The Role of Structural Dynamics in Protein Function and Evolvability

Muthahari, Yusran

DOI:
10.33612/diss.155494424

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

Introduction
Adopting from an old Greeks philosophy, we may agree that life is interconnected, and all creatures have to live their life by thriving their own Arete. While there are no direct translations for Arete, let us all assume that the Ancient Greek word resembles the notion of excellence in fulfilling the purpose of life. To achieve Arete, one has to act up according to its function by utilizing the available potential from the environment in an ideal way. This concept seems to be generic and applicable not only for an organism as a whole but we may associate the philosophy with the smaller microscopic system, i.e., the cell or even its biomolecules as an example.

Nature delegates the more specific functions to much smaller biomolecules (i.e., the DNA and mRNA polymer carrying genetic information, the tRNA is acting as the molecular adapter for protein translation) to gain the more complex function of the cell\textsuperscript{1,2}. Proteins, as another example, operating synergistically to preserve the dynamic equilibria of the cell\textsuperscript{3-5}.

For adapting to new environmental changes, the dynamic equilibria of the cells need to be adjusted, which can be facilitated by modulating the proteins function\textsuperscript{2,6,7}. However, it remains unclear how nature provides a solution to modulate protein function while maintaining cell fitness. As quoted from the neutral theory of molecular evolution, the mutation that mostly affects fitness is deleterious and removed from the population\textsuperscript{8-10}. As an alternative, nature may enable the mutation or recombination that does not significantly alter the original functions in which the protein has been designed. Such structural changes would result in an array of conformations (or dynamics) suitable for the cell to remain fit while emerging the new function.

In 2005, Dan Tawfik and co-workers advanced the breakthrough theory of protein evolvability, which describes protein evolution from the perspective of the structural dynamics and energetics\textsuperscript{11,12}. The proposed notion accommodated the neutral theory by explaining that protein may pose several promiscuous functions generated by different levels of dynamics, thus hypothesized to be the starting point of protein evolution for the emergence of a new function\textsuperscript{13-15}. The new dominant function arises from a repetitive round of synergistic neutral mutations and plastic adaptations that shift the equilibria towards one specific seemingly-promiscuous function. In their study, directed mutations on a fixed-length protein sequence are enough to evolve the protein in a short period, which is excellent to describe faithfully how an evolved protein can recognize a human-made substrate without losing its original function\textsuperscript{11,15}. However, it remains a mystery of how nature utilises the same repertoire to evolve new functionalities. To address these
specific questions, we need to understand the evolution and the mechanics of protein allosteric network.

Protein allostery defines the transmission of the effect from ligand/partners interaction to the protein functional sites\textsuperscript{16,17}. Such an effect is observable at both proximal and distal functional site that permits the regulation of protein function\textsuperscript{16}. Unlike the structure-based allostery model, the visionary perspective of Cooper and Dyden in 1984 revealed the correlation between the allostery and the conformational distribution, which often termed ‘dynamic allostery’\textsuperscript{18}. The more recent model (the ‘Ensemble Allosteric Model’ – EAM of Vincent Hilser), showed that the allosteric mechanism is pre-encoded in ‘\textit{the protein ensemble}’ well-described by the protein energetic landscape\textsuperscript{16,17}. The unique properties of EAM also allow us to explain: (i) the redistribution of the ensemble upon addition of the ligand and (ii) what is the structural basis of the allosteric effect\textsuperscript{17}.

Both the evolvability theory of Tawfik and the EAM theory of Hilser highlighted the importance of seeing protein as a ‘\textit{dynamic ensemble}’ rather than as a ‘\textit{static structure}’ as defined by the energetic landscape model\textsuperscript{19,20}. It is worth noting that the evolvability opens up the possibility to uncover the protein history and how they evolved, whereas the EAM provides the mechanical basis for how the protein work. Combining the two perspectives on the protein studies may decode the secret ingredients on how nature achieve the \textit{Arete} by efficiently designed a selected protein to perform specific functions that can still be further evolved. Such insight might benefit us not only to understand the properties of the extant proteins but also to rationalize the molecular basis of protein design as addressed by Rama Ranganathan\textsuperscript{21-23}.

