Preliminary analysis of protein catalytic ability and catalytic structure

Zheming Cao *

Posted Date: 18 July 2024

doi: 10.20944/preprints202407.1526.v1

Keywords: protein; high-energy state; catalysis; AlphaFold

Article

Preliminary Analysis of Protein Catalytic Ability and Catalytic Structure

Zheming Cao

Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization, Ministry of Agriculture, Freshwater Fisheries Research Center, Wuxi, China; caozm@ffrc.cn

Abstract: When water molecules or other particles collide with protein molecules in a solution, the protein molecules absorb some of the momentum and deform, forming a certain molecular potential energy. When a protein molecule in a high-energy state changes to a stable state, if there is a suitable substrate binding and the potential energy of the protein molecule is greater than the energy required to break the substrate chemical bond, catalytic action will occur. The α-helices and β-folds in protein molecules play a major role in converting the kinetic energy of water molecules into their own potential energy due to the large number of amino acids involved and their regular structure. The internal structural deformation of proteins caused by the impact of water molecules on α-helices and β-folds also has a profound impact on the structure of substrates.

Keywords: protein; high-energy state; catalysis; AlphaFold,

Proteins are the material basis of life, the fundamental organic compounds that make up cells, and the main carriers of life activities. Amino acids are the fundamental building blocks of proteins. Proteins have primary, secondary, tertiary, and quaternary structures. The primary structure refers to the type, number, and arrangement of amino acids in the polypeptide chain that makes up the protein; The secondary structure is a structure formed by hydrogen bonding at corresponding points between amino acid residues on the basis of the primary structure, forming helices or folds, mainly including α-helices and β-folds; The tertiary structure is a specific three-dimensional conformation formed by further coiling and folding on the basis of the secondary structure; The quaternary structure refers to the structural subunits of protein molecules formed by two or more peptide chains on the basis of their respective tertiary structures, which are linked by non covalent bonds to form more complex spatial structures. The key factors that maintain the structure at all levels include hydrogen bonding, hydrophobic interactions, van der Waals forces, and ionic bonds, which work together to ensure that proteins can fold correctly and perform their biological functions.

Protein structure determination usually requires steps such as gene expression, protein extraction and purification, crystallization, and X-ray diffraction analysis. Nearly 90% of protein structures are determined using X-ray crystallography. X-ray crystallography can determine the spatial distribution of electron density of protein molecules in crystals, and resolve the three-dimensional coordinates of all atoms in proteins at a certain resolution. Cryo electron microscopy is a method that has emerged in recent years to obtain low resolution protein structures. Cryo electron microscopy rapidly freezes aqueous biological samples and irradiates them with electron beams, and finally uses three-dimensional reconstruction techniques to obtain structural data of the target. The 2017 Nobel Prize in Chemistry was awarded to the inventors of cryoelectron microscopy, Jacques Dubochet, Joachim Frank, and Richard Henderson.

The main purpose of protein structure analysis is to screen for new drugs. Drug molecules must be tightly bound to target protein molecules, easy to synthesize, and have no toxic side effects. Traditional drug design relies less on the three-dimensional structure of the target protein by screening a large number of natural compounds, known substrates or ligand analogues, and biochemical studies to determine lead compounds. With the growth of protein structure data and the development of structure prediction technology, the information of the three-dimensional structure

 doi:10.20944/preprints202407.1526.v1

of target protein molecules plays an increasingly important role in the above process. In 2022, DeepMind, an artificial intelligence company under Google, further cracked almost all known protein structures. The database constructed by its AlphaFold algorithm now contains more than 200 million known protein structures, paving the way for the development of new drugs or technologies to address global challenges such as famine or pollution.

The method of screening new drugs based on protein structure has a bright prospect, but there are currently no successful reports. Brownian motion refers to the uninterrupted and irregular movement of particles suspended in a liquid or gas. This type of motion is caused by the irregular thermal motion of liquid or gas molecules, resulting in particles being subjected to collisions and imbalances from molecules in various directions. Proteins are difficult to maintain stability in solution due to the impact of water molecules, and the structure obtained through methods such as cryo electron microscopy cannot represent their true state. Below, we will explore this issue preliminarily.

We will now establish several hypotheses and analyze protein catalytic changes based on them:

1. Under protein crystals or cryo electron microscopy, protein is the lowest energy state, that is, the most stable state of the protein.

