SPA-PN,[38] respectively. The optimization strategy is to maximize the performances on intrinsic specificity and affinity predictions simultaneously by tuning the energy parameters according to the funnel-like energy landscape. In addition, Zheng *et al*. performed two-dimensional virtual screening with ISR and affinity evaluated by the scoring function SPA[36] against flexible Ras protein and protein–protein interactions involving the Ras protein.[138]

3.2. The kinetic specificity and its applications in drug discovery

In the equilibrium condition, the binding free energy *in vitro* can be computed by the kinetic rate constants according to the equation quantifying the relation between the kinetic constants and the thermodynamic features *in vitro*. However, the condition *in vivo* is usually a nonequilibrium condition. Consequently, the effectiveness of ligand binding may be affected not only by the thermodynamic properties, but also by the dissociation rate (off-rate constant) determining the residence time (half life).[139,140] The longer the residence time is, the more potent the drug is on the the receptor *in vivo*. Therefore, not only the thermodynamic affinity and specificity, but also the residence time of the drug on the receptor quantified as the kinetic specificity, determine the duration and effectiveness of the drug.

To calculate the residence time practically, a diffusion process on the energy landscape was used to model the ligand–receptor binding/unbinding kinetics.[141] In this way, the residence time can be obtained approximately from the kinetic time via the diffusion dynamics. The driving force of this process is determined by the gradient of the free energy landscape. By projecting the free energy onto a reaction coordinate or order parameter, the binding/unbinding kinetics on the energy landscape can be modeled by the following one-dimensional diffusion equation:[142,143]

$$\frac{\partial}{\partial t}P(r,t) = \frac{\partial}{\partial r}\left\{D(r)\left[\frac{\partial P(r,t)}{\partial r} + P(r,t)\frac{\partial (F(r)/k_B T)}{\partial r}\right]\right\}, \tag{4}$$

where $P(r,t)$ is the probability distribution of the binding conformation with specific r (RMSD, root mean square displacement) at time t, $D(r)$ denotes the diffusion coefficient which is in general a function of r (for simplicity, we consider constant diffusion coefficient D in the following calculation), and $F(r)$ represents the free energy of the binding system at a given r. The residence time (RT) for the native binding conformation can then be calculated by integrating the above diffusion equation

$$\mathrm{RT} = \int_r^{r_u} \mathrm{d}r \int_{r_n}^{r'} \mathrm{d}r' \frac{\exp\left[\frac{F(r)-F(r')}{k_B T}\right]}{D}. \tag{5}$$

At $r_n = 0$ where the system is in the native state, the reflecting boundary condition is imposed. At r_u where the binding is in the "non-native" states, the absorbing condition is used. Note that the choice of r_u is system-specific.

The kinetic specificity quantified by the calculated residence time showed moderate correlations with both thermodynamic stability and specificity,[141] indicating the thermodynamics and kinetics of ligand binding can be optimized simultaneously. Practically, the intrinsic specificity ratio ISR is less difficult to obtain computationally than experimentally. In contrast, the kinetic specificity is less difficult to obtain experimentally than computationally. The microscopic molecular modeling (molecular dynamics or Monte Carlo simulations) demands tremendous computational time. The coarse-grained diffusion approach here is relatively fast and provides qualitative/semi-quantitative results, although not as accurate as MD or Monte Carlo simulations. The connection of thermodynamics and kinetics bridges a gap in the understanding of drug efficacy. It is also helpful for drug discovery and design. In drug screening, both experimentally and computationally, efforts to identify new lead compounds often concentrated merely on the optimization of the affinity.[144] Such a strategy neglected the importance of kinetic and thermodynamic specificities. Based on the previous studies on two-dimensional virtual screenings relying on thermodynamic specificity and binding affinity against the Ras protein and Ache protein, Wang *et al.* have further developed three-dimensional virtual screening taking into account the residence time describing the kinetic specificity. In the drug–target binding system, what determines the residence time of a drug residing in its target is the gap of binding affinity between the native state and the transition state.[139,140,145–148] Therefore, improving the stability of the native state could decrease the drug potency *in vivo*, if the affinity change of the transition state is larger than that of the native state. Also, improvement of the affinity may decrease the thermodynamic specificity, if the affinity gap between the native and non-native modes is reduced. This demonstrates the importance of taking the thermodynamic specificity and kinetic specificity as additional criteria for drug discovery.

Based on the energy landscape theory, ISR as quantified thermodynamic specificity and the residence time as quantified kinetic specificity were both quantified computationally.[141] This gives a quantitative means to choose and optimize the lead compounds computationally with multi-dimensional drug screening strategy by considering both affinity and specificities. Yan *et al.*[69] has tested this strategy on the drug targets COXs. As seen in Fig. 10(a), two- or three-dimensional drug screenings (E+ISR, E+RT, or E+ISR+RT) are superior to any single parameter in discriminating the target COX-2 against COX-1 when binding with the selective drugs for COX-2. In addition, three-dimensional screening shows more capability than two-dimensional screening in discriminating COX-2 against COX-1. Also, two- or three-dimensional drug screening enhances discrimination between the selective and non-selective drugs when binding onto COX-2 (Fig. 10(b)). Therefore, it is promising to conduct multi-dimensional (such as two- or three-dimensional) drug screening and target identification by incorporating the thermodynamic and kinetic specificities as selection criteria, instead of using affinity alone. Specificity can be used as an indicator in addition to affinity in drug screening to avoid side effects and seek more specific lead compounds.

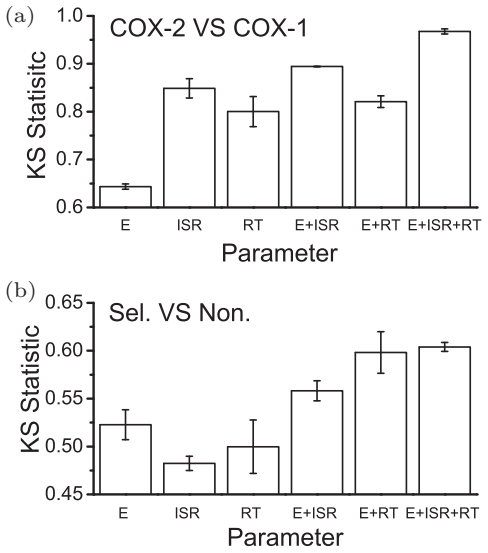


Fig. 10. Two- and three-dimensional drug and target screening. (a) Target screening: the KS statistic for the discrimination of the COX-2 against COX-1 when complexing with the selective drugs for COX-2. (b) Drug screening: the KS statistic of discrimination of the selective drugs against non-selective drugs when complexing with COX-2.[141]

4. Landscape and flux theory and its applications to biological networks

4.1. Landscape and flux theory for nonequilibrium systems

In general, biological systems are nonequilibrium dynamical systems with complex behaviors and varied functions.[9,10,13,15,40–43,72–78] The deterministic dynamics of a generic class of nonequilibrium biological networks can be mathematically modeled by a set of ordinary differential equations, which can be written in the compact form

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}), \tag{6}$$

where $\dot{\boldsymbol{x}}$ is a short notation for $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$, the vector $\boldsymbol{x}$ denotes the collection of variables characterizing the system state, and $\boldsymbol{F}(\boldsymbol{x})$ is the driving force governing the system's deterministic dynamics.

In reality, biological environments are noisy and full of intrinsic and external fluctuations. Taking into account the effects of fluctuations, the Langevin equation usually gives a reasonable description of the system's stochastic dynamics[40–43]

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{\xi}(\boldsymbol{x}, t). \tag{7}$$

In addition to the deterministic driving force $\boldsymbol{F}(\boldsymbol{x})$, there is also the stochastic force $\boldsymbol{\xi}(\boldsymbol{x},t)$. It has the statistical properties $\langle\boldsymbol{\xi}(\boldsymbol{x},t)\rangle = 0$ and $\langle\boldsymbol{\xi}(\boldsymbol{x},t)\boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x},t')\rangle = 2\varepsilon\boldsymbol{D}(\boldsymbol{x})\delta(t-t')$, where ε is a scale factor characterizing the fluctuation strength (not necessarily small) and $\boldsymbol{D}(\boldsymbol{x})$ is the diffusion matrix characterizing the fluctuation correlations.[40–43]

Corresponding to the stochastic dynamics described by the Langevin equation which traces a stochastic trajectory, the probability distribution evolves according to the Fokker–Planck equation (or diffusion equation), which in Ito's prescription reads[40–43]

$$\frac{\partial}{\partial t}P(\boldsymbol{x},t) = -\boldsymbol{\nabla}\cdot\{\boldsymbol{F}(\boldsymbol{x})P(\boldsymbol{x},t) - \varepsilon\boldsymbol{\nabla}\cdot[\boldsymbol{D}(\boldsymbol{x})P(\boldsymbol{x},t)]\}. \quad (8)$$

Interpreting this equation as a continuity equation $\partial_t P(\boldsymbol{x},t) + \boldsymbol{\nabla}\cdot\boldsymbol{J}(\boldsymbol{x},t) = 0$, which represents the probability conservation, the probability flux is identified as

$$\boldsymbol{J}(\boldsymbol{x},t) = \boldsymbol{F}(\boldsymbol{x})P(\boldsymbol{x},t) - \varepsilon\boldsymbol{\nabla}\cdot[\boldsymbol{D}(\boldsymbol{x})P(\boldsymbol{x},t)]. \quad (9)$$

In the long time limit, in general, the system settles in the steady state, $P_{\mathrm{ss}}(\boldsymbol{x})$, which does not vary with time: $\partial_t P_{\mathrm{ss}} = 0$.[40–43] Accordingly, the steady-state probability flux satisfies $\boldsymbol{\nabla}\cdot\boldsymbol{J}_{\mathrm{ss}} = 0$, namely, the divergence-free condition. There are two physically distinct ways for $\boldsymbol{J}_{\mathrm{ss}}$ to have vanishing divergence. One way is that the steady-state flux itself is zero, $\boldsymbol{J}_{\mathrm{ss}} = 0$, which indicates no flux in or out of the ensemble and characterizes the equilibrium condition satisfying detailed balance.[40–43] The other way is that the steady-state flux $\boldsymbol{J}_{\mathrm{ss}}$ itself does not vanish although its divergence is identically zero. Nonvanishing steady-state probability flux $\boldsymbol{J}_{\mathrm{ss}}$ quantifies the nonequilibrium characteristics of the steady state with detailed balance broken.[9,10,40–43,83] Due to the divergence-free condition, $\boldsymbol{J}_{\mathrm{ss}}$ has a curl nature; its field lines typically circulate in loops (though they can also extend into infinity).

Applying Eq. (9) to the steady state, the driving force decomposition follows[9,10,83,149]

$$\boldsymbol{F} = -\varepsilon\boldsymbol{D}\cdot\boldsymbol{\nabla}U + \boldsymbol{V}_{\mathrm{ss}} + \varepsilon\boldsymbol{\nabla}\cdot\boldsymbol{D}, \quad (10)$$

where $U = -\ln P_{\mathrm{ss}}$ is the potential landscape and $\boldsymbol{V}_{\mathrm{ss}} = \boldsymbol{J}_{\mathrm{ss}}/P_{\mathrm{ss}}$ is the flux velocity. Thus the driving force of nonequilibrium dynamical systems is decomposed into three parts: the gradient-like force of the potential landscape U related to the steady-state probability distribution P_{ss}; the curl-like force of the flux velocity $\boldsymbol{V}_{\mathrm{ss}}$ connected to the steady-state probability flux $\boldsymbol{J}_{\mathrm{ss}}$; and the fluctuation-induced force arising from the dependence of fluctuations on state variables, which disappears when $\boldsymbol{D}$ does not depend on $\boldsymbol{x}$. For equilibrium systems, in which $\boldsymbol{J}_{\mathrm{ss}}$ and thus $\boldsymbol{V}_{\mathrm{ss}}$ vanish, the equilibrium potential landscape characterizes the system's global stability of the system, while the gradient of the landscape determines the system's dynamics. For nonequilibrium systems, in which $\boldsymbol{J}_{\mathrm{ss}}$ and thus $\boldsymbol{V}_{\mathrm{ss}}$ do not vanish, the nonequilibrium potential landscape characterizes the system's global stability, while both the gradient force of the potential landscape and the curl force of the flux velocity determine the system's dynamics.

