LECTURE 12: NATURAL ENZYMES

DESIGNER ENZYME

INTRODUCTION

The success of inverse protein folding reached a milestone with the design and successful experimental proof of structure of a 93-residue □/□ protein called Top7.[1] The next great challenge of protein design is to predict and create a functional protein for a reaction for which no natural enzymes are known.

The first successful response to this challenge was achieved in a collaborative effort between Rothlisberger, D group and that of David Baker (University of Washington) protocol for enzyme design. This started with the design of an active site with the appropriate functionality for catalysis using quantum mechanical calculations. This is called as theozyme, for theoretical enzyme. The advantage of this computational method is the ability to tune optimal functionality that a protein can use to bind and stabilize the transition state. Combining chemistry, mathematics and physics, the scientists reported in the Nature paper that they have successfully created designer enzymes for a chemical reaction known as the Kemp elimination, a non-natural chemical transformation in which hydrogen is pulled off a carbon atom. other chemists reported another successful chemical reaction that uses designer enzymes to catalyze a retro-aldol

NATURAL ENZYMES

Natural enzymes, which are relatively large protein molecules, are the powerful catalysts that control the reactions that sustain life. They play a central role in the chemical reactions involved in the transformation of food into the essential nutrients that provide energy, among many other critical functions.

Although enzymes are superb catalysts, their range of reactions is limited to those that support life. Their repertoire could be expanded by a method that allows artificial enzymes to be made from scratch.

PROTEIN ENGINEERING

Enzymes are astoundingly good catalysts: they allow reactions to occur billions of times faster than would be possible without them, at temperatures much lower than those required by typical synthetic catalysts. But enzymes have evolved to accelerate only biological reactions, under the narrow set of conditions that are compatible with life. Two papers from the same group, one in this issue (Röthlisberger *et al.*[1], page 190) and another in *Science* (Jiang *et al.*[2]), show how these limitations can be overcome. They describe a method for designing enzymes that catalyse unnatural reactions, and demonstrate its use for two different chemical transformations.

Enzymes work by lowering the activation energy of reactions, specifically by confining substrates in binding sites that stabilize the highest-energy arrangement of atoms in the reaction pathway (known as the transition state). They also shield the reactants, thus preventing possible side reactions. The idea behind the latest work[1, 2] is simple — model the transition state for a reaction, stabilize it by surrounding it with carefully placed chemical groups, graft the resulting active site into an existing protein and then alter the amino-acid sequence of the protein to accommodate the changes. In practice, this is a complicated procedure. For starters, building an accurate model of a transition state requires a detailed understanding of the reaction's mechanism, which isn't always available. Furthermore, transition states are modelled using quantum-mechanical calculations, but currently available methods can handle only a limited number of atoms, and are often inadequate for modelling enzyme reactions.

Designing a protein that folds into a given structure is equally challenging. For a protein made of 100 amino acids, there are about 10^{130} possible sequences, each of which can adopt many different conformations. The thermodynamic stability of every sequence and conformation must therefore be calculated to find the lowest-energy structure (that is, the one most likely to form). Some simplifications can be made using advanced computational methods to quickly eliminate unfavourable combinations. This has resulted in several notable accomplishments, such as the complete redesign of a protein consisting of 28 amino acids[3], the design of an amino-acid sequence that forms a structure not found in nature[4], and the engineering of naturally occurring proteins into biosensors for trinitrotoluene (TNT) and other small molecules[5].

With these precedents, you might think that designing catalytic proteins should be straightforward, but success has been limited. Catalytically inactive proteins have been converted into modestly catalytic ones for two different reactions, but the observed enhancements of rate[6, 7]were only about a millionth of those produced by naturally occurring enzymes. It is also sometimes difficult to prove that designer enzymes are truly catalytic on the basis of biochemical observations, and some exciting claims have been found to be flawed.

But some reports of catalysis by designed enzymes have fared rather better — especially those that are based on sound crystallographic evidence[6, 7, 8]. An essential step in demonstrating the success of a designer enzyme, therefore, is the determination of a high-resolution crystal structure for the protein, to verify that the designed catalytic features are present. The results of Röthlisberger *et al.*[1] and Jiang *et al.*[2] are remarkable in the spectacular agreement between their computationally predicted enzyme models and the experimentally determined structures (Fig. 1).