Firstly, we need to understand why many types of proteins have catalytic effects. Proteins in their crystalline or frozen state are thermodynamically the lowest energy state and have no catalytic ability. We need to ask how the catalytic ability of enzymes is generated. If an enzyme maintains its most stable state at any temperature, it definitely has no catalytic ability. At higher temperatures, such as body temperature, due to the Brownian motion of water molecules, proteins cannot maintain their most stable state. When subjected to irregular impacts from water molecules, the protein structure undergoes deformation, resulting in structural stress (just like a spring being compressed and generating potential energy). Enzymes with this structural stress are in a high-energy state and need to return to a stable state at the corresponding temperature. If there is a suitable substrate, the release of this structural stress will break the key chemical bonds of the substrate and produce new substances.

2. Proteins undergo deformation under the impact of external water molecules, which generates structural stress

We assume that a protein is impacted by a certain number of water molecules or other particles (momentum: $M_{n1}V_{n1}$) and generates a momentum (M_PV_P) in a certain direction, while a portion of water or other particles gain additional momentum ($M_{n2}V_{n2}$). Therefore, part kinetic energy of the water molecules that impact the protein is converted into potential energy, and the formula is:

$$E_p = \sum_{k=1}^{n_1} \frac{1}{2} m_{n_1} {v_{n_1}}^2 - \frac{1}{2} m_p v_p - \sum_{k=1}^{n_2} \frac{1}{2} m_{n_2} {v_{n_2}}^{22}$$

Among them, E_P is the total potential energy obtained by the protein, and catalytic action occurs when this total potential energy is greater than or equal to the energy required for chemical bond breakage or connection required for a substrate change.

1. Calculation of enzyme catalytic ability

Due to the fact that not every deformation can generate catalytic activity, catalytic ability (equivalent to enzyme activity) is the number of times the total potential energy of a protein reaches or exceeds the minimum line per unit time. Enzyme activity refers to the ability of enzymes to catalyze certain chemical reactions. The magnitude of enzyme activity can be expressed by the conversion rate of a certain chemical reaction it catalyzes under certain conditions. The formula for calculating enzyme activity is:

$$U=\Delta A/（0.01\times t）\times D$$

(U represents enzyme activity unit; Δ A is the change in absorbance value during the reaction time; T is the reaction time; D is the dilution factor, which means the total enzyme solution extracted is a multiple of the enzyme solution in the reaction system)

 doi:10.20944/preprints202407.1526.v1

Enzyme activity is influenced by various factors, including enzyme concentration, substrate concentration, temperature, pH value, activators, and inhibitors. These factors affect the speed and efficiency of enzymatic reactions through different mechanisms. These factors work together to determine the specific conditions and outcomes of enzymatic reactions.

The effect of temperature on enzyme activity is twofold: an increase in salt concentration and solution viscosity both affect the movement of water molecules and the number of times the total potential energy of proteins reaches or exceeds the minimum line per unit time.

On the one hand, as the temperature increases, the number of activated molecules increases. As the temperature rises, the speed of water molecule movement increases, the impact force increases, and the number of times the total potential energy of the protein reaches or exceeds the minimum line per unit time increases, resulting in an accelerated enzymatic reaction rate; On the other hand, as the temperature increases, the enzyme gradually denatures, reducing its reaction rate. It should also be mentioned that the catalytic domain changes more vigorously with increasing temperature, which hinders substrate binding to the appropriate region of the enzyme and reduces its catalytic efficiency.

2. Changes in catalytic domains

Proteins undergo deformation within a certain range under the impact of external water molecules. Since this deformation is irregular, the answer is obviously not whether all of these deformations can produce catalytic effects. In addition to reaching the minimum cell potential energy, it is also necessary to have a suitable substrate structure, which requires a specific analysis of how the total potential energy is generated.

There are four types of protein secondary structures (α-helices ,β-folds, β-angle and irregular curl). We believe that the deformation of β-angle and irregular curl may alter the structure of the catalytic domain, but these regions are designed with fewer amino acids and do not form large regular structures. Their deformation makes it difficult to generate sufficient potential energy to intervene in catalytic activity. Their function is to form a suitable catalytic domain, maintain the basic tertiary structure, and have little effect in the catalytic process. Due to the involvement of numerous amino acids and their distinct structures, we hypothesize that the deformation of α-helices ,β-folds can generate significant stress. Moreover, α-helices ,β-folds can also generate super secondary structures, which may in some way more efficiently convert the energy brought by water molecule collisions into the total potential energy of proteins. So the deformation of the α-helices ,β-folds determines both the catalytic ability and the structure of the catalytic substrate.