In this thesis, we will focus on uncovering one side of the story of protein evolution/design by correlating the role of the structural dynamics of the non-fixed length peptides to their evolvability. The study of ‘the evolution of the allosteric network’ of such peptides will be the main subject of our future studies. Here we employed a rigorous comparative structural analysis followed by the biophysical experiments to indicate the evolution of the biomechanical properties of the non-fixed length peptides that lead to functional specialization. To better understand on how the ‘\textit{evolution}’ reshapes the protein function, we provided a brief introduction in this chapter concerning the basic principles of (i) the protein structures, (ii) the energetic landscape/funnel, (iii) the multi-tiers dynamics, and (iv) the protein evolvability.
The Structure of a Protein

Proteins are bio-polymers constituted of amino acids, possessing an ensemble of conformations that fulfil one function at least. Any polymer of an array of amino acids that does not fulfill the previously-mentioned criteria, is referred to as a peptide. Proteins may resemble several levels of structural complexity, depending on the interactions that clench the corresponding structure:

1. The primary structure of proteins is stabilized by the covalent interaction between atoms, forming the amide functional group of the backbone (Figure 1.1A). This polymeric structure begins with the first amine group of the first amino acid (N-termini) and ends at the carboxylate group of the last amino acid (C-termini). Following the previous criterion, one can determine the protein sequence, which is the linear arrangement of different amino acids that compose the polypeptide chain.

2. The secondary structure mainly sustained by the hydrogen bonds between the nitrogen-attached hydrogens and the oxygen atoms of the stacked backbone amide groups (Figure 1.1B). These interactions are responsible for the formation of the two widely known secondary structures: beta-strands and alpha-helices. By considering its geometry and orientation, one can determine the protein topology, which is the sequence-based arrangement of the secondary structures.

3. The ensemble of the relatively-weak interactions of the amino acid side chains establishes the protein tertiary structure (Figure 1.1C). Except for the disulfide bonds between cysteines, all of these interactions are non-covalent, including: the van der Waal, hydrophobic, phi-phi, and electrostatic interactions. The tertiary structure defines the spatial arrangement of each amino acid, which is crucial for dictating the protein dynamics and function.
Figure 1.1 | Cartoon Representation of the Multi-level Structures of a Protein. (A) Covalent interactions connect amino-acids to form a polymer, termed the Primary Structure of a Protein. The sequence begins at the amino (N-) termini and ends at carboxy (C-) termini. (B) The formation of the α-helix (red) and β-sheet (blue) structures facilitated by hydrogen bonds between the backbone amine and carboxyl functional groups of amino-acids. (C) The tertiary structure of Proliferating Cell Nuclear Antigen (PCNA) formed mainly by relatively-weak interactions of the side chain of amino acids. (D) Quaternary structure of PCNA (PDB: 1AXC) composed of three tertiary-structure subunits. Figure adapted from Wikimedia Commons website by Thomas Shafee, 2018, retrieved from https://commons.wikimedia.org/wiki/File:Protein_structure_(full).png. Licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0).
4. Complex proteins require more dynamical properties that are not provided by the tertiary level conformational changes. Thus, these proteins utilize their quaternary structure to achieve conformational diversity and develop a wide repertoire of functions. Such proteins require various synergistic interactions between their domains to form this structure (Figure 1.1D).

The ensemble of the sub-structural conformations defines protein dynamics. These conformations differ by at least one rotatable part between them, which also describe the protein spatial arrangement. These rotational changes are observable in a small scale (e.g. backbone amide bonds) or even a large scale (e.g. the assembly of the protein folds). Because of the high degree of freedom, a protein might adopt numerous conformers, although not all are functional. Out of all conformers, the native state is the most dominant and physiologically relevant one.

The Path of Protein Folding

In 1969 at Allerton House, Cyrus Levinthal gave his famous talks, which has been noted and titled as ‘How to Fold Graciously’ on the Mossbauer Spectroscopy in Biological System Proceedings25. He argued on how a protein folds into the native structure from the unfolded one. The argument, widely known as ‘The Levinthal Paradox,’ reasoned that there are so many ‘haystacks’ conformations that a protein has to sample before reaching the native ‘needle’ state. He also proposed that the protein folding process has to be kinetically ‘speeded’ and thermodynamically ‘guided’ to find the needle. Following his arguments, many studies relied on the existence of a specific protein folding ‘pathway.’ To decrypt the “secrets” of protein folding, many scientists put an effort to decrypt the on-pathway intermediates during the folding process. They assumed that if we could follow the kinetics of the intermediate states, we would find the answer on how nature can rapidly find the ‘needle’ out of ‘haystacks.’