Figure 1: Enzymes by design.

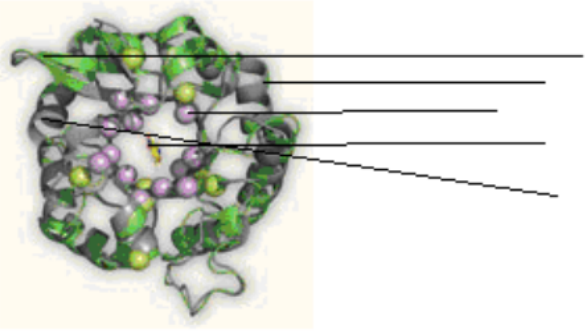


Fig 1 description: Röthlisberger *et al.*[1] have computationally designed and prepared the first enzyme capable of catalysing a non-biological reaction. Here, the computational model (grey) is overlaid with the crystal structure of the actual protein (green); the two overlap almost perfectly. The substrate is shown at the centre of the structure. The design process involved modifying the amino-acid sequence of a naturally occurring protein. Residues selected computationally to form the active site are shown as purple spheres.

Additional mutations that were introduced *in vitro* to optimize the enzyme's performance are shown as green spheres.

EXAMPLE 1 Röthlisberger *et al.*[1] made an enzyme that catalyses the Kemp elimination reaction. The Kemp elimination is initiated by the removal of a hydrogen ion from a carbon–hydrogen bond in the substrate; the minimum requirement for catalysis of the reaction is the presence of a base to perform this step. The authors therefore identified two amino acids — aspartic acid and histidine — that have side chains that can act as bases under physiological conditions, and used these as the starting points of their putative active sites. They decorated models of the proposed active sites with other chemical groups found in proteins, choosing those that could interact favourably with groups in the substrate. They then used state-of-the-art quantum-mechanical methods to precisely place all the groups in the models to maximize stabilization of the transition state of the substrate. The authors thus obtained a large ensemble of designs for catalytic sites in enzymes.

Next, Röthlisberger *et al*. selected about 100 proteins that could be used as scaffolds for their proposed active sites. The criteria for selection were the availability of high-resolution crystal structures and the presence of pre-organized cavities, with a preference for proteins that behave well in experiments (that is, those that have good solubility, are expressed easily in cells, and so on). The authors then used computational methods to search each of the proteins for specific regions that could accommodate the sites, narrowing down the vast number of possibilities to about 100,000 promising leads. These were whittled down further using an automated modelling technique to find the optimal amino-acid sequence in defined shells around the active site, selecting sequences that maintained protein stability and integrity.

This computational screening method picked out 59 candidate enzymes, which the authors expressed in cells and evaluated for their ability to catalyse the target reaction. Only eight of the proteins had measurable catalytic activity. The team then used *in vitro* evolution to further optimize one of their successful leads (designated KE07), mutating the amino-acid sequence in both random and directed locations. After several rounds of

mutation and screening, Röthlisberger *et al*. obtained improved enzymes that were up to 200 times more active than KE07. The best two of these mutants accelerate the rate of the Kemp elimination reaction to about a million times that of the uncatalysed version.

The strategy used by Röthlisberger *et al*.[1] promises to be general, as the same group[2] has successfully applied the procedure to another chemical transformation known as the retro-aldol reaction, which is very different from the Kemp elimination. The complexity of the design procedure is underlined by the number of interdisciplinary groups involved in the work, and by the huge amount of computational power required to solve the problem — donated from hundreds of thousands of idling computers around the world as part of a project known as Rosetta@home[9].