The inward displacement （Δl）caused by the deformation of α-helices and β-folds is directly generated by the impact force of water molecules, which conforms to the following formula:

$$\sum mv = K\Delta L$$

Among them, K is the characteristic coefficient of a certain α-helices or β-folds, which is related to the length of the α-helices or β-folds and the composition of the amino acid species. The calculation of specific coefficients requires analysis of experimental results.

I cannot find any literature to deduce when substrates bind to proteins, so I can only make assumptions. Because proteins are constantly changing in solution. In theory, as long as the protein catalytic domain changes to the appropriate shape, the substrate can enter. From the current research reports on protein catalysis, it can be seen that most substrates do not bind to proteins in their most stable state. I believe that most substrates bind to proteins in a high-energy state, and I cannot rule out the possibility that a few proteins may only begin to deform and produce catalytic effects after binding to substrates.

Due to the lack of catalytic ability in stable proteins, predicting protein structures based on crystal structures and designing drugs based on these structures is meaningless. If the stress of α-helices and β-folds is the basis of catalysis, then their deformation on the structural domain is the key to catalysis, and α-helices and β-folds occupy the vast majority of the enzyme's surface. Therefore, it is necessary to calculate the influence of the deformation of α-helices and β-folds on the enzyme catalytic domain structure in order to obtain the correct substrate structure. The true catalytic ability

 doi:10.20944/preprints202407.1526.v1

comes at the moment of protein deformation, when the catalytic domain can accurately design the structure of drugs.

Discussion:

It once caused a sensation that Chunyu Han 's discovery that Ngago has the ability to cut DNA, but as the vast majority of people were unable to replicate the experiment, Han was accused of fraud and the paper was withdrawn. We found through analysis that Ngago has the ability to bind guiding DNA and target DNA, but does not have the ability to cleave DNA. The above theory can reasonably explain this phenomenon. At the experimental temperature, NgAgo's deformation can bind guiding DNA and target DNA, but the potential energy of this deformation is lower than the energy required to cut the chemical bond of the target DNA. I think Chunyu Han may not have noticed a certain factor affecting Ngago's deformation potential energy (or the quantum tunneling effect producing a cutting reaction with a potential energy lower than that of chemical bonds) in the experiment, thus producing the ability to cut DNA (so I don't think Han Chunyu fabricated it, because this phenomenon is theoretically possible). If the temperature is increased, the deformation potential energy of NgAgo increases, which can exceed the energy required to cleave the target DNA chemical bond. But the problem is that the increase in NgAgo's deformation potential energy may result in NgAgo no longer having the ability to bind to guiding DNA and target DNA, which can only be solved through experimental methods. But the efficiency of this enzyme should not be very high, and its practical value is not significant.

In May 2024, researchers from DeepMind and Isomorphic Labs published a research paper titled "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3" in Nature. This study introduced AlphaFold3, which covers unprecedented breadth and accuracy, and can accurately predict the structure of protein interactions with various other biomolecules. Being able to predict the structure of complexes containing almost all types of molecules in protein databases, including how ligands, proteins, and nucleic acids aggregate and interact with each other, as well as predict the structural effects of post-translational modifications and ions on these molecular systems, thus helping us to accurately observe the structure of biological molecular systems at the elementary level. I think it's still an old problem. Proteins are in a constantly changing state in solution, and the various properties of proteins predicted based on their crystal structure may not necessarily match reality.

The twenty amino acids that make up proteins are very puzzling. Why did biological evolution determine these twenty amino acids instead of other types? Why do aα-helices and β-folds become the main structures of most proteins? If RNA is the earliest molecule with self replication function, why did RNA ultimately evolve into a comprehensive system such as DNA and proteins instead of a single organism? The theories presented in this article provide some implications for these issues, and more experimental evidence is needed to obtain specific results. If the experiment proves that the theory is correct, we can design proteins based on this theory.

Reference:

1. Feng Gao, Xiao Shen, Feng Jiang, Chunyu Han, (2016) DNA-guided genome editing using the Natronobacterium gregoryiArgonaute, Nat. Biotechnol.34, 768-773;
2. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature (2024) doi:10.1038/s41586-024-07487-w.
3. Helen R. Saibil, Cryo-EM in molecular and cellular biology ,molecular cell, 2022,82(2):274-284