In early 1971, Atsushi Ikai and Charles Tanford followed the (re)folding reactions kinetics of two different proteins, i.e., Lysozyme and Cytochrome C26. On their published article, they observed an unsurprising result on the ‘usual’ two-states kinetics of Lysozyme folding. However, the Cytochrome C folding seemed to be incompatible with the currently known kinetical models featuring the on-pathway intermediates. Such experimental data reveal the first evidence on the existence of the off-pathway intermediates, which added more complexity to the Levinthal
hypothesis. While the on-pathway intermediates might teach us to decode the protein folding process, the off-pathway states seemed to be trivial and uninteresting. However, several critical questions remain: Is the off-pathway intermediate absolutely trivial? Why would the protein sample the off-pathway intermediates if those merely hinder the formation of the final native state?

**The Energetic Tunnel vs. The Energetic Funnel**

The comparative study on protein folding has been performed rigorously by Baldwin, resulting in the distinction between ‘the classical view’ and ‘the new view’\(^{27,28}\). In the early 70s, the available model to study protein folding relied on kinetic experiments by following the population of the native (N) and denatured (D) species over time. The produced data fitted with the hypothetical model, which often matched with the two-state kinetics (Equation 1.1). If the data are incompatible with the two-state fitting, another kinetic model is needed to resolve them by taking into consideration the involvement of the on- or off-pathway intermediate (I) species (Equation 1.2 or 1.3)\(^{26,29}\). In this classical experiment of protein folding, the elucidation of the folding ‘pathway’ depends on the observation of the population shift between all species over time. However, the data do not provide any structural details on the protein conformations, as the classical view relied their experiment merely on the distribution of the observable protein species.

\[
D \rightleftharpoons N \quad \text{(Equation 1.1. Two-state Model)}
\]

\[
D \rightleftharpoons I \rightleftharpoons N \quad \text{(Equation 1.2. Model with the 'ON'-Pathway Intermediate)}
\]

\[
D \rightleftharpoons N \rightleftharpoons I \quad \text{(Equation 1.3. Model with the 'OFF'-Pathway Intermediate)}
\]

In contrast, ‘the new view’ of protein folding is based on the observation of the microstates of the protein. The protein microstates, obtained from the structural information delivering atomic details of each of the species, is critical to decode all available conformers. A proposed model, dubbed as the ‘energetic funnel’ model, which relied on the highly simplified statistical mechanics, is built upon two critical parameters: (i) the conformational states of the protein and (ii) the internal energy of each corresponding state (Figure 1.2A-B)\(^{28,30,31}\). The model is based on the ‘energetic landscape’ concept, which describe “the potential energy (\(E_c\)) of the protein as a function of the conformational coordinates. It is a hypersurface in the high-dimensional
space of the coordinates of all atoms in Mb (Myoglobin – the protein)” as explained by Frauenfelder in 1991\textsuperscript{32}. On the horizontal field of the funnel, several conformational states with minor differences in the 3D space located close to each other, assuming that each dot represents one conformer (\textbf{Figure 1.2B}). The vertical axis represents the internal free energy of each conformation, which derived from all interactions that stabilize the corresponding state. The differences of the free energy generate a 3D contour map, which consists of multiple energetic ‘hills’ and ‘valleys’ (\textbf{Figure 1.2C}). Because the internal free energy also depends on the external condition (i.e., temperature or solvent), the contour map of the energetic funnel is highly dynamic in response to environmental changes\textsuperscript{30}. The conformational sampling during protein folding is analogous to rolling a ball onto the energetic funnel to reach its deepest point. During the folding process, the protein may sample several the more-stable ‘valley’ (intermediates) and the unstable ‘hill’ (transition states) before reaching the native state\textsuperscript{28}.