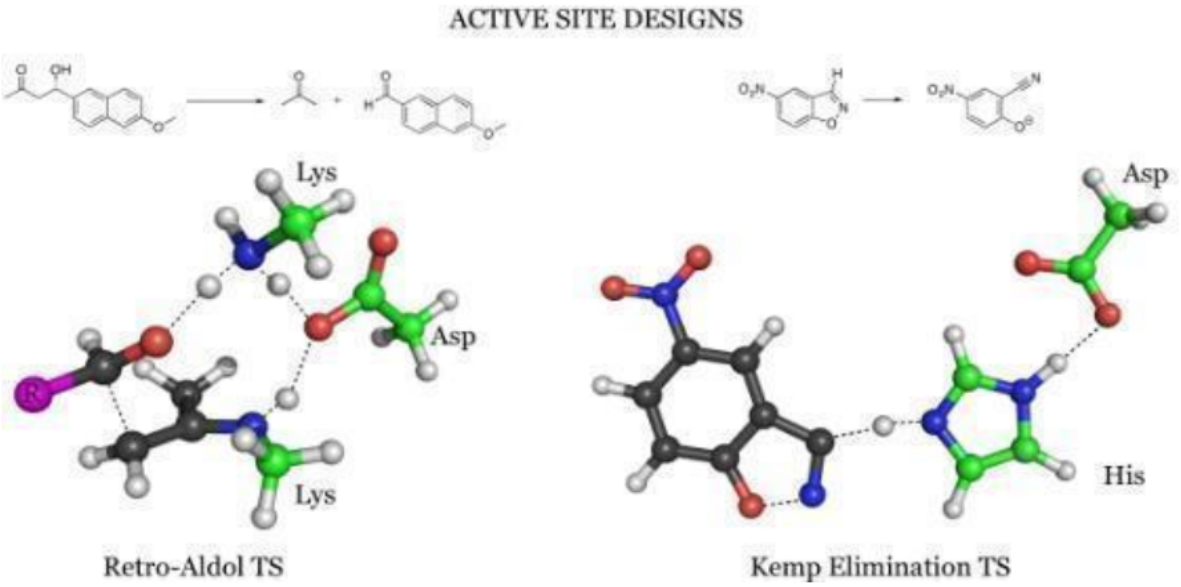
Those in the know might say that the performance of the designed enzymes is far from impressive — the reaction-rate enhancements for typical, naturally occurring enzymes are anywhere between 10,000 and 1 billion times higher than those of the artificial enzymes described in these papers[1, 2]. Furthermore, the chosen reactions are relatively easy targets. The Kemp elimination is accelerated by several catalysts, including various synthetic compounds, catalytic antibodies and even serum albumin. Similarly, the retro-aldol reaction is catalysed by antibodies[10] and by various peptides[11, 12]. Indeed, the rate enhancements reported by Röthlisberger *et al*.[1] are equivalent to those of only the most sophisticated catalytic antibodies[13, 14]; the enhancements obtained by Jiang *et al*.[2] for the retro-aldol reaction are even more modest.

Another limitation of the design process is that, although naturally occurring enzymes have evolved to optimize steps other than just catalysis (such as the binding of substrates and the release of products), the model used by the authors[1, 2] to design their enzymes doesn't attempt to address these factors. This is understandable, because many of the finer features that provide enzymes with their unique properties are not yet understood. For example, the mutations introduced by the authors into their enzymes by directed evolution did not modify the active site itself, but occurred at neighbouring positions (Fig. 1). The effect of some of these mutations can be easily understood with hindsight,

but others are much less obvious. It was therefore wise of the authors to let nature lend a helping hand in their designs.

Nevertheless, these results[1, 2] are a milestone in biochemistry. For the first time, artificial enzymes have been designed for non-biological reactions, providing rate accelerations that are about 1,000 times faster than previous examples of computationally designed enzymes. Biochemists have long wanted to build artificial enzymes to identify and validate the minimal requirements for enzyme-like catalysis. These reports provide an accurate framework for this enterprise to which further features can be added. As Röthlisberger *et al.*[1] note, the ability to design enzymes will truly test our understanding of enzyme catalysis.

EXAMPLE 2 Other chemists reported another successful chemical reaction that uses designer enzymes to catalyze a retro-aldol



reaction, which involves breaking a carbon-carbon bond. The aldol reaction is a key process in living organisms associated with the processing and synthesis of carbohydrates. This reaction is also widely used in the large-scale production of commodity chemicals and in the pharmaceutical industry, Houk said.