In the small fluctuation regime, the stochastic dynamics of the system can be investigated using the WKB method. The Fokker–Planck equation is expanded in terms of the fluctuation strength parameter ε. The zero-fluctuation limit results in the following Hamilton–Jacobian equation:[22,23,41,83,150]

$$\boldsymbol{F}\cdot\boldsymbol{\nabla}\phi_0 + \boldsymbol{\nabla}\phi_0\cdot\boldsymbol{D}\cdot\boldsymbol{\nabla}\phi_0 = 0, \quad (11)$$

where $\phi_0 = (\varepsilon U)|_{\varepsilon\to 0}$ is called the intrinsic potential. Following the deterministic dynamics $\dot{\boldsymbol{x}} = \boldsymbol{F}$, it can be shown that ϕ_0 decreases monotonically with time

$$\mathrm{d}\phi_0(\boldsymbol{x})/\mathrm{d}t = \boldsymbol{F}\cdot\boldsymbol{\nabla}\phi_0 = -\boldsymbol{\nabla}\phi_0\cdot\boldsymbol{D}\cdot\boldsymbol{\nabla}\phi_0 \le 0. \quad (12)$$

The equality holds when $\boldsymbol{\nabla}\phi_0 = 0$, i.e., ϕ_0 has reached its extremum. This means ϕ_0 is a Lyapunov function, well suited to studying the global stability of deterministic nonequilibrium systems.[22,23,41,83,150] Moreover, $\boldsymbol{V}_{\mathrm{ss}}$ can also be expanded in terms of the fluctuation strength ε. The leading order is shown to be $\boldsymbol{V}_0 = \boldsymbol{F} + \boldsymbol{D}\cdot\boldsymbol{\nabla}\phi_0$,[22,23,83,150] where $\boldsymbol{V}_0 = \boldsymbol{V}_{\mathrm{ss}}|_{\varepsilon\to 0} = (\boldsymbol{J}_{\mathrm{ss}}/P_{\mathrm{ss}})|_{\varepsilon\to 0}$ is the intrinsic flux velocity. Equation (11) shows that

$$\boldsymbol{V}_0\cdot\boldsymbol{\nabla}\phi_0 = 0, \quad (13)$$

which means the gradient of the intrinsic potential and the intrinsic flux velocity are perpendicular to each other. In the zero-fluctuation limit, the driving force decomposition, in terms of the intrinsic potential ϕ_0 and the intrinsic flux velocity $\boldsymbol{V}_0$, reads[22,23,41,83,150]

$$\boldsymbol{F} = -\boldsymbol{D}\cdot\boldsymbol{\nabla}\phi_0 + \boldsymbol{V}_0. \quad (14)$$

Hence, both the intrinsic potential landscape and the intrinsic flux velocity contribute to the deterministic dynamics of nonequilibrium systems. Furthermore, the dynamical decomposition of the driving force into the potential landscape and the flux velocity has been further generalized to accommodate transient processes, time-dependent external conditions, multiple state-transition mechanisms, and spatially extended systems, of which we refer to Refs. [82] and [83] for a more complete presentation.

Based on the potential–flux dynamical decomposition, we investigated and extended the fluctuation–dissipation theorem, nonequilibrium thermodynamics, and gauge field formulation.[15,83] When generalized to the nonequilibrium regime,[15,151–154] the fluctuation–dissipation theorem for an observable Ω and perturbation on component i of the driving force reads[15]

$$\begin{aligned} R_i^{\Omega}(t-t') &= -\langle\Omega(t)\partial_i \ln P\rangle \\ &= -\langle\Omega(t)D_{ik}^{-1}(t')\widetilde{F}_k(t')\rangle \end{aligned}$$

$$+\langle \Omega(t) D_{ik}^{-1}(t') V_k(t') \rangle, \tag{15}$$

where $\partial_i \equiv \partial/\partial x_i$, $\widetilde{F}_k = F_k - \partial_j D_{kj}$ is the effective driving force, and $V_k = J_k/P$ is the transient flux velocity.[83] The Einstein notation of summing over repeated indices is used. $\langle\cdot\rangle$ represents the average over the ensemble distribution P. We have also set $\varepsilon = 1$ (the same as below unless specified otherwise). Note that the probability flux (velocity) defined here in the Fokker–Planck equation differs from that used in Ref. [15] by a minus sign (the sign used here agrees with the convention in physics). The left-hand side of the equation is the response function of the observable to perturbation. On the right-hand side of the equation, the first term is the correlation function of the observable with the effective driving force, representing the fluctuation properties of the system (similar to the equilibrium case); the second term arises from the flux force breaking detailed balance which characterizes the nonequilibrium condition (including the nonequilibrium transient process).[83]

To make contact with the nonequilibrium thermodynamics, we choose the observable $\Omega = V_i$ and take the equal time limit $t = t'$ in Eq. (15), which gives[15]

$$\dot{\mathscr{S}} = -\langle V_i \partial_i \ln P \rangle = -\langle V_i D_{ik}^{-1} \widetilde{F}_k \rangle + \langle V_i D_{ik}^{-1} V_k \rangle, \tag{16}$$

where the index i is also summed over. The left side is the rate of change of system entropy $\mathscr{S} = \langle -\ln P \rangle$. The first term on the right (without the minus sign) is identified as the rate of entropy flowing from the system to the environment $\dot{\mathscr{S}}_{\rm m} = \langle V_i D_{ik}^{-1} \widetilde{F}_k \rangle$. Its relation to the average heat dissipation rate in the environment (medium) is given by $\dot{\mathscr{Q}}_{\rm m} = T\dot{\mathscr{S}}_{\rm m}$ when the environmental temperature T is constant. The second term on the right, which is non-negative, is identified as the entropy production rate $\dot{\mathscr{S}}_{\rm tot} = \langle V_i D_{ik}^{-1} V_k \rangle \geq 0$, driven by the flux velocity. Therefore, equation (16) can be interpreted as the entropy balance equation $\dot{\mathscr{S}} = \dot{\mathscr{S}}_{\rm tot} - \dot{\mathscr{S}}_{\rm m}$.[15,83]

Further, taking the observable as the relative flux velocity $V_i^{\rm rel} \equiv V_i - V_i^{\rm ss}$[83] and following similar procedures as above, we obtain[15,83,151–154]

$$\begin{aligned} -\dot{\mathscr{F}}/T &= -\langle V_i \partial_i \ln(P_{\rm ss}/P) \rangle \\ &= \langle V_i D_{ik}^{-1} V_k \rangle - \langle V_i^{\rm ss} D_{ik}^{-1} V_k \rangle, \end{aligned} \tag{17}$$

where $\mathscr{F}$ is the nonequilibrium free energy defined as $\mathscr{F} = T\langle \ln(P_{\rm ss}/P) \rangle = \mathscr{U} - T\mathscr{S}$, with the nonequilibrium internal energy $\mathscr{U} = T\langle -\ln P_{\rm ss} \rangle$. These definitions apply when the environmental temperature is constant and there are no time-dependent external conditions.[83] The first term on the right side is the total entropy production rate $\dot{\mathscr{S}}_{\rm tot}$. The second term, which is non-negative, is identified as the adiabatic entropy production rate $\dot{\mathscr{S}}_{\rm ad} = \langle V_i^{\rm ss} D_{ij}^{-1} V_j \rangle = \langle V_i^{\rm ss} D_{ij}^{-1} V_j^{\rm ss} \rangle \geq 0$. It is related to the house keeping heat via $\dot{\mathscr{S}}_{\rm ad} = \dot{\mathscr{Q}}_{\rm hk}/T$ when the environmental temperature is constant. The term on the left side, $-\dot{\mathscr{F}}/T$, is also non-negative and identified as the non-adiabatic entropy production rate $\dot{\mathscr{S}}_{\rm na} = -\dot{\mathscr{F}}/T \geq 0$. Hence, equation (17) can be interpreted as the entropy production decomposition equation $\dot{\mathscr{S}}_{\rm tot} = \dot{\mathscr{S}}_{\rm ad} + \dot{\mathscr{S}}_{\rm na}$.[83] The entropy production rate can be decomposed into the contribution of spontaneous relaxation to the steady state from the transient state and that of detailed balance breaking in the steady state driven by the steady-state flux sustaining the nonequilibrium environment.[15,83,155–158]

Moreover, a more general set of nonequilibrium thermodynamic equations, taking into account time-dependent external conditions, can be constructed directly from the potential–flux dynamical decomposition, which includes the above two equations as special cases[83]

$$\begin{cases} \dot{\mathscr{U}} &= \dot{\mathscr{W}}_{\rm ed} - \dot{\mathscr{Q}}_{\rm ex}, \\ \dot{\mathscr{S}} &= \dot{\mathscr{S}}_{\rm tot} - \dot{\mathscr{S}}_{\rm m}, \\ \dot{\mathscr{F}} &= \dot{\mathscr{W}}_{\rm ed} - \dot{\mathscr{A}}_{\rm sd}, \\ \dot{\mathscr{S}}_{\rm tot} &= \dot{\mathscr{S}}_{\rm ad} + \dot{\mathscr{S}}_{\rm na}, \\ \dot{\mathscr{S}}_{\rm m} &= \dot{\mathscr{S}}_{\rm ad} + \dot{\mathscr{S}}_{\rm ex}, \end{cases} \tag{18}$$

where $\dot{\mathscr{S}}_{\rm m} = \dot{\mathscr{Q}}_{\rm m}/T$, $\dot{\mathscr{S}}_{\rm ad} = \dot{\mathscr{Q}}_{\rm hk}/T$, $\dot{\mathscr{S}}_{\rm ex} = \dot{\mathscr{Q}}_{\rm ex}/T$, $\dot{\mathscr{S}}_{\rm na} = \dot{A}_{\rm sd}/T$ and we have adapted the notations to those used here. The effect of time-dependent external conditions is represented by the thermodynamic power (work per unit time) $\dot{\mathscr{W}}_{\rm ed}$. A further extension of this set of equations and its application in a spatial stochastic neuronal model can be found in Ref. [83].

The connection to the gauge field theory[15] is realized by the covariant derivative $\boldsymbol{\nabla}_i = \partial_i + A_i$, with the gauge field identified as $A_i = -D_{ij}^{-1}\widetilde{F}_j$. The probability flux can be rewritten as $J_i = -D_{ij}\boldsymbol{\nabla}_j P$. Associated with the Abelian gauge field A_i, there is an internal space with curvature $R_{ij} = [\boldsymbol{\nabla}_i, \boldsymbol{\nabla}_j] = \partial_i A_j - \partial_j A_i$. R_{ij} is gauge invariant, meaning for a gauge transformation $A_i \to A_i + \partial_i \Lambda$, we have $R_{ij} \to R'_{ij} = R_{ij}$. For systems with detailed balance $\boldsymbol{V}_{\rm ss} = 0$, $A_i = \partial_i U$ is a gradient field and the curvature $R_{ij} = 0$, indicating a flat internal space. For nonequilibrium systems with detailed balance breaking, $A_i = \partial_i U - D_{ij}^{-1} V_j^{\rm ss}$ is not a gradient field and $R_{ij} = -[\partial_i (D_{jk}^{-1} V_k^{\rm ss}) - \partial_j (D_{ik}^{-1} V_k^{\rm ss})]$ is non-zero indicating a curved internal space. This gauge-invariant curvature is intimately connected to the entropy flow along any closed loop[15]

$$\begin{aligned} \Delta s_{\rm m}^C &= -\oint_C A_i(\boldsymbol{x})\,{\rm d}x_i \\ &= \oint_C D_{ij}^{-1}(\boldsymbol{x}) V_j^{\rm ss}(\boldsymbol{x})\,{\rm d}x_i \\ &= -\frac{1}{2}\int_\Sigma R_{ij}\,{\rm d}\sigma_{ij}. \end{aligned} \tag{19}$$

Here Σ represents the surface spanning the closed loop C and ${\rm d}\sigma_{ij}$ denotes the area element on Σ.

A further connection can be made with $\Omega = x_j$ in Eq. (15), giving for constant D_{ij}[15]

$$R_i^j(t-t') = -D_{ik}^{-1}\frac{\rm d}{{\rm d}t}\langle x_k(t) x_j(t') \rangle$$

$$- D_{ik}^{-1} \int \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}' x_j F_k(\boldsymbol{x}') A(\boldsymbol{x},\boldsymbol{x}',t-t') - D_{ik}^{-1}\langle x_j(t) V_k^{\mathrm{ss}}(t')\rangle, \quad (20)$$

where $A(\boldsymbol{x},\boldsymbol{x}',t-t') = P_{\mathrm{ss}}(\boldsymbol{x}')\int \mathscr{D}[\boldsymbol{x}] W(\boldsymbol{x},t|\boldsymbol{x}',t')[1-\exp(-\Delta s_{\mathrm{m}})P_{\mathrm{ss}}(\boldsymbol{x})/P_{\mathrm{ss}}(\boldsymbol{x}')]$ with $W(\boldsymbol{x},t|\boldsymbol{x}',t')$ denoting the transition probability. The expression $U(\boldsymbol{x},\boldsymbol{y}) = \exp(\Delta s_{\mathrm{m}}) = \exp(-\int_P A_i \mathrm{d}x_i)$, where $\int_P$ represents integration on the path from $\boldsymbol{x}$ to $\boldsymbol{y}$, is analogous to the Wilson line or Wilson loop (for closed path) in the Abelian gauge theory. It characterizes nonequilibrium irreversibility in the current context. Under the gauge transformation, $U(\boldsymbol{x},\boldsymbol{y})$ transforms as $U(\boldsymbol{x},\boldsymbol{y}) \to \exp[\Lambda(\boldsymbol{x})]U(\boldsymbol{x},\boldsymbol{y})\exp[-\Lambda(\boldsymbol{y})]$. It also satisfies $\dot{y}_i \nabla_i U(\boldsymbol{x},\boldsymbol{y}) = 0$, which means the covariant gradient of the Wilson line integral is perpendicular to the dynamical path.[15]

4.2. Landscape and flux of the cell cycle

The process of the cell cycle takes place in cell replication and division. It is important for cell growth, cell development, cell proliferation, and cell death. Studying the nature of the cell cycle process is important for exploring the cell functions. The cell cycle has some certain phases, the G1 phase, S phase, G2 phase, M phase, and some check points regulating the process of the cell cycle. It was shown that cell cycle gene regulatory networks control the process of a cell cycle.[10,11,20,21,159,160]

The cell cycle gene regulatory networks have been proposed already.[20,21,159] We explored the dynamics of the cell cycle in the landscape and flux theoretical framework,[10,11,160] using the cell cycle gene regulatory network proposed by Goldbeter.[11,20,21] The state of the cell cycle gene regulatory network was characterized by a vector with 44 components. The deterministic dynamics of the network was modeled by a set of nonlinear ordinary differential equations.[11]

Due to the existence of extrinsic fluctuations and intrinsic fluctuations, the dynamics of the cell cycle network is stochastic and governed by the Langevin and Fokker–Planck equation. To decrease the high dimensionality that stands as an obstacle for solving the Fokker–Planck equation practically, we applied the self-consistent mean field approximation.[10,11,18] Then we got the probability of the steady state P_{ss} and the corresponding potential landscape $U = -\ln P_{\mathrm{ss}}$.