The concept of the energetic funnel is sufficient to untangle a little misconception of the ‘\textit{classical view}’ regarding the ‘Levinthal Paradox’ and the protein folding pathway, caused by the unavailability of the energetic information and the protein structural details\textsuperscript{28}. To describe the Levinthal Paradox, ‘\textit{the new view}’ visualize such problem as a two-states funnel (\textbf{Figure 1.2D}), in which the protein has to sample multiple isoenergetic ‘haystacks’ conformations before falling into the ‘needle’ native state. These ‘haystacks’ conformations generate a kinetic trap, adequate to delay the folding process. As the Levinthal paradox contradicts the fact that protein folding occurs rapidly, many studies subsequently relied on the concept of the protein folding \textit{pathway} that may explain the speeded process of protein folding, that according to ‘\textit{the new view}’ is described as an energetic \textit{tunnel inside a funnel} (\textbf{Figure 1.2E}). In the energetic \textit{tunnel} representation, the native (N) and denatured state (D), in which each of them is described as an ensemble of conformations, are connected by a fixed ‘\textit{pathway}.’ In this concept, the ‘\textit{on}’-\textit{pathway} intermediates are defined as any intermediate conformations that falls inside the tunnel.

Although the tunnel concept might explain how the protein folding is ‘speeded’ and ‘guided,’ it remains unclear how the folding has to begin from one single point on the funnel. To resolve this issue, the energetic funnel generalized the idea that folding may start with any conformations by using any available pathways (even though that one preferred path may be frequently selected), as presented in \textbf{Figure 1.2C}. During the process, protein may also sample...
occasionally any alternative conformers, previously dubbed as the *off*-intermediates according to the protein *tunnel* model.

**Figure 1.2 | Artistic-rendering of Protein Energetic Funnels.** (A) Idealized Energetic Funnel. The depth of the funnel represents the potential energy as a function of each conformational state. When the protein is optimizing the number of contacts, its internal free energy and the conformational freedom are lowered. The native state (N) is represented as the global minima of the funnel. (B) The slice of the idealized energetic funnel represented by the lattice model where the black and white beads embody the hydrophobic and the polar monomer, respectively. The more compact conformations are holding a high number of hydrophobic contacts \((h)\). As shown in this panel, the native (N) is the only conformation that poses the highest number of contact \((h=5)\); while the other mode of contacts \((h<5)\) might be represented by the more abundant conformations. (C) Rugged energetic funnel comprising multiple transition states (hills – energy barriers), intermediates (valleys – local minima), and kinetic traps (flat surfaces). Many alternative pathways are available to be selected before the protein reaches its native state. (D) Levinthal Paradox represented as a ‘golf-course’ landscape in the energetic funnel model. A random sampling of conformations leads to the kinetic traps during the searching process of the native state (N). (E) Folding pathway exemplified as a tunnel inside the energetic funnel, reasoning the means by which the folding is ‘speeded’ and ‘guided.’ Figure adapted by permission from Springer Nature, *Nature Structural Biology*, ‘*From Levinthal to Pathways to Funnels*’ (Figure 1-5) by K.A. Dill and H.S Chan28, Copyright 1997.

**Multi-tier Protein Dynamics**

Other than the protein folding investigations, the concept of the energetic landscape is also suitable to explain the dynamics of the folded protein. This was extensively shown by Frauenfelder and co-workers by studying the dynamics of Myoglobin during ligand binding30. Any
Conformational changes underlie the protein dynamics, which can be represented as a ‘moving’ dot of the native state in the energetic funnels. In protein dynamics studies, the ensemble of protein microstates are observed under equilibrium conditions using multiple high-resolution instruments. In regards to the microstate changes in the energetic funnel, Dorothee and Kern introduced a multi-tier classification of the protein dynamics. The corresponding classification will help us understand the dynamic personalities represented on the energetic funnel.