PROTEIN ENGINEERING

The implementation of the aldol reaction in the active site of an enzyme has been an important challenge. The reaction involves at least six chemical transformations, requiring UCLA scientists to compute all six chemical steps with their corresponding transition states. The structures were then combined in such a way to allow all six steps to occur.

Using algorithms and supercomputers, the UCLA chemists design the active site for the enzymes — the area of the enzymes in which the chemical reactions take place — and give a blueprint for the active site to their University of Washington colleagues. Baker and his group then use their computer programs to design a sequence of amino acids that fold to produce an active site like the one designed by Houk's group; Baker's group produces the enzymes.

UCLA uses modern computational methods based on the physical laws of quantum mechanics to study in detail the mechanisms of chemical reactions. They have been involved in the DARPA-funded Protein Design Processes program, whose goal is to develop the technology that would make possible the design and creation of man-made working enzymes. The role of UCLA chemists has been the design of the active sites of the enzymes. By exploring multiple combinations of chemical groups, they can determine those that are most suitable to facilitate any given chemical transformation. Then, they determine the precise three-dimensional arrangement of these chemical groups, which is critical for the specificity and activity of the enzyme, with an accuracy of less than a hundredth of a nanometer.

MODEL MAKE UP AND COMPUTER GRAPHICS

Molecular modelling encompasses all theoretical methods and computational techniques used to model or mimic the behaviour of molecules. The techniques are used in the fields ofcomputational chemistry, drug design, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modelling of any reasonably sized system. The common feature of molecular modelling techniques is the atomistic level description of the molecular systems. This may include treating atoms as the smallest individual unit (the Molecular mechanics approach), or explicitly modeling electrons of each atom (the quantum chemistry approach).

Molecular mechanics is one aspect of molecular modelling, as it refers to the use of classical mechanics/Newtonian mechanics to describe the physical basis behind the models. Molecular models typically describe atoms (nucleus and electrons collectively) as point charges with an associated mass. The interactions between neighbouring atoms are described by spring-like interactions (representing chemical bonds) and van der Waals forces. The Lennard-Jones potential is commonly used to describe van der Waals forces. The electrostatic interactions are computed based on Coulomb's law. Atoms are assigned coordinates in Cartesian space or in internal coordinates, and can also be assigned velocities in dynamical simulations. The atomic velocities are related to the temperature of the system, a macroscopic quantity. The collective mathematical expression is known as a potential function and is related to the system internal energy (U), a thermodynamic quantity equal to the sum of potential and kinetic energies. Methods which minimize the potential energy are known as energy minimization techniques (e.g., steepest descent and conjugate gradient), while methods that model the behaviour of the system with propagation of time are known as molecular dynamics.

$$E = E_{\text{bonds}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{non-bonded}}$$

$$E_{\text{non-bonded}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}$$

This function, referred to as a potential function, computes the molecular potential energy as a sum of energy terms that describe the deviation of bond lengths, bond angles and torsion angles away from equilibrium values, plus terms for non-bonded pairs of atoms describing van der Waals and electrostatic interactions. The set of parameters consisting of equilibrium bond lengths, bond angles, partial charge values, force constants and van der Waals parameters are collectively known as a force field. Different implementations of molecular mechanics use different mathematical expressions and different parameters for the potential function. The common force fields in use today have been developed by using high level quantum calculations and/or fitting to experimental data. The technique known as energy minimization is used to find positions of zero gradient for all atoms, in other words, a local energy minimum. Lower energy states are more stable and are commonly investigated because of their role in chemical and biological processes. A molecular dynamics simulation, on the other hand, computes the behaviour of a system as a function of time. It involves solving Newton's laws of motion, principally the second law, $\mathbf{F} = m\mathbf{a}$. Integration of Newton's laws of motion, using different integration algorithms, leads to atomic trajectories in space and time. The force on an atom is defined as the negative gradient of the potential energy function. The energy minimization technique is useful for obtaining a static picture for comparing between states of similar systems, while molecular dynamics provides information about the dynamic processes with the intrinsic inclusion of temperature effects.