We projected the landscape and flux, as shown in Fig. 11, onto the two-dimensional space spanned by the concentrations of the two proteins: cyc E and cyc A. In Fig. 11(a), the process of the cell cycle is illustrated with the G1 phase, S phase, G2 phase, and M phase, as well as the checkpoints.[11,20,21] In Fig. 11(b), the potential landscape U is displayed, on which there is a closed cell cycle oscillation.[11] Figure 11(c) shows the two-dimensional potential landscape and flux. The curl flux force and the negative gradient of the landscape are indicated, respectively, by white arrows and red arrows.

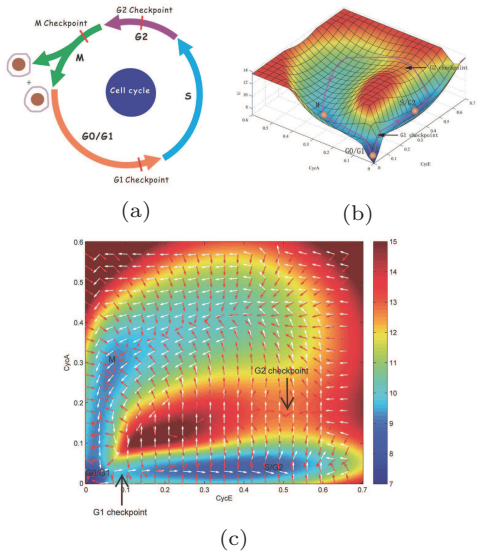


Fig. 11. (a) The phases with checkpoints for the cell cycle system: the G1 phase, S phase, G2 phase, M phase. (b) G1 phase, S/G2 phase, M phase, G1 checkpoint, and S/G2 checkpoint on the potential landscape. (c) The two-dimensional landscape (the non-zero curl flux is shown in white arrows; the negative gradient of the landscape is shown in red arrows). The diffusion parameter is set as $D = 0.05$.[11]

The driving force is composed of the flux and the landscape in this nonequilibrium cell cycle dynamical system. The curl flux force (flux velocity) is an essential component of the nonequilibrium driving force for the cell cycle, which characterizes the nonequilibrium nature of the cell cycle network.[9–11,13–15,23] It is a rotating vector maintaining the cell cycle oscillation. The input of energy is supplied by the cell nutrition in the cell cycle. The potential landscape quantifies the global robustness and stability for the nonequilibrium cell cycle system.[9–11,13–15,23,149,160]

We can see from Fig. 11 that the cell states get attracted to the Mexican-hat-like limit cycle valley by the force of the negative gradient of the potential landscape. The local attractors on the cell cycle landscape are identified as the G1 phase, S/G2 phase, and M phase. The transition states between different attractors (cell cycle phases) on the paths are identified as the checkpoints. The dynamics is governed by the curl flux force and local barriers between the basins along the oscillation paths in the oscillation valley. This provides a global and quantitative approach to study the dynamics of the cell cycle process.

We also investigated the global sensitivity analysis and the function of this system. The barrier and the flux reflecting genes and regulatory connections in the cell cycle network can determine the cell cycle function. We show in Fig. 12(a) the variations in the central barrier ($\mathrm{Barrier}_{\mathrm{Center}}$), flux, and period when parameters are changed. In Fig. 12(b), we show the changes in $\mathrm{Barrier}_{\mathrm{G1/S}}$ and $\mathrm{Barrier}_{\mathrm{G2/M}}$ as the parameters vary.[11]

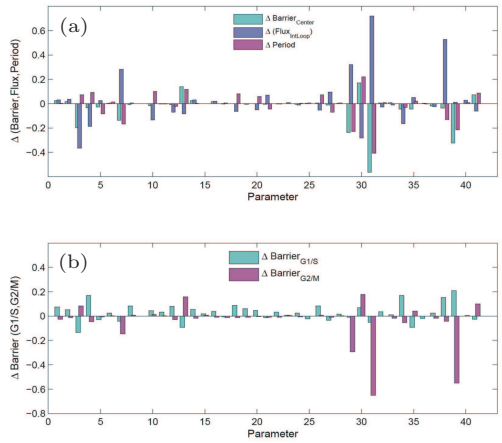


Fig. 12. Global sensitivity analysis. The barrier ($\text{Barrier}_{\text{Center}}$ is for global stability, $\text{Barrier}_{\text{G1/S}}$ is for the G1 checkpoint, $\text{Barrier}_{\text{G2/M}}$ is for the checkpoint of S/G2), flux and period change as the parameters vary. The horizontal coordinates display 41 parameters. Each parameter is changed 10%.[11]

Using the method of global sensitivity analysis, we have identified important regulations and genes affecting the cell cycle function. On the experimental aspect, a few key genes and regulations of the cell cycle have been discovered and verified. The genes and regulations we identified as being important from global sensitivity analysis are consistent with the results from experiments. The pRB activation can decrease the effect of the curl flux and increase the barrier height $\text{Barrier}_{\text{G2/M}}$. It can also make the process of the cell cycle longer.[11]

The cancer cell cycle is an uncontrollable cycling. Its period is much shorter than that of the normal cell. The two components of the driving force, the barrier force and the flux force, are not balanced in the cancer cell cycle. The impact of curl flux overcomes the landscape barrier on the oscillation path. Therefore, decreasing the flux or increasing the barriers may restore the balance between the landscape barrier and curl flux so that the cancer cells can return to the normal state. This process may be realized by varying the genes, the nutrition supply or regulatory connections. We can apply the global sensitivity analysis to identify these variations of gene or environment.

The potential landscape and curl flux theory has also been applied to investigate the mammalian cell cycle.[10,11,160] The barrier height on the landscape is a measurement of the global cell cycle robustness and stability. There is a periodic oscillation with four phases and three checkpoints on the cycle path. The gradient of landscape guides the cell cycle system to settle on the cell cycle oscillation. Along the limit cycle, the system is driven by the curl flux. Both landscape and flux direct the cell cycle. Investigation on the mechanism of checkpoints has also been conducted in the context of the landscape and flux framework.[9–11] Global sensitivity analysis of the cell cycle has identified important genes and regulations that determine the global cell cycle function.[9–11] Some of our results have been confirmed in experiments, and some findings can be used to predict potential anticancer strategies.

4.3. Landscape, paths, and kinetic rates: stem cell differentiation, development, and reprogramming

4.3.1. Landscape of stem cell differentiation, development, and reprogramming

In the 1940s, Waddington proposed a famous emblem of the epigenetic landscape in the cell development and differentiation.[161] The growing cell is analogous to a ball rolling from the top of the landscape to the bottom in Waddington's image. When the ball encounters a bifurcation of the landscape every time, it will choose one of the alternative trajectories eventually. This process was considered as a metaphor of cell differentiation. However, the picture only provides a qualitative understanding of the developmental process of cells. Recently, considerable theoretical effort has been devoted to the cell fate problem.[14,16,18,162–168]

To analyze the process of cell development and differentiation quantitatively, we have utilized the landscape and flux framework. Waddington's landscape picture can be quantified through investigating an important example of the cell development network consisting of a pair of self-activating and interactively repressing genes. The gene regulatory network governing a binary cell fate decision model has been found in the multipotent stem cells of various tissues. The cell fate determining model masters transcription factors, gene X_1 and gene X_2. The relative expression levels x_1 and x_2 of genes X_1 and X_2 characterize the state of the network. The dynamics of the model is represented by a set of ordinary differential equations.[14,165] For stochastic dynamics with fluctuations, we can obtain the steady-state probability distribution and derive the potential landscape $U = -\ln P_{\text{ss}}$. The developmental/reprogramming paths are quantified by the dominant paths using the path integral method.

In Fig. 13, we have quantified the Waddington landscape of cell development and differentiation. A certain stage of the cell development is indicated by a specific value for the self-activation parameter a. The full landscape and dynamical paths are depicted by decreasing a. Our quantified Waddington landscape is qualitatively similar to the original Waddington picture. However, there are some important differences. Firstly, the stem cell state in our quantified landscape is stable with a basin of attraction, while in the Waddington picture, this state is regarded as an unstable state at the top of the landscape. Secondly, our development/differentiation pathways and reprogramming paths are irreversible, while in the Waddington picture, the paths are expected to be reversible. Thirdly, in our picture, the development/differentiation process is initiated by induction, fluctuations, and/or genetic/environmental changes, while in the Waddington picture, the development and differentiation process is spontaneous. Finally, we have specified

the direction of development by quantifying the changes of specific gene regulation parameter in the developmental process (self activation in this case), while in the Waddington picture, the direction of development is not specified. In short, our approach provides the quantitative foundation and specifies the key ingredients for the Waddington landscape.[14]

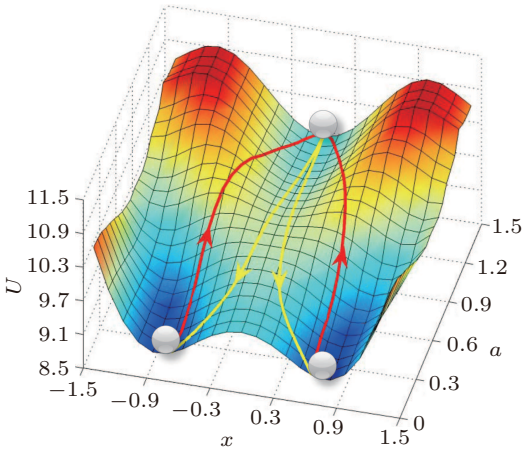


Fig. 13. The quantified Waddington developmental landscape and pathways.[14]

In addition, we have also studied the human embryonic stem cell network, which is composed of nine nodes after integrating previous known networks.[169,170] The dynamics of this network is represented as ordinary differential equations with stochastic forces when fluctuations are taken into account. Within these nine nodes, NANOG is a primary stem cell marker and its higher expression level indicates more stemness. GATA6 and CDX2 are considered as two differentiation markers. Higher GATA6 or CDX2 expression levels indicate higher differentiation. Different combinations of gene expression levels give rise to the stem cell state (attractor), differentiation state, or intermediate states. This human embryonic stem cell network can also be studied quantitatively in the landscape and flux framework.[166]

4.3.2. Paths of stem cell differentiation, development, and reprogramming

The biological pathways of stem cell differentiation/development and reprogramming can be studied using the path integral approach. The path integral method has found wide applications in different areas of physics and chemistry since it was proposed. It has now been developed and applied to biological systems. Generally speaking, the dynamics of a spatially homogeneous biological system is represented by a set of ordinary differential equations with stochastic forces accounting for fluctuations. An insightful approach to study stochastic dynamical systems is to consider the transition probability from the initial state $\boldsymbol{x}_{\rm i}$ at $t = t_{\rm i}$ to the final state of $\boldsymbol{x}_{\rm f}$ at time $t = t_{\rm f}$. To avoid complications, we consider a constant diffusion matrix. The transition probability is then given by[13,171,172]

$$P(\boldsymbol{x}_{\rm f}, t_{\rm f}; \boldsymbol{x}_i, t_i) = \int \mathscr{D}[\boldsymbol{x}(t)] \exp\{-S[\boldsymbol{x}(t)]\} = \int \mathscr{D}[\boldsymbol{x}(t)] \exp\left\{-\int_{t_{\rm i}}^{t_{\rm f}} L(\boldsymbol{x}(t))\,{\rm d}t\right\}, \quad (21)$$

where $S[\boldsymbol{x}(t)]$ is the action and $L(\boldsymbol{x}) = ({\rm d}\boldsymbol{x}/{\rm d}t - \boldsymbol{F}(\boldsymbol{x}))/4 \cdot \boldsymbol{D}^{-1} \cdot ({\rm d}\boldsymbol{x}/{\rm d}t - \boldsymbol{F}(\boldsymbol{x}))] + \boldsymbol{\nabla} \cdot \boldsymbol{F}(\boldsymbol{x})/2$ is the Lagrangian. Roughly speaking, $\int \mathscr{D}[\boldsymbol{x}(t)]$ denotes summing over all possible paths with fixed end points $\boldsymbol{x}(t_{\rm i}) = \boldsymbol{x}_{\rm i}$ and $\boldsymbol{x}(t_{\rm f}) = \boldsymbol{x}_{\rm f}$; the precise definition of $\mathscr{D}[\boldsymbol{x}(t)]$, however, requires care. This formula is called the Onsager–Machlup functional path integral; it identifies the probability weights for paths. The action $S[\boldsymbol{x}(t)]$ represents the weight contribution of a specific trajectory $\boldsymbol{x}(t)$. Since it is on the exponent (with a minus sign), a path with minimal action has maximal probability. The dominant (or optimal) paths with maximal probability can be obtained by minimizing the action. Equivalently, the dominant paths follow the Euler–Lagrange equation associated with the action. We can approximate the path integral with a set of dominant paths in certain circumstances to study the paths of stochastic biological systems.[167]