**Figure 1.3 | Multi-tiers Dynamics of the Folded Protein on the Energetic Funnel.** The model protein exhibits two-states (A and B) tier-0 dynamics as summarized by Dorothee Kern, according to the tiers classification of Frauenfelder and co-workers. On the presented energetic landscape, a state is represented as the minimum while the maximum between the minima exemplifies the transition state. The energy barriers exemplify the interconversion rate between states, while the distribution of two states (A and B) depends on the energy difference between them. Changes in environmental conditions (i.e., temperature, solvent, or the presence of ligand) will affect the energetic landscape, shifting the equilibria between the two states. While the interconversion of Tier-0 states are exhibited on micro- to millisecond timescales, the Tier-1 and -2 are displayed on a faster period. Figure adapted by permission from Springer Nature, Nature, ‘Dynamic Personalities of Protein’ (Figure 1A) by K. Henzler-Wildman and D. Kern. Copyright 2007.
There are three tiers of dynamics which imply two critical aspects: (i) the timescale of transitions; and (ii) the type of a structural changes and its directionality\(^{35,38}\). Both aspects correlate tightly with two fundamental characteristics of the energetic funnel that exemplify the internal free energy and the conformational state of the protein. The multi-tier dynamics classification relied on the height of the energetic barrier, which is the manifest of the timescale of each transition (Figure 1.3). Dynamics that are separated by several \(k_B T\) (where \(T\) and \(k_B\) define temperature and Boltzmann constant, respectively) exhibit slow structural fluctuations on a microsecond to millisecond timescale at physiological temperature\(^{35}\). The corresponding dynamics, called Tier-0 dynamics, involve the tertiary conformational changes as imposing the global collective motions of the protein. Protein that observed to be static on Tier-0 levels; fluctuate on a lower-tier, affecting local protein flexibilities occurring on a faster timescale. Indeed, Tier-1 and Tier-2 dynamics, separated by a lower energetic barrier, have a faster time scale ranging from picosecond to nanosecond. The structural changes involved in these two tiers include secondary structure changes or even smaller conformational changes such as rotation of amino-acid’s side chain.

Multi-tiers dynamics can explain the significant conformational changes needed to occur for protein function. Because the changes of the external condition affect the dynamics and the energetic funnel\(^{30}\), addition of substrates or ligands may also reforms the whole energetic landscape and drastically stabilize the novel native state of the new holo funnel (Figure 1.3 - Dark Blue Line to Light Blue Line). As previously mentioned, a single protein in equilibria has a highly dynamic energetic funnel and always sample the most stable conformation\(^{28}\). In holo conditions, the ligand may stabilize the unfavorable apo conformation by facilitating the formation of several key interactions locally on the active site. The following changes can also be allosterically transmitted further to another distal site and thus initiating the global collective motions of a protein. The more detailed analysis on the protein allosteric network, which will not be covered in this thesis, requires: (i) the rigorous inspection on the ensemble of the protein microstates that could be addressed with the NMR-style dynamics and (ii) the implementation of the Ensemble Allosteric Model that also relied on the energetic landscape. By following the conformational changes at a single molecule level and determining their kinetics and energetic barriers, we might be able to elucidate the key multi-tier dynamics which exist as a response to specific environmental changes leading to function.
**Protein Evolution**

Despite the ‘classical view’ presumption that proteins exhibit one dominant native state that is correlated with a single physiologically relevant function, other isoenergetic alternative conformations are usually available to be sampled. The availability of these protein alternative conformations is dubbed as the conformational variability on the avant-garde ‘new view’ theory of protein dynamics. While the conformational variability seemed to be trivial to be inspected for protein dynamics studies, it may contain critical information to trace back the protein evolutionary trajectory. On a related article regarding protein evolvability, Dan Tawfik and co-workers illustrated how nature selects the most physiologically relevant protein conformers as a response to environmental changes while maintaining the fitness. By efficiently eliminating the alternative conformers, the protein evolved towards a more specific function.

**Protein Conformational Variability and Functional Promiscuity**

Conformational variability is an inherent property of any protein, which defines its dynamical flexibility and evolutionary adaptability. According to this definition, proteins are expected to exist at an equilibrium in the protein ensemble between their native state and the alternative conformers. As previously mentioned, the ensemble of these numerous conformers describes the protein dynamics. During evolution, the minimization of the non-functional conformers might result in a more pronounced native state that leads to restricted dynamics and specific functions. On the other hand, protein functional promiscuity defines the degree of various coincidental functions that contradict the original one for which the protein has been evolved. Any modification that might be deleterious for the alternate conformers would reduce the functional promiscuity. Recent studies hypothesised that functional promiscuity is the starting point for the divergence and acquisition of new protein functions.