Variables

Molecules can be modeled either in vacuum or in the presence of a solvent such as water. Simulations of systems in vacuum are referred to as *gas-phase*simulations, while those that include the presence of solvent molecules are referred to as *explicit solvent* simulations. In another type of simulation, the effect of solvent is estimated using an empirical mathematical expression; these are known as *implicit solvation* simulations.

Coordinate Representations

Most force fields are distance-dependent, making the most convenient expression for these Cartesian coordinates. Yet the comparatively rigid nature of bonds which occur

between specific atoms, and in essence, defines what we mean by the molecule itself, make an internal coordinate system the most logical representation. In some fields the IC representation (bond length, angle between bonds, and twist angle of the bond as shown in the figure) is known as the Z-matrix or torsion angle representation. Unfortunately, continuous motions in Cartesian space often require discontinuous angular branches in internal coordinates making it relatively hard to work with force fields in the internal coordinate representation and conversely a simple displacement of an atom in Cartesian space may not be a straight line trajectory due to the prohibitions of the interconnected bonds. Thus, it is very common for computational optimization programs to flip back and forth between representations during their iterations; this can dominate the calculation time of the potential itself and in long chain molecules introduce cumulative numerical inaccuracy. While all conversion algorithms produce mathematically identical results, they differ in speed and numerical accuracy.[1] Currently, the fastest and most accurate torsion to Cartesian conversion is the Natural Extension Reference Frame (NERF) method.

Applications

Molecular modelling methods are now routinely used to investigate the structure, dynamics, surface properties and thermodynamics of inorganic, biological and polymeric systems. The types of biological activity that have been investigated using molecular modelling include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes.

PROTEIN ENGINEERING

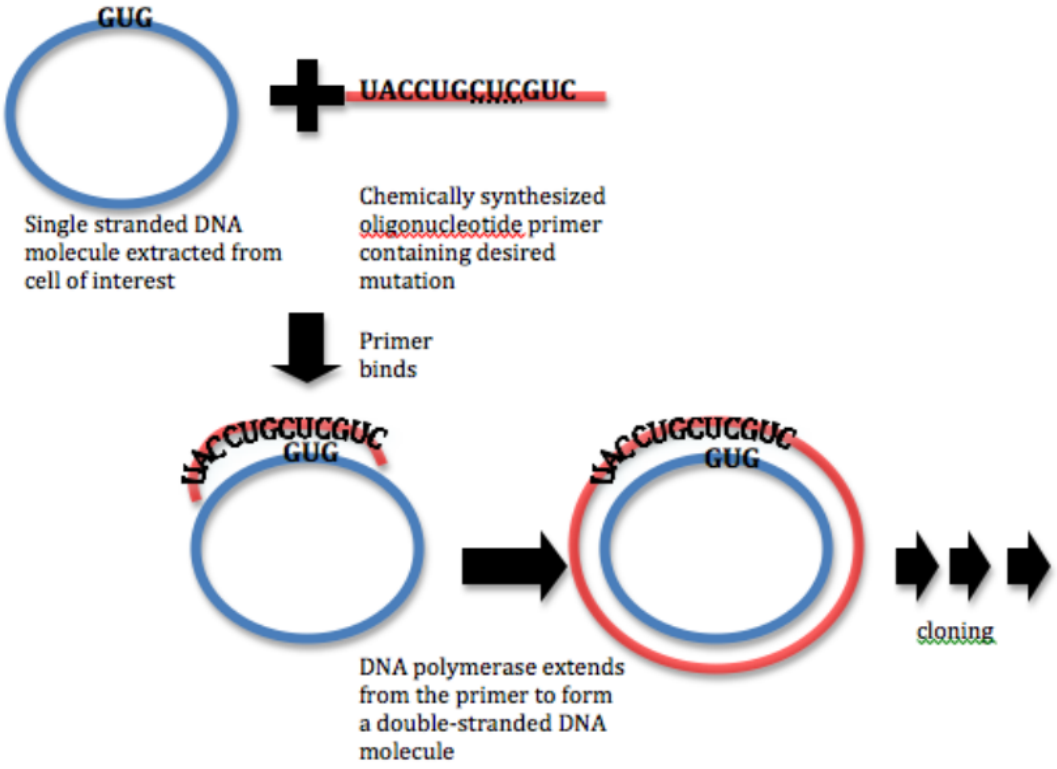
SITE-DIRECTED MUTAGENESIS USED FOR PROTEIN STUDIES

Mutagenesis is a broad term that is defined as the alteration of the genetic material of an organism in a stable manner.