The term $\int {\rm d}\boldsymbol{x} \cdot \boldsymbol{D}^{-1}\boldsymbol{F}$ in the action is actually $-\int A_i {\rm d}x_i$ for constant $\boldsymbol{D}$, where A_i is the gauge field discussed previously. It plays an important role in distinguishing nonequilibrium systems from equilibrium systems. In equilibrium systems, this is a surface term only depending on the initial and the finial states which are fixed, making no contribution to the dynamics. As a result, the optimal equilibrium paths are reversible and pass through the saddle point of the equilibrium potential. In contrast, the detailed balance is broken by non-zero flux in nonequilibrium systems. As a result, $\int {\rm d}\boldsymbol{x} \cdot \boldsymbol{D}^{-1}\boldsymbol{F}$ is not a surface term and it contributes to the dynamics in nonequilibrium systems. This is where time irreversibility comes into play. The resulting dominant back and forth paths are not reversible; they have different configurations in the state space. In general, nonequilibrium dominant paths do not go through the saddle point of the nonequilibrium landscape. Yet we can define the "global maximum along the dominant path" characterized by the projection of $\boldsymbol{D}^{-1}\boldsymbol{F}$ along the dominant path being zero, analogous to the saddle point of equilibrium systems.[167]

4.3.3. Kinetic rates of stem cell differentiation, development, and reprogramming

As is known, the equilibrium transition state or Kramers rate is an important concept in equilibrium systems.[173] The forward and backward dominant paths have identical configurations and go through the saddle point of the equilibrium landscape. The equilibrium transition rate from one basin to another for one-dimensional systems with constant diffusion

is given by $r_{\rm K}^{\rm eq} = (D\sqrt{U''(S)|U''(\hat{S})|}/2\pi)\,{\rm e}^{-[U(\hat{S})-U(S)]}$, where S is the bottom of the attractor basin and $\hat{S}$ is the saddle point. $U(x) = -\ln P_{\rm eq}$ is the equilibrium potential landscape and the driving force $F(x) = -DU'(x)$. The exponent factor is associated with the action function. For nonequilibrium biological systems, the driving force is not only the gradient of a potential. For the zero fluctuation limit, researchers have obtained a formula similar to Kramers' original one and the transition state is the saddle point of the landscape.[174,175] But we have found that in the finite fluctuation regime, the back and forth dominant paths are not reversible and do not necessarily get through the saddle point. Therefore, we derived a new analytical formula for the transition state rate by the path integral method[167]

$$r_{\rm K}^{\rm neq} = \frac{\lambda_{\rm u}(\hat{S}')}{2\pi}\sqrt{\frac{{\rm det}\boldsymbol{M}(S)}{|{\rm det}\boldsymbol{M}(\hat{S}')|}}\,{\rm e}^{-S_{\rm HJ}^{\rm DOM}}. \tag{22}$$

As illustrated in Fig. 14, S is the minimum of an attractor, S' is the minimum of another attractor, $\hat{S}$ is the saddle point, and $\hat{S}'$ is a new "saddle point" defined as the "global maximum along the dominant path". $S_{\rm HJ}^{\rm DOM} = \int_S^{\hat{S}'} \boldsymbol{p}\cdot{\rm d}\boldsymbol{x}$ is an action which is an integral along the optimal path from the basin S to the point $\hat{S}'$. And $\lambda_{\rm u}(\hat{S}')$ is the positive eigenvalue of the Jacobian matrix of the driving force. The pre-factor can be solved by the matrix $\boldsymbol{M}(S)$ which satisfies an algebraic equation and matrix $\boldsymbol{M}(\hat{S}')$ which satisfies a dynamic equation.

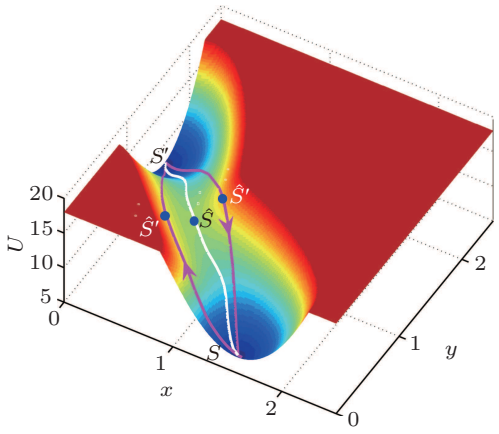


Fig. 14. The three-dimensional nonequilibrium landscape including the gradient path (white line) and the irreversible dominant transition pathways between the basins S and S' (purple lines). Here, $\hat{S}$ denotes the saddle point and $\hat{S}'$ denotes "the global maximum along the dominant path".[167]

4.3.4. Direction of differentiation/development and time reversal symmetry breaking

Both the landscape and flux guide the processes of stem cell differentiation.[14,16,18,19] The probability curl flux is a curling force which breaks detailed balance, leading to the nonequilibrium state of the system. The flux force also drives the biological dominant paths of stem cell differentiation to deviate from the path of steepest descent. Furthermore, due to the flux's curl nature, the forward and backward dominant paths of the stem cell differentiation and reprogramming process are irreversible. They have different configurations and usually do not pass the saddle point. Therefore, the curl flux breaks time reversal symmetry and provides the direction of time. The flux can even be the major driving force for differentiation and reprogramming of the stem cell as the potential landscape is relatively flat. Thus "flux-directed" process can occur in cell development. The arrow of time in cell development cannot be imposed by the potential landscape alone.[12,14,19] It is the curl flux that indicates the arrow of time in cell development. This suggests a resolution of a long standing puzzle of the direction of time in cell development.

4.4. Landscapes and paths of cancer

Conventionally, cancer has been considered as a disease arising from mutation. The landscape and flux theory provides a different perspective on this issue and forms a quantitative foundation for viewing cancer as a diseased state and associated state transformations.[176] The hallmarks of cancer[176–178] have been proposed by Weinberg; these hallmarks can be characterized by key genes of cancer. Cancer states originate from the interactions among genes in a cancer gene regulatory network. These networks are general nonequilibrium dynamical systems. We can study these systems using the potential landscape and flux theory.[39] Our investigation shows that cancer can be viewed as a specific state related to the gene regulatory network. There have also been other studies supporting the view that cancer is a disease associated with a network rather than a disease due to a single mutation.[77,168,179–183]

We investigated the cancer mechanisms and dynamics by the gene network.[39] We focused on the cancer marker genes P53, RB, P21, and PTEN. After a literature search, we constructed a cancer gene regulatory network with 32 gene nodes and 111 regulation edges (shown in Fig. 15).[39] Arrows denote activation and round end points denote repression. Three types of marker genes are shown in the figure: green nodes represent apoptosis marker genes, magenta nodes represent cancer marker genes, and light blue nodes represent tumor repressor genes. Other genes are represented by brown nodes.[39]

Ordinary differential equations can be used to describe the deterministic cancer gene dynamics. The activation and repression interactions among these genes are modeled by Hill functions, which specify the 32 ODEs for this network. We have also considered the stochastic dynamics of this cancer network with fluctuations.[9,10,14] The distribution of the steady-state probability and the corresponding landscape of the cancer regulatory network can be obtained by the method

of the self-consistent mean field approximation.

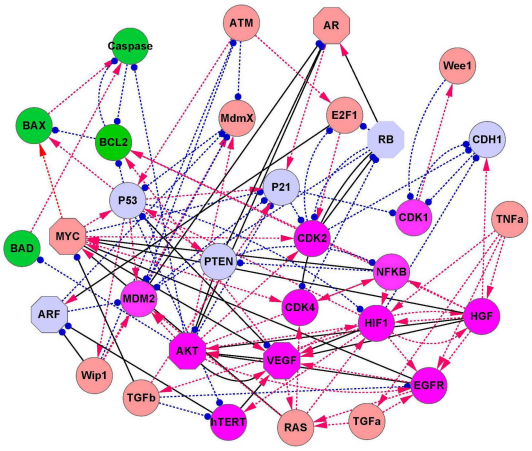


Fig. 15. The diagram of the cancer network which has 32 nodes and 111 edges. Red arrows denote activation while blue end points represent repression. Green nodes represent apoptosis marker genes; magenta nodes represent cancer marker genes; and light blue represent nodes tumor repressor marker genes. The brown nodes represent other genes. The solid black lines denote the key links, while the octagon nodes denote key genes found by the method of global sensitivity analysis.[39]

We reduced the 32-dimensional landscape to a 2-dimensional plane for clarity of presentation. The three-dimensional and two-dimensional landscapes are shown in Fig. 16. The two-dimensional state plane is spanned by the expression levels of two genes, AKT and RB. Figure 16(a) shows three stable states on the tri-stability landscape. The top, middle, and bottom attractors are identified as the cancer, normal, and apoptosis states, respectively.[39] These three attractors correspond to the normal, cancer, and apoptosis states in the cancer gene regulatory network.[176–178] The basins in the gene expression space represent some specific cell types. These attractor states representing different cell types cannot convert to each other easily due to the landscape barriers. We also explored the landscape by altering the nodes (imitating mutations) and regulation strengths (imitating environmental influences).[39]

We explored the dominant pathways connecting the cancer cell state, apoptosis cell state, and the normal cell state using the path integral method.[13,14,18,39] The dominant kinetic pathway from the normal state to the cancer state is shown by the yellow path; the magenta path denotes the dominant kinetic pathway from the cancer state to the normal state in Fig. 16. The dominant pathways between the cancer state and normal states are irreversible. The black pathways represent the dominant paths of the apoptosis processes from normal to death and from cancer to death. We quantified the flux shown in Fig. 16(b). The negative gradient of the landscape and the curl flux are represented by the red arrows and white arrows, respectively. The dynamics of the cancer system is guided by the curl flux and potential landscape.[9]

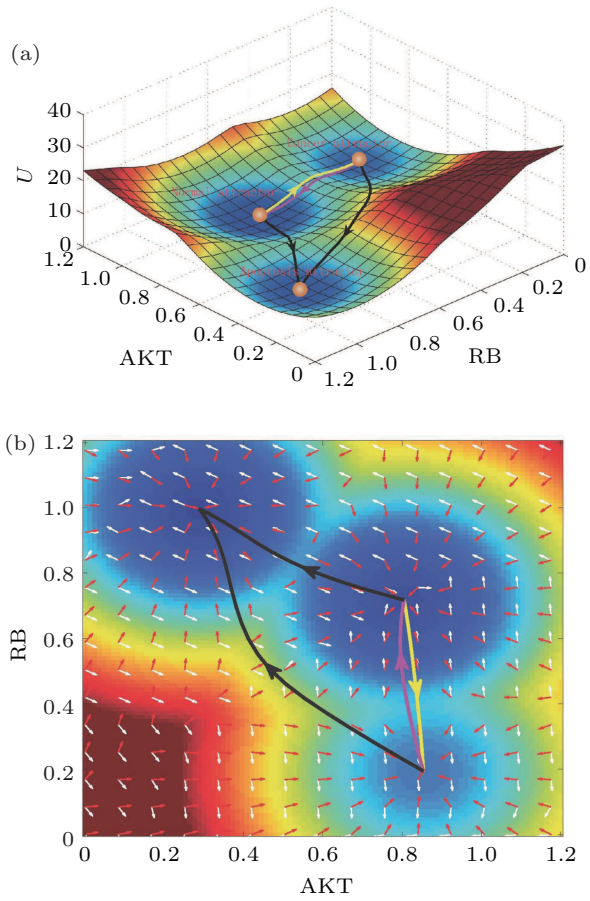


Fig. 16. The potential landscape for the cancer network. The parameters are evaluated at: degradation constant $k = 1$, activation strength $a = 0.5$, repression strength $b = 0.5$, and diffusion constant $D = 0.01$. (a) The potential landscape and the kinetic dominant pathways. (b) Two-dimensional landscape of the cancer network.[39]