The conformational variability might be observed from any tiers of protein dynamics ranging from: (i) the local fluctuations of amino-acid side chains, (ii) the secondary-structure reformations, or even (iii) the global tertiary dynamics. These conformational variabilities are correlated with protein functional promiscuity. One example of conformational variability that leads to functional promiscuity is the existence of intrinsic dynamics observed in the monoclonal antibody (SPE7) which adopts two conformers generating distinct binding sites, the ‘flat’ and ‘narrow and deep’. While the presence of the dominant ‘flat’ binding site facilitates the interactions of SPE7
with the protein antigen, the existence the ‘narrow and deep’ binding site allows the antibody to sense the hapten molecule (dinitrophenol – DNP). This study demonstrates that a single protein with a fixed sequence can select the pre-existing conformers to interact with different ligands. Interestingly, recent studies proposed that directed evolution is proficient to shift the dominant original function to the promiscuous one by altering a defined numbers of residues on a fix length protein\textsuperscript{41,45}.

The ancestral protein, which is presumed to have a high degree of conformational variability, is believed to pose increased functional promiscuity\textsuperscript{14,46}. Such characteristics of the ancestral protein facilitate the evolutionary divergence towards a new specific function\textsuperscript{47}.

![Figure 1.4 | Protein Conformational Variability and Functional Promiscuity. The represented protein model has a structural flexibility to pose many alternative conformers. One of those conformers is the native state, which is the most dominant and physiologically relevant one. The alternative conformers, which are available to be sampled, might lead to the promiscuous functions.](image)

**Protein Evolvability**

Evolution is the natural fixation of structural and functional changes which will result in a high degree of functional specificity by minimizing the population of the minor unfunctional state(s)\textsuperscript{13}. The native state of a protein has been evolved and fixated by ‘natural selection’ as a result of environmental adaptations. The evolved protein, possibly gained its primary function by selecting and optimizing the dominant conformer, a process facilitated by mutations, additions, or deletion of the protein sequence.
The evolvability, representing the ability of a protein to evolve, depends on two main factors: its plasticity and neutrality. Protein plasticity describes the structural adaptability of protein in a series of complementary adjustments to mutations. However, some mutations may radically alter the original protein function, which is essential for survival. How then the evolving protein accommodate a series of evolutionary mutations? Interestingly, another contradictory element to plasticity needed during protein evolution, is represented by the concept of neutrality. Neutrality defines the ability to retain a neutral mutation, which causes no observable effect on the original structure and function of the protein and mostly affecting the promiscuous conformation. If a neutral mutation causes a deleterious effect on the promiscuous conformation, it might evolve the protein by shifting the equilibria towards the more specific native state. However, the neutral mutation might also introduce a new conformation and increase protein fitness, which beneficial to facilitate future protein evolution in response to environmental adaptions. As to the latter case, the protein might be able to emerge a novel function with several additional evolutionary mutations. While neutrality does not implicate the structural novelty, plasticity often demands a radical change to accommodate evolutionary modifications. However, these two elements synergistically pinpoint the evolution problem as described extensively on Dan Tawfik article in 2005:

“Our hypothesis provides a possible solution to these conflicting features at the single-protein level: rapid adaptability is inherent to the promiscuous, accidental functions of the protein, and their plasticity need not be at the expense of the protein’s original activity. ………. Results from directed laboratory evolution experiments indicate that the evolution of a new function is driven by mutations that have little effect on the native function but large effects on the promiscuous functions that serve as starting point. Thus, an evolving protein can initially acquire increased fitness for a new function without losing its original function. Gene duplication and the divergence of a completely new protein may then follow.”

Based on the evidence of recent studies, the more evolved proteins pose a higher level of specificity and activity. However, several questions remain unclear concerning the protein evolvability: Is there any trade-off between activity and evolvability? How did the first proteins
evolved while the structural novelty does not always give any selective advantage for the protein? Furthermore, does the global structural flexibility provide higher evolvability of the fold?