Site-directed mutagenesis is when the amino acid sequence of a given enzyme molecule or other protein may be altered by deliberately and precisely mutating the cloned gene encoding that molecule. It is a very useful technique that can be used in the study of protein function, the identification of enzymatic active sites, and the design of novel proteins. With this technique, it is possible to exchange a single amino acid in the sequence of a protein for another amino acid with different chemical properties. In this way, the function of the specific amino acid at this site can be examined. The basic protocol for this process was developed by Michael Smith, who was awarded the Nobel Prize in Chemistry in 1993 "for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies"

In order to carry out site-directed mutagenisis, a DNA primer must be designed at the site of interest. The primer should contain the necessary nucleotide differences in order to affect the change in the protein sequence. For example, consider the case where a protein sequence reads Tyr-Leu-His-Val, corresponding to a genetic sequence of UACCUGCACGUC. If the experimentalist intends to mutate the histidine residue to a leucine residue, they might design a primer with the sequence UACCUGCUCGUC. This primer is then hybridized to the complementary single stranded DNA molecule and extended using a DNA polymerase. A mutated double stranded DNA molecule encoding the (mutated) protein is obtained in this way, and this DNA molecule is cloned into a host cell. Host cells are allowed to grow and mutants are selected. In this way, proteins that are generated by this DNA sequence will contain the desired mutation. A similar

procedure can be carried out with the use of PCR (PCR site-directed mutagenisis).



An example of site-directed mutagenesis: catalytic triad

Site-Directed Mutagenesis is a method used to dissect the amount of catalytic power by each of the catalytic triad in a enzyme. This is done by converting each of the triad into a common amino acid and measure the catalytic power differences.

For example, in subtilisin, the catalytic triad that is studied by this method are aspartic acid 32, histidine 64, and serine 221. Each of the amino acids in the triad will be converted into alanine and thus, the ability of the mutant enzyme to cleave a substrate is examined. The result by this method shows that serine 221 into alanine reduces the catalytic power as much as histidine 64 into alanine. The value of the k_{cat} for serine 221 and histidine 64 becomes one-millionth of its original value. As for aspartic acid 32, the catalytic power is reduced but not as much as for serine 221 and histidine 64. The k_{cat} value is 0.005% of the original enzyme. Thus, the result of using the site-directed mutagenesis shows that the serine-histidine pair makes a nucleophile that has the capability to attack the carbonyl carbon atom in the peptide bond. Site-directed mutagenesis can be used to change particular base pairs in a piece of DNA. There are a

number of methods for achieving this. The approach described here is adapted from the Stratagene site-directed mutagenesis kit, the manual can be found here. Even when using a kit it will be necessary to design primers that are suitable for the specific changes you want to make to your DNA. Most of the contents of the kit can be found in your favorite labs stocks so you may not need to buy the kit itself. If you have problems with this procedure, you can try 'Round-the-horn site-directed mutagenesis which uses PCR to amplify the desired mutant product.

Mutagenesis has been used in terms of DNA recombinant technique.

The **Protein Data Bank** (**PDB**) is a crystallographic database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography,NMR spectroscopy, or, increasingly, cryo-electron microscopy, and submitted by biologists and biochemists from around the world, are freely accessible on the Internet via the websites of its member organisations (PDBe,[1]PDBj,[2] and RCSB[3]). The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

The PDB is a key resource in areas of structural biology, such as structural genomics. Most major scientific journals, and some funding agencies, now require scientists to submit their structure data to the PDB. Many other databases use protein structures deposited in the PDB. For example, SCOP and CATH classify protein structures, while PDBsum provides a graphic overview of PDB entries using information from other sources, such as Gene ontology