Furthermore, we reduced the system to 21 marker genes from the original 32 marker genes and used a binary code to represent the qualitative characteristic of gene expression levels (high or low). The states and paths of the network can be visualized using the simplified binary gene expression levels (2^{21} combinations in totality).[39] Figure 17 shows the cell states and transition jumps on the discrete landscape. The occurrence probabilities of the states and paths are depicted, respectively, by the sizes of nodes and edges. The blue, green, and red nodes denote the cell states that are approaching the normal, apoptosis, and cancer states, respectively. The largest red, blue, and green nodes indicate, respectively, the most significant state of cancer, normal and apoptosis states. The green path in Fig. 17 shows the cancerization process which proceeds from MDM2 on, CDK2 on, RB off, and then to the cancer state. In the reverse pathway of the cancerization process (the magenta path from the cancer state to the normal state in Fig. 17), the cell proceeds from RB on, CDK2 off, CDK4 off, and then to MDM2 off. Hence, the cell may turn on the gene RB at first when transforming from the cancer state to the normal state.[39]

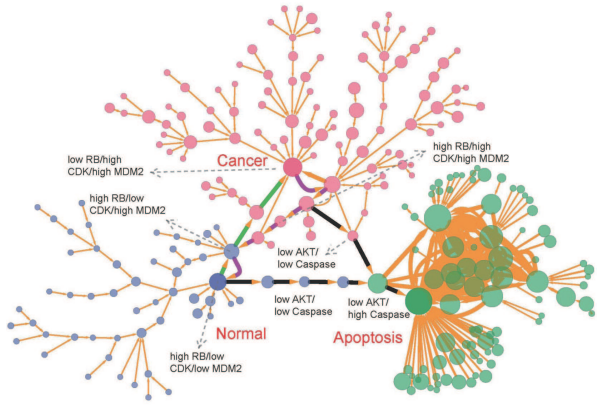


Fig. 17. The discrete landscape for the cancer network with parameter $a = 0.5, b = 0.5, k = 1$. Each node represents a specific cell state characterized by the qualitative expression levels of the corresponding marker genes. The nodes of blue, red, and green represent, respectively, the cell states evolving into the normal, cancer, and apoptosis states. The apoptosis pathways are represented by the black paths. The green and magenta paths denote, respectively, the pathways from the normal state to the cancer state and from the cancer state to the normal state.[39]

4.5. The potential and flux landscape framework of evolution

Fisher's fundamental theorem of natural selection[185] and Wright's fitness landscape[184] are widely used to interpret the evolutionary adaptation as the mean fitness maximization.[75–77] However, the evolutionary dynamics of non-equilibrium biological systems is complex and may no longer follow maximization of the mean fitness. The co-evolving systems may enter an endless cycle rather than reaching the mean fitness maximum. We studied the general evolutionary dynamics with the potential and flux landscape theory. We found that the conventional Fisher's fundamental theorem of natural selection and Wright's fitness landscape are based on equilibrium systems. Hence, they cannot explain many phenomena related to the nonequilibrium nature of systems. A typical example is Van Valen's Red Queen hypothesis, which indicates that an ecosystem can enter endless evolution processes even if the physical environment remains unchanged.[23]

4.5.1. Adaptive landscape and generalized fundamental theorem of natural selection

The mathematical framework of evolutionary theory is based on the description of the change in allele frequencies x_i. If the fitness w_{ij} is not dependent on the allele frequencies, the mean fitness is a Lyapunov function as indicated in Fisher's fundamental theorem of natural selection: $\mathrm{d}\bar{w}/\mathrm{d}t = V_{\mathrm{A}}(w)/\bar{w} \geq 0$ (where the additive genetic variance $V_{\mathrm{A}}(w) = 2\sum_{i=1}^{n} x_i(\sum_{j=1}^{n} x_j w_{ij} - \bar{w})^2 \geq 0$). In this sense, the evolutionary adaptation can be seen as the mean fitness maximization. This picture is Wright's fitness landscape, which has a gradient nature corresponding to equilibrium systems. In this frequency-independent selection case, the driving force of natural selection is determined by the gradient of the mean fitness $\bar{w} = \sum_{i,j=1}^{n} x_i x_j w_{ij}$ in the form of $\boldsymbol{F}^{\mathrm{S}} = (1/2)\boldsymbol{G}\cdot\boldsymbol{\nabla}\ln\bar{w}$, where the genetic drift matrix $G_{ij} = x_i(\delta_{ij} - x_j)$. The intrinsic flux velocity $\boldsymbol{V}_0$ vanishes in this case, indicating the system is in equilibrium for frequency-independent selections.

For general evolutionary dynamics, the system enters the nonequilibrium regime and the mean fitness does not always increase. The conventional evolutionary theory based on mean fitness then breaks down.[23] Some researchers believe that no general potential function exists whose optima are searched for in the evolution dynamics.[186] One may ask if there is a general potential function for evolution and if so what it is. The solution can be obtained from the potential–flux landscape theory. As a Lyapunov function, the intrinsic potential ϕ_0 can be used to redefine the adaptation for general evolutionary dynamics.[23]

By including selection and genetic drift, the nonequilibrium intrinsic potential ϕ_0 satisfies the Hamilton–Jacobi equation as given in Eq. (11)

$$\boldsymbol{F}^{\mathrm{S}}\cdot\boldsymbol{\nabla}\phi_0 + \boldsymbol{\nabla}\phi_0\cdot\boldsymbol{G}\cdot\boldsymbol{\nabla}\phi_0 = 0. \tag{23}$$

The adaptive landscape ϕ_0 for the group-help model[23] in the nonequilibrium regime is shown in Fig. 18. We can see the

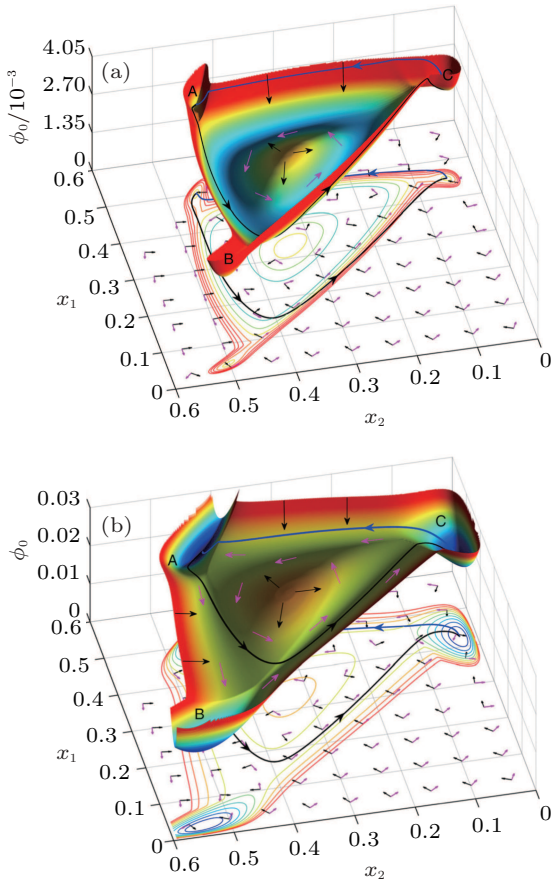


Fig. 18. Landscapes of different phases (with parameters (a) $\beta = 0.11$, (b) $\beta = 0.17$). The arrows are schematic, showing the two components of the driving force: the gradient of potential $-\boldsymbol{\nabla}\phi_0$ (black arrows) and the flux $\boldsymbol{V}_0$ (purple arrows).[23]

additional curl flux emerges and affects the dynamics of evolution. We redefined the adaptive rate by $\mathrm{d}\phi_0/\mathrm{d}t$ and obtained

$$\mathrm{d}\phi_0/\mathrm{d}t = -(1/2)V_{\mathrm{A}}(w)/\bar{w}^2 + \boldsymbol{V}_0 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{V}_0 \leq 0. \quad (24)$$

This is our picture about the evolutionary adaptation in which the adaptive landscape is measured by the intrinsic potential ϕ_0 and the adaptive rates measured by $\mathrm{d}\phi_0/\mathrm{d}t$.[23]

4.5.2. Red Queen hypothesis

An evolutionary system reaches its optima when $\mathrm{d}\phi_0/\mathrm{d}t = 0$. In the frequency-independent selection case, the system is in detailed balance with $\boldsymbol{V}_0 = 0$ and thus the genetic variance $V_{\mathrm{A}}(w) = 0$. In the general nonequilibrium regime, $\boldsymbol{V}_0 \neq 0$ leads to[23]

$$V_{\mathrm{A}}(w) = 2(\bar{w})^2[\boldsymbol{V}_0 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{V}_0] \geq 0, \quad (25)$$

which implies that natural selection has effects even if the system has reached the optima.[23]

What is the cause of the non-zero intrinsic flux velocity? If there is no biotic interaction, the natural selection comes from the external physical environment, as frequency independent selection would be a gradient force given by Wright's fitness landscape. Biotic interactions can sustain a genetic variation in fitness and thus give rise to endless evolution. The effect caused by biotic interactions was indicated by Van Valen in the Red Queen hypothesis.[187] According to this hypothesis, endless evolution can be sustained by biotic interactions even if the physical environment remains unchanged, that is, biotic interaction can lead to endless evolution.

In the potential–flux landscape framework, we can clearly see the origin of the Red Queen phenomena. The endless evolution comes from the non-zero flux of biotic interactions even when the evolution potential has reached its optima. Therefore, Van Valen's Red Queen hypothesis is the biological expression of the intrinsic flux: the system lies in the optimum where the gradient of potential is zero, while the intrinsic flux drives the endless evolution. As Van Valen himself remarked, this cannot be explained by the traditional evolutionary theory, since the flux is the missing part in the traditional evolutionary theory.[23]

4.6. The potential and flux landscape of ecology

Ecological systems interact with each other through cooperation and competition.[188] Exploring the stability of ecological systems is essential for understanding the mechanisms governing the dynamics of species and populations.[188] Ecological systems involve numerous interactions among the species which are generally non-linear complex dynamics. Much effort has been devoted to studying the stability of ecological systems.[189–192] Most of them focused on the analysis of local linear stability. It is thus important to study the global stability of ecological systems.

Previously, researchers have concentrated on deterministic models. In reality, both extrinsic and intrinsic fluctuations may exist in ecological systems, which can affect the global behaviors of ecological systems. The intrinsic fluctuations generally arise from the finite size of the ecological systems, while the extrinsic fluctuations typically come from the environment. The analysis of global stability is important for investigating the robustness of ecological systems under the influence of extrinsic and intrinsic fluctuations.[191,192]

Nonlinear ordinary differential equations can be used as a mathematical model of the deterministic dynamics of ecological systems with the interactions of predation, competition, and mutualism.[189,190] There are two basic types of relationships between species in ecological systems: a positive beneficial interaction (activation) and a negative compromising interaction (repression) as shown in Fig. 19. The predation model shows the repression and activation relationship $(+/-)$ in Fig. 19(a). It indicates that species S_{A} is compromised while species S_{B} benefits. Competition shows a relationship of mutual repression $(-/-)$ in Fig. 19(b). It indicates that both species S_{A} and S_{B} are compromised by interactions with each other. Mutualism shows a relationship of mutual activation (+/+) in Fig. 19(c). It indicates that species S_{A} and S_{B} benefit from interaction with each other.[188,191,192]

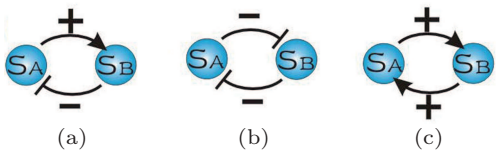


Fig. 19. (a) Relationship of predation. (b) Relationship of competition. (c) Relationship of mutualism.[22]

The landscape and flux theory has been applied to investigate the ecological systems with predation, competition, and mutualism for two species. We also added some restrictions and modifications on these ecological models so that they may be more realistic.[22,188,193] We acquired the steady-state probability distribution P_{ss} of ecological models with stochastic fluctuations as in real ecological systems. Then the potential landscape is obtained from the relation: $U = -\ln P_{\mathrm{ss}}$. It can be applied to investigate the global robustness and stability of ecological systems.

In the past studies, it was essential to address the global stability of the system using Lyapunov functions. In 1931, Volterra proposed an analytical Lyapunov function for a specific model of predation.[190] Then the analytical Lyapunov function[191,192] was explored in many ecological systems. For some simple ecological models, Lyapunov functions were obtained. However, there were no general methods to find the

Lyapunov function of more realistic ecological systems. A more systematic approach has emerged to fill the gap of identifying the Lyapunov function of general ecological systems.[22] Hence, we can now use the Lyapunov function to investigate the global robustness of a wider range of ecological systems.[22]

Figure 20 shows the population potential landscape U and intrinsic potential landscape ϕ_0 for the models of predation, competition, and mutualism. On the top row, black arrows represent the negative gradient of the population potential landscape $-\nabla U$. Purple arrows indicate the steady-state flux velocity $\boldsymbol{V}_{\rm ss}$. On the bottom row, the negative gradient of the intrinsic potential landscape $-\nabla\phi_0$ is represented by black arrows. The intrinsic flux velocity $\boldsymbol{V}_0$ is denoted by purple arrows. The bottom of each sub-figure shows the contour map of the landscape and the projection of the forces represented by arrows. In Figs. 20(a)–20(c), the purple flux velocity is almost perpendicular to the black gradient of potential U. In Figs. 20(d)–20(f), the purple intrinsic flux velocity is exactly perpendicular to the black gradient of ϕ_0.[22]

In the predation model, the potential landscape U with a Mexican hat shape for small fluctuations (characterized by a small diffusion coefficient) is shown in Fig. 20(a). It is an inhomogeneous oscillation. The population potential landscape U is distributed inhomogeneously along the oscillation ring. In contrast, the intrinsic potential landscape ϕ_0 (with the zero fluctuation limit taken) is distributed homogeneously along the oscillation ring shown in Fig. 20(d). The value of ϕ_0 is constant along the limit cycle.