**Tracing Back the Long-period Evolution**

The evolvability theory can faithfully explain how the short period evolution is “introduced” on a fix-length protein\(^{11,13}\). Multiple rounds of evolutionary mutations are sufficient to alter the lower tiers' dynamics of the native state by minimally affect the whole protein sequence. To extract the protein structural information, the dynamic ensemble of the native state has been captured by using an atomic-resolution technique such as Nuclear Magnetic Resonance (NMR)\(^{37,49,50}\). Additionally, Molecular Dynamics (MD) performed to explain the protein conformational changes at an atomic resolution, which use either a model or the high-resolution structural information as the starting point of the simulations\(^{51-53}\). The results are then subsequently correlated by tracing the protein activity and identifying the new specificity\(^{54}\). However, it is challenging to recreate the energetic landscape representation of an evolved protein since such involves conformational changes of the lower tiers, thus extremelly fast processes. Consequently, the energetics of each atomic-resolution ensemble is difficult to be traced.

In this thesis, we aim to reveal the evolution of proteins that alter mostly the high-tier dynamics, which hypothetically arose during more extended evolutionary periods, facilitated by the gene recombination on both of the protein N- and C-termini. To confirm such hypotheses, a class of proteins is introduced that possess secondary and tertiary structural similarity, which we dubbed as Cherry-Core Proteins (CCPs). Multi-tier dynamics were traced by using the single-molecule technique such as Alternating Laser Excitation (ALEX) single-molecule Forster Resonance Energy Transfer (smFRET)\(^{55-57}\) and the high-resolution ensemble technique of Hydrogen-Deuterium Exchange Mass Spectroscopy (HDX-MS)\(^{58-60}\) coupled with MD Simulations. As the ligand-induced conformational changes occur on the higher-tier dynamics, it is possible to calculate the energetics on the ensemble of the native states using Isothermic Titration Calorimetry (ITC)\(^{61}\). Therefore, the complete perspective, which includes the protein energetic landscape needed to view the protein evolvability, is highly plausible to be traced.

In this thesis, we will focus to untangle the evolvability and the structural dynamics of the Cherry-core proteins that might occurred during the long-evolutionary period. To present the outcomes of our study, we curated such results as the chapters of this thesis as shown below:
Chapter 2

A new class of proteins was identified, which dubbed as Cherry-core (CC) Proteins, to study the long-period evolution and their evolvability. These proteins mainly comprise a core composed of two simple Rossmann-fold domains connected by a single β-strand. CCPs pose a unique symmetry and flexibility, which is hypothesized to contribute to the evolvability of the core. By acquiring the N- or and C-termini extensions, these proteins develop various functions ranged throughout diverse organisms in the kingdoms of life. In this chapter, we present an extensive structural analysis based on the high-resolution structural information of CCPs.

Chapter 3

Following the previous chapter, distinct protein functions of the cherry-core emerged and were correlated with the presence of distinct termini. In this chapter, the critical multi-tiers dynamics of the CCPs needed for their function were determined by using various techniques such as single-molecule FRET (Tier-0) and HDX-MS (Tier-1). We assessed the role of the termini in dictating these fundamental dynamics by combining the results of structural analysis and experimental data, on the following ‘model’ proteins: (i) Cyanate Transcriptional Regulator-CynR, (ii) Glutamine Binding Protein-SBD2, (iii) Carbonate Binding Protein-CmpA, and (iv) Maltose Binding Protein-MalE.

Chapter 4

In this chapter, we aim to identify the quaternary dynamics of LTTR that is hypothesized to be modulated by N-termini region. We focused our study on the full-length CynR and sought to elucidate if the effector (i.e., azide) might play a pivotal role in altering the quaternary dynamics. We employed Intrinsic Tryptophan Fluorescence and Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) in combination with extensive structural analysis to deciphering the dynamics of CynR.

Chapter 5

The multi-tiers dynamics on MalE during ligand binding were identified using multiple techniques such as smFRET (Tier-0), HDX-MS (Tier-1), and Molecular Dynamic Simulation (Tier-2). We address at the residue level, the means by which the CC of MalE co-evolved by utilizing...
their ‘modules’ to complement the presence of the its tail during evolution required for gain function. Structural details presented on this chapter reveals the insight of the presence of molecular ‘magnets’ to ‘create’ the conformational state able to bind the ligand. Subsequently, we also address the asymmetric ligand binding process occurring on the rigid domain in the early phase of ligand recognition that is necessary to trigger the conformational transition. We tried to reconstruct the energetic landscape of the CCPs by complementing structural data data reporting on the multi-tier conformational ensemble and the thermodynamics of binding.

Bibliography