The population potential landscape U and intrinsic potential landscape ϕ_0 for the competition model with the parameters evaluated at $a_1 = a_2 = 0.1$, $L_1 = L_2 = 0.3$, $\alpha = 1.0$, $D = 0.01$ are shown in Figs. 20(b) and 20(e). It is clear that the population potential landscape (Fig. 20(b)) and intrinsic potential landscape (Fig. 20(e)) have four locally stable attractors.[22]

The population potential landscape U and intrinsic potential landscape ϕ_0 for the mutualism model are shown in Figs. 20(c) and 20(f). The parameters are $D = 0.01$, $a_1 = a_2 = 0.1$, $\alpha = 1.0$, $L_1 = L_2 = 0.5$. The population potential landscape in Fig. 20(c) and intrinsic potential landscape in Fig. 20(f) have four attractors around locally stable states. We can see the gradient of the potential landscape attracts the ecological system to the four attractors, while the curl flux curls around in these four attractor basins.[22]

In the above, we studied three crucial ecological modes: predation, competition, and mutualism. We can see the ecological system may have multiple basins of attraction and also a limit cycle which has a landscape of the Mexican hat shape. We demonstrated the two key components determining the nonequilibrium dynamics of these ecological systems: the non-zero curl flux and the negative gradient of the landscape.[22]

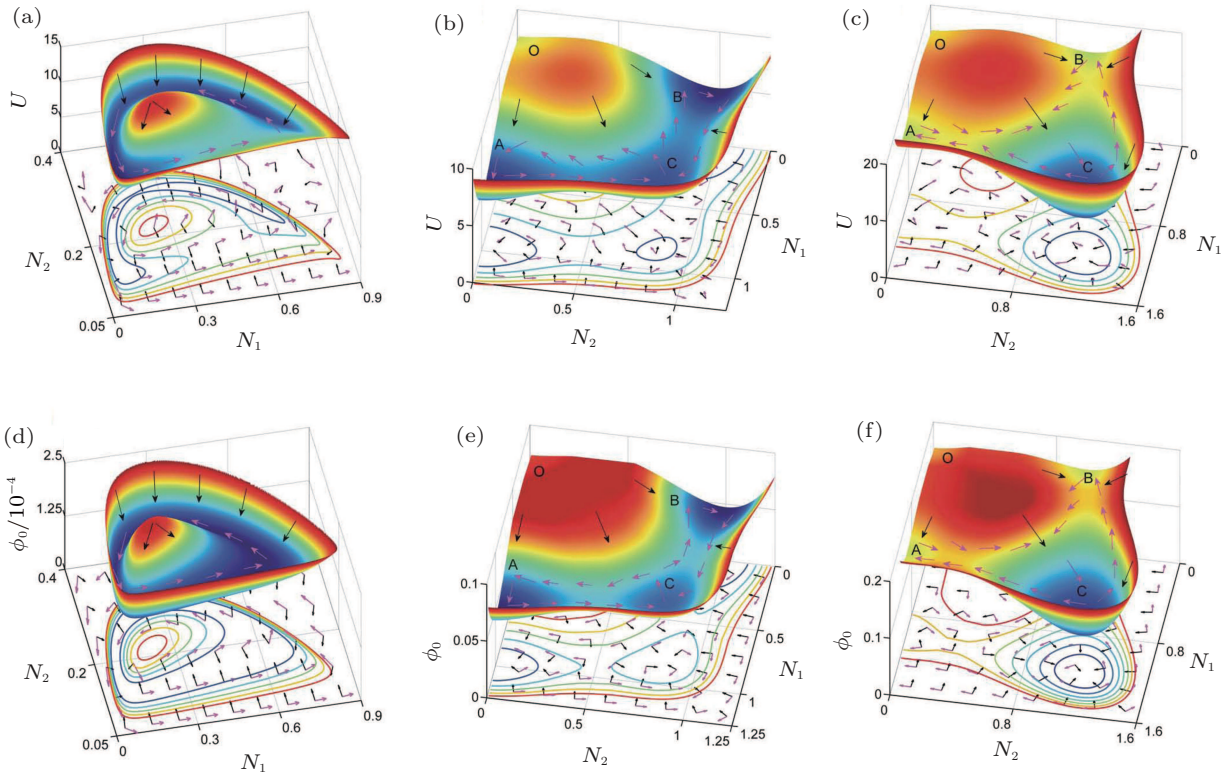


Fig. 20. (a)–(c) The population potential landscape U ((a) model of predation; (b) model of competition; (c) model of mutualism). (d)–(f) The intrinsic potential landscape ϕ_0 ((d) model of predation; (e) model of competition; (f) model of mutualism). On the top row, the flux velocity and the negative gradient of the population potential are denoted, respectively, by the purple arrows and black arrows. On the bottom row, the intrinsic flux velocity and the negative gradient of the intrinsic potential landscape are denoted, respectively, by the purple arrows and black arrows.[22]

We have also applied our theory to quantify the robustness of ecological systems through changing parameters that control the underlying landscape topography. We can investigate a variety of parameters to search for systems with higher stability quantified by the barrier heights of the underlying landscape topography. Thus the stability and decision making strategy for ecological systems can be achieved by modulating the control parameters. The global sensitivity analysis can be applied to investigate the effects of different values of parameters on the stability and robustness of ecological systems. We can also identify the species and interactions that are important for determining the global robustness of the ecological system by studying the topography of the potential landscape. The method we proposed for investigating the global robustness and stability can be used to explore more realistic ecological systems.[22]

4.7. Nonequilibrium landscape and flux theory of neural networks

Understanding the functions of the human brain is always a grand goal of neuroscience. As a complex dynamical system, the brain is responsible for various vital functions.[194–199] Although individual neurons have all sorts of behaviors, complicated physiological and cognitive functions in the brain are implemented by the neural circuits consisting of a large number of neurons rather than individual neurons. The cognitive functions such as decision-making, sensory, learning, and memory are related to special complex patterns of activities generated by neural circuits.[200–202] Despite the significant efforts in both theoretical and experimental approaches, global and quantitative understanding of the nature of large neural networks is still challenging.

Many models have been suggested for exploring the behaviors of neural circuits and corresponding biological functions.[195,197–199] The behavior of a single neuron can be quantitatively described by the Hodgkin–Huxley model based on its biological features.[197] Hopfield developed a model[198,199] to uncover the global characteristics of large neural circuits. Assuming the neural circuits are symmetrically connected, he constructed an energy landscape that decreases with time.[24,199] The dynamics of the symmetric neural circuit is dominated by the gradient force of the energy function. Then the associative memory function of the symmetric neural networks can be quantitatively described as shown in Fig. 21. Different memories (complete information) are stored in distinct basins of attraction (local minimum of energy) in the underlying energy landscape. After being given a cue, the system will always go downhill to the nearest attractor from the initial state with incomplete initial information. The associative memory is retrieved.[24]

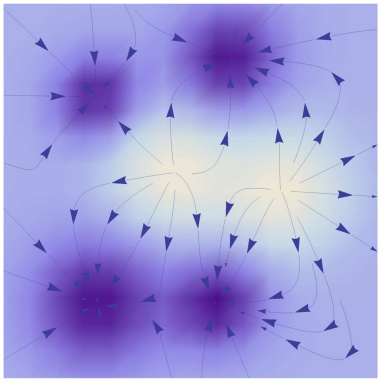


Fig. 21. The schematic drawing of the energy landscape of a symmetric neural network.[24]

The Hopfield model is able to show a good associative memory property. However, this model fails to apply in general neural networks since the connections among neurons are mostly asymmetric rather than symmetric in realistic neural circuits. Besides, there is evidence showing that oscillatory behaviors with different rhythms in neural circuits play crucial roles in cognitive processes.[201,202] Yet in Hopfield's symmetric neural networks, oscillations can never exist because the driving force here is a pure gradient of the potential. Fortunately, unlike the energy defined by Hopfield, the intrinsic potential ϕ_0 we introduced is always a Lyapunov function regardless of the symmetry or asymmetry of the circuit.[24] We have shown that the dynamics of neural networks is driven by both the gradient force of the potential and the curl flux force.[9,10,24] The flux can provide the main driving force for oscillations.[24]

We applied our landscape and flux framework to a network consisting of twenty model neurons.[24] Then we quantified the intrinsic potential ϕ_0, which is a Lyapunov function of the system, through taking the zero fluctuation limit of the potential landscape U. For the sake of visualizing the multi-dimensional results in the state space, we chose two state variables from the twenty model neurons and integrated the other eighteen variables to map out the landscape.[24] For the symmetrically connected neural circuit, the intrinsic potential landscape ϕ_0 is shown in Fig. 22(a). Similar to the Hopfield symmetric network model, we can see there are several basins of attraction. After a cue is given, the potential ϕ_0 of the system will decrease with time until a local minimum is reached eventually. However, in more general and realistic neural networks, the neurons are always asymmetrically connected. The difference between Hopfield's computational energy and the intrinsic potential ϕ_0 is that ϕ_0 is a Lyapunov function for both symmetric and asymmetric networks. In Fig. 22(b), we show that oscillations can occur for asymmetric networks. The landscape has a Mexican-hat topography. After being attracted by the Mexican-hat shape landscape down to the oscillation ring (white arrows represent the driving force from the negative

gradient of the potential landscape), the system is mainly dominated by the curl flux shown by green arrows in Fig. 22(b).

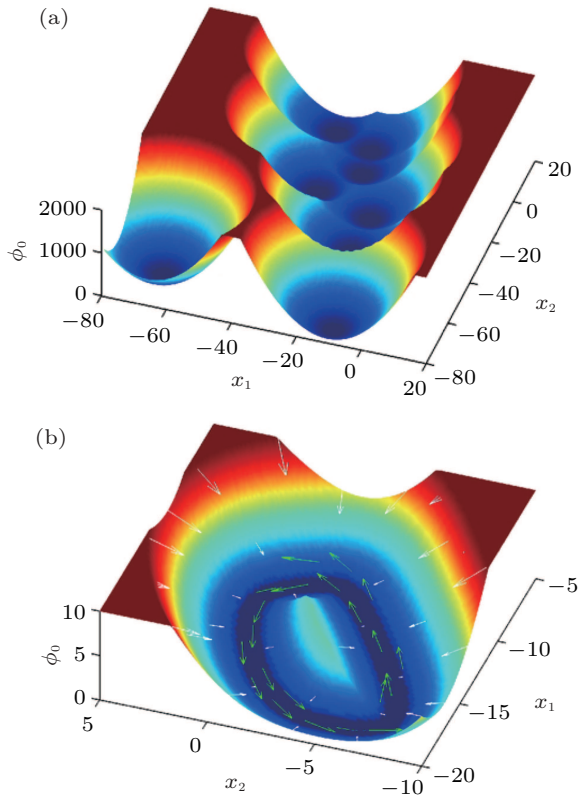


Fig. 22. (a) The intrinsic potential landscape ϕ_0 of the symmetric neural circuit. (b) The intrinsic potential landscape ϕ_0 and the corresponding driving force: white arrows represent the negative gradient force from the potential landscape and green arrows represent the curl flux force.[24]

As discussed, in Hopfield's symmetric neural network, the dynamics is only driven by the gradient force and there is no flux. Such a symmetric network corresponds to the equilibrium case of neural systems. In reality, the neural networks are open systems which constantly exchange materials and energy with the environment. We found that the nonzero flux in the asymmetric neural circuits are closely associated with the asymmetric part of the driving force.[24] The flux which breaks the detailed balance of the dynamical system is the signature indicating that the neural network is in a nonequilibrium state. Our results suggest that although the number of point attractors storing memories decreases when the network becomes more asymmetric, continuous attractors, for example, limit cycles, may occur, which are able to store memory in a sequential form. Different memories can be associated by the flux. The flux may provide a quantitative measure to explore the competition between memory associativity and storage capacity. We also found that the flux plays the key role in controlling repetitive physiological activities and biological rhythms because of its close association with the frequency of oscillations. To test this, we applied our nonequilibrium landscape and flux theory to a neural network model which can describe the dynamics of REM sleep cycle.[203,204] Based on the global quantification of the landscape, we made a global sensitivity analysis to investigate the effects of key factors on the stability and function of the system.[24] Our theoretical results agree well with experimental recordings.

5. Conclusion

This review has presented a theoretical framework of landscape and flux theory applied to a variety of biological systems. The energy landscape has become a powerful tool to understand biomolecular dynamics. However, it is often difficult to directly measure the energy landscape topography in reality. A novel way to explore the energy landscape is presented by the quantified energy landscape approach through thermodynamic and kinetic measurements. With the quantified energy landscapes, we are able to depict the biomolecular dynamics globally and physically. Based on the energy landscape theory, SBM has made a significant progress on simulating protein folding and binding processes. Even for IDPs, the energy landscapes of which are expected to be weakly funneled or anti-funneled, SBM can make excellent predictions consistent with experiments. With some adjustments, SBM can also be applied to investigate the large-scale conformational dynamics in an enzyme during its realization of function. Accumulating evidence stresses the crucial role of the energy landscape in various biomolecular dynamics, strengthening the importance of the energy landscape theory in modern molecular biology. A variety of simulation approaches based on the energy landscape theory require development and the applications are expected to have wide use in the future.

Biomolecular recognition is fundamental to the biological activities mediated by forming complexes between ligands and receptors. It is thus critical to study biomolecular recognition in order to understand the nature of these biological activities. Besides, such study can also find its direct use in drug discovery and design. Generally, both affinity and specificity are required for biomolecular recognition to realize its function, since affinity determines the stability of the complex and specificity controls the discrimination against competitive binding partners. The specificities of ligand binding can be calculated quantitatively using the energy landscape theory. Therefore, the energy landscape theory offers a practical way to predict and optimize interactions of ligand binding by considering the positive design (affinity) and negative design (specificity). In addition, with the thermodynamic and kinetic specificities quantified, multi-dimensional drug screening and target identification are promising means to avoid the side effects of drugs and seek specific targets.

The landscape and flux theoretical framework quantifies the global nature of the stability and dynamics of nonequilibrium biological systems, facilitating the discovery of physi-

cal mechanisms and basic principles behind complex dynamical behaviors of biological processes. The central idea of the framework consists of two essential components: (i) the potential landscape, obtained from the steady-state probability distribution, which gives a global characterization of the system's stability; (ii) the curl flux, quantified by the steady-state probability flux, which characterizes detailed balance breaking in nonequilibrium biological systems. Both the potential landscape and the curl flux contribute to the nonequilibrium dynamics of biological systems. The potential landscape drags the system downward to the valleys, while the curl flux drives the circulating process. For some, this may be reminiscent of the Chinese Yin–Yang philosophy. The potential landscape (Yin) and the curl flux (Yang), like a complementary pair promoting and constraining each other, are both needed to give a complete characterization of the global stability and dynamics of nonequilibrium biological systems. We have demonstrated the applications of the landscape and flux theory in multiple biological systems across a wide range, including cell cycle, stem cell, cancer, evolution, ecosystems, and neural networks.

It is challenging to understand the complex behaviors of nonequilibrium biological systems. In this review, we have presented our landscape and flux theoretical framework for quantitative global study of biological systems. We have focused on the theoretical aspects of the subjects here. It is worth noting that the theoretical explorations may be limited by our understanding of the topics and more experiments need to be carried out. Only through the combination of theoretical investigations and experimental studies can the field advance. We hope our theoretical framework can further contribute to the understanding of complex nonequilibrium biological processes through extending its applications in both depth and width and by making more contact with experiments.

References

[1] Bryngelson J D and Wolynes P G 1987 *Proc. Natl. Acad. Sci. USA* **84** 7524
[2] Bryngelson J D and Wolynes P G 1989 *J. Phys. Chem.* **93** 6902
[3] Bryngelson J D and Wolynes P G 1990 *Biopolymers* **30** 177
[4] Onuchic J N, Wolynes P G, Lutheyschulten Z and Socci N D 1995 *Proc. Natl. Acad. Sci. USA* **92** 3626
[5] Onuchic J N, LutheySchulten Z and Wolynes P G 1997 *Annu. Rev. Phys. Chem.* **48** 545
[6] Dill K A and Chan H S 1997 *Nat. Struct. Biol.* **4** 10
[7] Dobson C M, Šali A and Karplus M 1998 *Angew. Chem. Int. Ed.* **37** 868
[8] Wang J 2015 *Adv. Phys.* **64** 1
[9] Wang J, Xu L and Wang E K 2008 *Proc. Natl. Acad. Sci. USA* **105** 12271
[10] Wang J, Li C H and Wang E K 2010 *Proc. Natl. Acad. Sci. USA* **107** 8195
[11] Li C H and Wang J 2014 *Proc. Natl. Acad. Sci. USA* **111** 14130
[12] Wang J, Xu L, Wang E K and Huang S 2010 *Biophys. J.* **99** 29
[13] Wang J, Zhang K and Wang E K 2010 *J. Chem. Phys.* **133** 125103
[14] Wang J, Zhang K, Xu L and Wang E K 2011 *Proc. Natl. Acad. Sci. USA* **108** 8257
[15] Feng H D and Wang J 2011 *J. Chem. Phys.* **135** 234511
[16] Feng H D and Wang J 2012 *Sci. Rep.* **2** 550
[17] Zhang K, Sasai M and Wang J 2013 *Proc. Natl. Acad. Sci. USA* **110** 14930
[18] Li C H and Wang J 2013 *PLoS Comput. Biol.* **9** e1003165
[19] Xu L, Zhang K and Wang J 2014 *PLoS One* **9** e105216
[20] Gerard C and Goldbeter A 2009 *Proc. Natl. Acad. Sci. USA* **106** 21643
[21] Gerard C and Goldbeter A 2012 *PLoS Comput. Biol.* **8** e1002516
[22] Xu L, Zhang F, Zhang K, Wang E K and Wang J 2014 *PLoS ONE* **9** e86746
[23] Zhang F, Xu L, Zhang K, Wang E K and Wang J 2012 *J. Chem. Phys.* **137** 065102
[24] Yan H, Zhao L, Hu L, Wang X, Wang E K and Wang J 2013 *Proc. Natl. Acad. Sci. USA* **110** E4185
[25] Koretke K K, Luthey-Schulten Z A and Wolynes P G 1998 *Proc. Natl. Acad. Sci. USA* **95** 2932
[26] Wang J and Verkhivker G M 2003 *Phys. Rev. Lett.* **90** 188101
[27] Levy Y, Wolynes P G and Onuchic J N 2004 *Proc. Natl. Acad. Sci. USA* **101** 511
[28] Wang J, Zheng X, Yang Y, Drueckhammer D, Yang W, Verkhivker G and Wang E 2007 *Phys. Rev. Lett.* **99** 198101
[29] Chu X, Gan L, Wang E and Wang J 2013 *Proc. Natl. Acad. Sci. USA* **110** E2342
[30] Goldstein R A, Luthey-Schulten Z A and Wolynes P G 1992 *Proc. Natl. Acad. Sci. USA* **89** 4918
[31] Goldstein R A, Luthey-Schulten Z A and Wolynes P G 1992 *Proc. Natl. Acad. Sci. USA* **89** 9029
[32] Hardin C, Eastwood M P, Luthey-Schulten Z and Wolynes P G 2000 *Proc. Natl. Acad. Sci. USA* **97** 14235
[33] Takada S 2001 *Proteins* **42** 85
[34] Fain B and Levitt M 2003 *Proc. Natl. Acad. Sci. USA* **100** 10700
[35] Liu Z, Dominy B N and Shakhnovich E I 2004 *J. Am. Chem. Soc.* **126** 8515
[36] Yan Z and Wang J 2012 *Sci. Rep.* **2** 309
[37] Yan Z, Guo L, Hu L and Wang J 2013 *Bioinformatics* **29** 1127
[38] Yan Z and Wang J 2013 *PloS ONE* **8** e74443
[39] Li C H and Wang J 2014 *J. R. Soc. Interf.* **11** 20140774
[40] Van Kampen N G 2007 *Stochastic Processes in Physics and Chemistry* (Amsterdam: Elsevier)
[41] Hu G 1995 *Stochastic Force and Nonlinear Systems* (Shanghai: Shanghai Science Education)
[42] Gardiner C W 2004 *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences* (Berlin: Springer-Verlag)
[43] Risken H 1996 *The Fokker–Planck Equation: Methods of Solution and Applications* (Berlin: Springer-Verlag)
[44] Levinthal C 1969 *Proceedings in Mossbauer Spectroscopy in Biological Systems. Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois.* p. 22
[45] Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 *Proteins* **21** 167
[46] Leopold P E, Montal M and Onuchic J N 1992 *Proc. Natl. Acad. Sci. USA* **89** 8721
[47] Wang J, Onuchic J and Wolynes P 1996 *Phys. Rev. Lett.* **76** 4861
[48] Frauenfelder H, Sligar S G and Wolynes P G 1991 *Science* **254** 1598
[49] Tsai C J, Kumar S, Ma B Y and Nussinov R 1999 *Protein Sci.* **8** 1181
[50] Wang J, Xu L and Wang E 2007 *Biophys. J.* **92** L109
[51] Oliveberg M and Wolynes P G 2005 *Q. Rev. Biophys.* **38** 245
[52] Wolynes P G 2005 *Philos. T. Roy. Soc. A* **363** 453
[53] Wolynes P G 2005 *Q. Rev. Biophys.* **38** 405
[54] Wang J, Oliveira R J, Chu X, Whitford P C, Chahine J, Han W, Wang E, Onuchic J N and Leite V B P 2012 *Proc. Natl. Acad. Sci. USA* **109** 15763
[55] Chu X and Wang J 2014 *PLoS Comput. Biol.* **10** e1003782
[56] Wang Y, Chu X, Suo Z, Wang E and Wang J 2012 *J. Am. Chem. Soc.* **134** 13755
[57] Wang J, Lu Q and Lu H P 2006 *PLoS Comput. Biol.* **2** e78
[58] Lu Q, Lu P H and Wang J 2007 *Phys. Rev. Lett.* **98** 128105
[59] Wang J, Wang Y, Chu X, Hagen S J, Han W and Wang E 2011 *PLoS Comput. Biol.* **7** e1001118
[60] Chu X, Wang Y, Gan L, Bai Y, Han W, Wang E and Wang J 2012 *PLoS Comput. Biol.* **8** e1002608
[61] Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E and Wang J 2013 *Proc. Natl. Acad. Sci. USA* **110** E3743
[62] Wang Y, Longhi S, Roche P and Wang J 2014 *Proc. Natl. Acad. Sci. USA* **111** E1559

[63] Lu Q and Wang J 2008 *J. Am. Chem. Soc.* **130** 4772
[64] Lu Q and Wang J 2009 *J. Phys. Chem. B* **113** 1517
[65] Lai Z Z, Lu Q and Wang J 2011 *J. Phys. Chem. B* **115** 4147
[66] Wang Y, Tang C, Wang E and Wang J 2012 *PLoS Comput. Biol.* **8** e1002471
[67] Wang Y, Gan L, Wang E and Wang J 2013 *J. Chem. Theory Comput.* **9** 84
[68] Chu X, Liu F, Maxwell B A, Wang Y, Suo Z, Wang H, Han W and Wang J 2014 *PLoS Comput. Biol.* **10** e1003804
[69] Yan Z Q, Zheng X L, Wang E K and Wang J 2013 *Chem. Sci.* **4** 2387
[70] Yang Y L, Li G H, Zhao D Y, Yu H Y, Zheng X L, Peng X D, Zhang X E, Fu T, Hu X Q, Niu M S, Ji X F, Zou L B and Wang J 2015 *Chem. Sci.* **6** 2812
[71] Zheng X and Wang J 2015 *PLoS Comput. Biol.* **11** e1004212
[72] Graham R 1989 *Macroscopic Potentials, Bifurcations and Noise in Dissipative Systems in Noise in Nonlinear Dynamical Systems* (Cambridge: Cambridege Unviersity Press)
[73] Sasai M and Wolynes P G 2003 *Proc. Natl. Acad. Sci. USA* **100** 2374
[74] Qian H and Reluga T C 2005 *Phys. Rev. Lett.* **94** 028101
[75] Ao P 2005 *Phys. Life Rev.* **2** 117
[76] Ao P 2008 *Commun. Theor. Phys.* **49** 1073
[77] Ao P, Galas D L H and Zhu X 2008 *Med Hypotheses* **70** 678
[78] Qian H 2014 *Phys. Lett. A* **378** 609
[79] Feng H D, Han B and Wang J 2011 *J. Phys. Chem. B* **115** 1254
[80] Li C H, Wang E K and Wang J 2012 *J. Chem. Phys.* **136** 194108
[81] Wu W and Wang J 2013 *J. Phys. Chem. B* **117** 12908
[82] Wu W and Wang J 2013 *J. Chem. Phys.* **139** 121920
[83] Wu W and Wang J 2014 *J. Chem. Phys.* **141** 105104
[84] Lepzelter D and Wang J 2008 *Phys.Rev. E* **77** 041917
[85] Kumar S, Bouzida D, Swendsen R H, Kollman P A and Rosenberg J M 1992 *J. Comput. Chem.* **13** 1011
[86] Plotkin S S, Wang J and Wolynes P G 1997 *J. Chem. Phys.* **106** 2932
[87] Shakhnovich E I 1994 *Phys. Rev. Lett.* **72** 3907
[88] Iben I E T, Braunstein D, Doster W, Frauenfelder H, Hong M K, Johnson J B, Luck S, Ormos P, Schulte A, Steinbach P J, Xie A H and Young R D 1989 *Phys. Rev. Lett.* **62** 1916
[89] Wright P E and Dyson H J 1999 *J. Mol. Biol.* **293** 321
[90] Dunker A K, Lawson J D, Brown C J, Williams R M, Romero P, Oh J S, Oldfield C J, Campen A M, Ratliff C R, Hipps K W, Ausio J, Nissen M S, Reeves R, Kang C H, Kissinger C R, Bailey R W, Griswold M D, Chiu M, Garner E C and Obradovic Z 2001 *J. Mol. Graph. Model.* **19** 26
[91] Tompa P 2002 *Trends. Biochem. Sci.* **27** 527
[92] Dyson H J and Wright P E 2002 *Curr. Opin. Struct. Biol.* **12** 54
[93] Wright P E and Dyson H J 2009 *Curr. Opin. Struct. Biol.* **19** 31
[94] Dunker A K, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C and Villafranca J E 1998 *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 473
[95] Shoemaker B A, Portman J J and Wolynes P G 2000 *Proc. Natl. Acad. Sci. USA* **97** 8868
[96] Dunker A K, Silman I, Uversky V N and Sussman J L 2008 *Curr. Opin. Struct. Biol.* **18** 756
[97] Smock R G and Gierasch L M 2009 *Science* **324** 198
[98] Tantos A, Han K H and Tompa P 2012 *Mol. Cell. Endocrinol* **348** 457
[99] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937
[100] Levy Y, Cho S S, Onuchic J N and Wolynes P G 2005 *J. Mol. Biol.* **346** 1121
[101] Ganguly D, Zhang W and Chen J 2013 *PLoS Comput. Biol.* **9** e1003363
[102] Turjanski A G, Gutkind J S, Best R B and Hummer G 2008 *PLoS Comput. Biol.* **4** 2008 *PLoS Com-put. Biol.* **4** e1000060
[103] Ganguly D and Chen J 2011 *Proteins: Struc. Funct. Bioinform.* **79** 1251
[104] Huang Y and Liu Z 2009 *J. Mol. Biol.* **393** 1143
[105] Hansen D F, Zhou Z, Fen H Q, Jenkins L M M, Bai Y W and Kay L E 2009 *J. Mol. Biol.* **387** 1
[106] Pappu R V, Mao A H, Crick S L, Vitalis A and Chicoine C L 2010 *Proc. Natl. Acad. Sci. USA* **107** 8183
[107] Muller-Spath S, Soranno A, Hirschfeld V, Hofmann H, Ruegger S, Reymond L, Nettels D and Schuler B 2010 *Proc. Natl. Acad. Sci. USA* **107** 14609
[108] Haran G and England J L 2010 *Proc. Natl. Acad. Sci. USA* **107** 14519
[109] Shaw D E, Maragakis P, Lindorff-Larsen K, Piana S, Dror R O, Eastwood M P, Bank J A, Jumper J M, Salmon J K, Shan Y, *et al.* 2010 *Science* **330** 341
[110] Lindorff-Larsen K, Piana S, Dror R O and Shaw D E 2011 *Science* **334** 517
[111] Narayanan R, Ganesh O K, Edison A S and Hagen S J 2008 *J. Am. Chem. Soc.* **130** 11477
[112] Longhi S, Receveur-Bréchot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S and Canard B 2003 *J. Biol. Chem.* **278** 18638
[113] Bourhis J M, Johansson K, Receveur-Bréchot V, Oldfield C J, Dunker K A, Canard B and Longhi S 2004 *Virus Res.* **99** 157
[114] Bourhis J M, Receveur-Bréchot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, Canard B, Finet S and Longhi S 2005 *Protein Sci.* **14** 1975
[115] Jensen M R, Communie G, Ribeiro E A, Martinez N, Desfosses A, Salmon L, Mollica L, Gabel F, Jamin M, Longhi S *et al.* 2011 *Proc. Natl. Acad. Sci. USA* **108** 9839
[116] Piana S, Lindorff-Larsen K and Shaw D E 2011 *Biophys. J.* **100** L47
[117] Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren W F and Mark A E 1999 *Angew. Chem. Int. Ed.* **38** 236
[118] Koshland D E 1958 *Proc. Natl. Acad. Sci. USA* **44** 98
[119] Ma B Y, Kumar S, Tsai C J and Nussinov R 1999 *Protein Eng.* **12** 713
[120] Boehr D D, Nussinov R and Wright P E 2009 *Nat. Chem. Biol.* **5** 789
[121] Okazaki K i, Koga N, Takada S, Onuchic J N and Wolynes P G 2006 *Proc. Natl. Acad. Sci. USA* **103** 11844
[122] Okazaki K i and Takada S 2008 *Proc. Natl. Acad. Sci. USA* **105** 11182
[123] Whitford P C, Miyashita O, Levy Y and Onuchic J N 2007 *J. Mol. Biol.* **366** 1661
[124] Berry M B, Meador B, Bilderback T, Liang P, Glaser M and Phillips G N 1994 *Proteins* **19** 183
[125] Berry M B, Bae E, Bilderback T R, Glaser M and Phillips G N 2006 *Proteins* **62** 555
[126] Wong J H, Fiala K A, Suo Z and Ling H 2008 *J. Mol. Biol.* **379** 317
[127] Ohmori H, Friedberg E, Fuchs R, Goodman M, Hanaoka F, Hinkle D, Kunkel T, Lawrence C, Livneh Z, Nohmi T, Prakash L, Prakash S, Todo T, Walker G, Wang Z and Woodgate R 2001 *Mol. Cell* **8** 7
[128] Janin J 1995 *Biochimie* **77** 497
[129] Havranek J J and Harbury P B 2003 *Nat. Struct. Mol. Biol.* **10** 45
[130] Kortemme T, Joachimiak L A, Bullock A N, Schuler A D, Stoddard B L and Baker D 2004 *Nat. Struct. Mol. Biol.* **11** 371
[131] Bolon D N, Grant R A, Baker T A and Sauer R T 2005 *Proc. Natl. Acad. Sci. USA* **102** 12724
[132] Liu Z, Zheng X, Yang X, Wang E and Wang J 2009 *Biophys. J.* **96** 3917
[133] Li D, Liu Z, Zhao W, Zheng X, Wang J and Wang E 2013 *Eur. J. Pharm. Sci.* **48** 658
[134] Liu Z, Li D, Zhao W, Zheng X, Wang J and Wang E 2012 *PLoS ONE* **7** e37841
[135] Zhao W, Li D, Liu Z, Zheng X, Wang J and Wang E 2013 *PLoS ONE* **8** e66362
[136] Zheng X, Liu Z, Li D, Wang E and Wang J 2013 *Curr. Pharm. Des.* **19** 2246
[137] Liu Z, Li D, Zheng X, Wang E and Wang J 2013 *Curr. Pharm. Des.* **19** 2259
[138] Zheng X, Gan L, Wang E and Wang J 2013 *The AAPS. Journal* **15** 228
[139] Vauquelin G and Van Liefde I 2006 *Trends Pharmacol. Sci.* **27** 355
[140] Tummino P J and Copeland R A 2008 *Biochemistry* **47** 5481
[141] Yan Z, Zheng X, Wang E and Wang J 2013 *Chem. Sci.* **4** 2387
[142] Socci N D, Onuchic J N and Wolynes P G 1996 *J. Chem. Phys.* **104** 5860
[143] Wang J 2006 *Chem. Phys. Lett.* **418** 544
[144] Shoichet B K 2004 *Nature* **432** 862
[145] Swinney D C 2004 *Nat. Rev. Drug Discov.* **3** 801
[146] Swinney D C 2006 *Lett. Drug Des. Discov.* **3** 569
[147] Copeland R A, Pompliano D L and Meek T D 2006 *Nat. Rev. Drug Discov.* **5** 730
[148] Lu H and Tonge P J 2010 *Curr. Opin. Chem. Biol.* **14** 467
[149] Xu L F, Shi H L, Feng H D and Wang J 2012 *J. Chem. Phys.* **136** 165102
[150] Xu L, Zhang F, Wang E K and Wang J 2013 *Nonlinearity* **26** 69
[151] V Blickle T Speck L H U S and Bechinger C 2006 *Phys. Rev. Lett.* **96** 070603
[152] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
[153] Seifert U and Speck T 2010 *Europhys. Lett.* **89** 10007

[154] Speck T and Seifert U 2005 *J. Phys. A* **38** L581
[155] Esposito M and Van den Broeck C 2010 *Phys. Rev. Lett.* **104** 090601
[156] Esposito M and Van den Broeck C 2010 *Phys. Rev. E* **82** 011143
[157] Esposito M and Van den Broeck C 2010 *Phys. Rev. E* **82** 011144
[158] Ge H and Qian H 2010 *Phys. Rev. E* **81** 051133
[159] Singhania R, Sramkoski R M, Jacobberger J W, and Tyson J J 2011 *PLoS Comput. Biol.* **7** e1001077
[160] Han B and Wang J 2007 *Biophys. J.* **92** 3755
[161] Waddington C H 1957 *The Strategy of the Genes* (London: Allen and Unwin)
[162] Zhang B and Wolynes P G 2014 *Proc. Natl. Acad. Sci. USA* **111** 10185
[163] Walczak A M, Sasai M and Wolynes P G 2005 *Biophys. J.* **88** 828
[164] Ferrell J E 2012 *Curr. Biol.* **22** R458
[165] Huang S, Guo Y P, May G and Enver T 2007 *Dev. Biol.* **305** 695
[166] Li C H and Wang J 2013 *J. R. Soc. Interf.* **10** 20130787
[167] Feng H D, Zhang K and Wang J 2014 *Chem. Sci.* **5** 3761
[168] Huang S, Ernberg I and Kauffman S 2009 *Semin Cell Dev. Biol.* **20** 869
[169] Kalmar T, Lim C, Hayward P, Munoz-Descalzo S and Nichols J 2009 *PLoS* **Biol.** 7
[170] Singh A M, Hamazaki T, Hankowski K E and Terada N 2007 *Stem Cells* **25** 2534
[171] Onsager L and Machlup S 1953 *Phys. Rev.* **91** 1505
[172] Hunt K L C and Ross J 1981 *J. Chem. Phys.* **75** 976
[173] Kramers H A 1940 *Physica* **7** 284
[174] Maier R S and Stein D L 1997 *SIAM J. Appl. Math.* **57** 752
[175] Hänggi P, Talkner P and Borkovec M 1990 *Rev. Mod. Phys.* **62** 251
[176] Weinberg R A 2013 *The Biology of Cancer* (New York: Taylor & Francis)
[177] Hanahan D and Weinberg R A 2011 *Cell* **144** 646
[178] Hanahan D and Weinberg R A 2000 *Cell* **100** 57
[179] Kauffman S 1971 *J. Theor. Biol.* **31** 429
[180] Lu M, Onuchic J N and Ben-Jacob E 2014 *Phys. Rev. Lett.* **113** 078102
[181] Lu M, Jolly M K, Gomoto R, Huang B, Onuchic J N and Ben-Jacob E 2013 *J. Phys. Chem. B* **117** 13164
[182] Lu M, Jolly M K, Levine H, Onuchic J N and Ben-Jacob E 2013 *Proc. Natl. Acad. Sci. USA* **110** 18144
[183] Jolly M K, Huang B, Lu M, Mani S A, Levine H and Ben-Jacob E 2014 *J. R. Soc. Interface* **11** 20140962
[184] Wright S 1932 *Proceedings of the Sixth International Congress on Genetics* **1** 356
[185] Fisher R A 1930 *The Genetical Theory of Natural Selection* (New York: Clarendon and Oxford)
[186] Rice S H 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations* (Sinauer Associates)
[187] Van Valen L 1973 *Evol Theory* **1** 1
[188] Murray J 1998 *Mathematical Biology* (New York/Berlin: Springer)
[189] Lotka A 1925 *Elements of Physical Biology* (Baltimore: Williams and Wilkins)
[190] Volterra V 1931 *Lecons sur la Theorie Mathematique de la Lutte pour la Vie* (Paris: Gauthier–Villars)
[191] Goh B 1976 *J. Math. Biol.* **3** 313
[192] Hastings A 1977 *J. Math. Biol.* **5** 399
[193] Bazykin A 1985 *Mathematical Biophysics of Interacting Populations* (Moscow: Nauka) (in Russian)
[194] Abbott L F and Regehr W G 2004 *Nature* **431** 796
[195] Abbott L F 2008 *Neuron* **60** 489
[196] Vogels T P, Rajan K and Abbott L F 2005 *Annu. Rev. Neurosci.* **28** 357
[197] Hodgkin A L and Huxley A F 1952 *J. Physiol.* **117** 500
[198] Hopfield J J 1982 *Proc.Natl. Acad. Sci. USA* **79** 2554
[199] Hopfield J J and Tank D W 1986 *Science* **233** 625
[200] Arnsten A F T, Wang M, Gamo N J, Yang Y, Jin L E, Wang X J, Laubach M, Mazer J A and Lee D 2011 *Nature* **476** 210
[201] Buzsaki G and Draguhn A 2004 *Science* **304** 1926
[202] Wang X J, Tegner J, Constantinidis C and Goldman-Rakic P S 2004 *Proc. Natl. Acad. Sci. USA* **101** 1368
[203] McCarley R W and Massaquoi S G 1986 *Am. J. Physiol-Reg. I* **251** R1011
[204] Massaquoi S G and McCarley R W 1992 *J. Sleep. Res.* **1** 138