

Lecture 22: Protein Engineering

Proteins have important role in physiological processes and they are involved in movement, catalysis, recognition, regulation etc. Moreover, proteins also have several therapeutical and industrial applications. Advances in Molecular Biology have enabled us to manipulate DNA and express a foreign gene in other organism (heterologous expression). This has made advancement in the process of making changes in proteins at genetic level. Proteins are not always optimized for their properties for various applications and usefulness of a protein may be limited by low stability and /or undesired substrate specificity. Protein Engineering is the process of developing proteins with desired function by manipulating stability and specificity of a protein. There are two main approaches for protein engineering, rational design and directed evolution (irrational design). In case of rational design, knowledge of the structure and function of the protein is taken into consideration and a rational gene mutation is planned (Fig.1). Mostly, this is done by making rationally designed changes in the gene of the protein cloned in expression vector of heterologous expression. The production of protein molecules is altered by site directed or site-specific mutagenesis of their genes. However, in some cases protein structure is not available and directed evolution method is required. In this method, random changes (mutation) are done in the protein and a mutant form with desired properties is chosen.

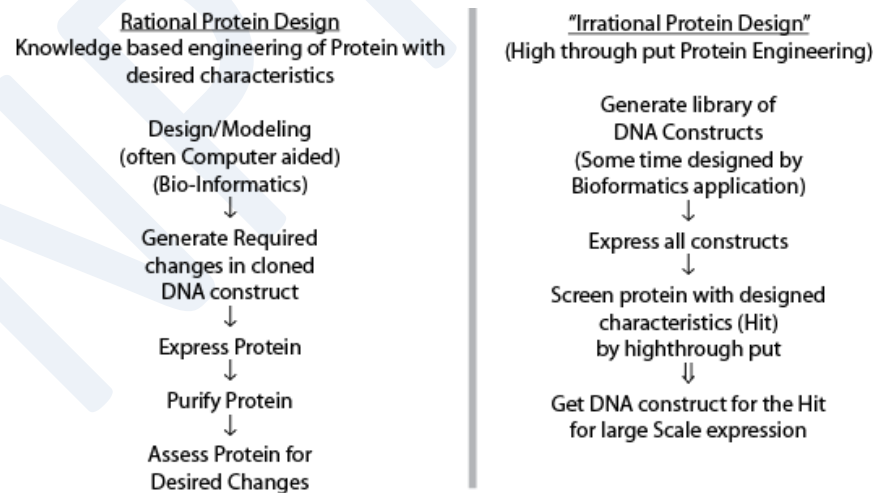


Figure 1: Two different approaches of protein engineering.

One very simple example of rational protein design: Designing an inactive form of Pancreatic ribonuclease A

Pancreatic ribonuclease A is an enzyme comprising of 124 amino acids that cleaves the covalent bonds that join ribonucleic acids (RNA). From available literature about the protein we know that histidine at position 119 is necessary for catalysis. Thus, if at position 119 in the sequence the naturally occurring histidine is replaced with an alanine, the mutant protein is referred to as a histidine 119 → alanine (H119A) mutant of ribonuclease A. This mutant protein is expected to have little or no biological activity, because histidine 119 is important for catalytic activity. A simple example of mutation is shown in Fig. 2.

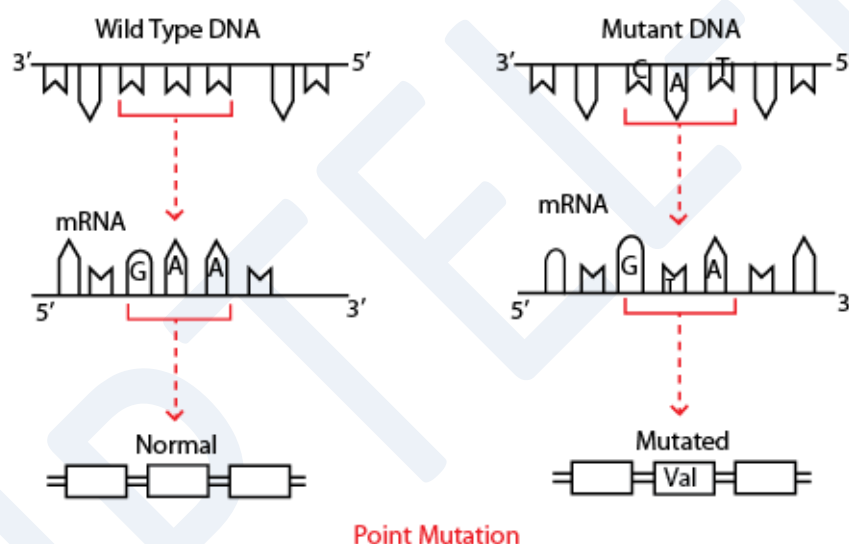
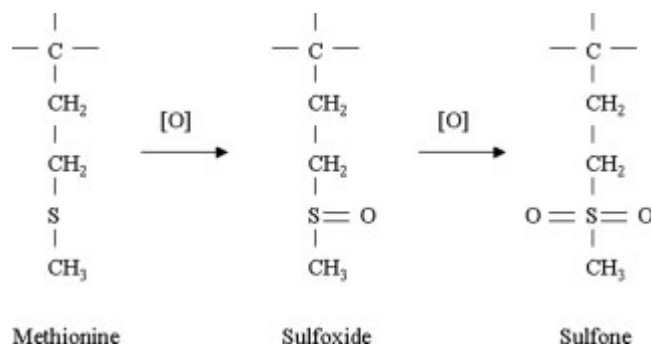


Figure 2: Diagrammatic representation of a designed mutation in protein at DNA level.

Another example of rational protein design: A very common example of protein engineering is engineering of enzyme subtilisin, a protease (a protein-digesting enzyme). To improve the efficiency of laundry detergents, subtilisin is added in the detergent. However, subtilisin is inactivated by bleach. Experimental and structural analysis revealed that this inactivation was due to oxidation of the amino acid methionine at position 22 of the subtilisin molecule. Using site-directed mutagenesis technique the subtilisin gene in *E. coli* was mutated and methionine was changed by alanine. This engineered subtilisin showed high activity and stability, and now many laundry detergents contain cloned, engineered subtilisin.



Methionine is immediately adjacent to a catalytic serine residue and therefore is located at a particularly sensitive structural position. Methionine oxidation results in increased side chain bulk or perhaps even the introduction of strongly electronegative oxygen atom(s) in the immediate vicinity of the active site negatively effects the catalytic activity.

For several other case studies about rational protein designing students may read following article.

© 2007 by The International Union of Biochemistry and Molecular Biology

BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 35, No. 1, pp. 2-8, 2007

Articles

Protein Engineering

CASE STUDIES OF COMMERCIALIZED ENGINEERED PRODUCTS

Received for publication, February 1, 2006, and in revised form, August 11, 2006

Gary Walsh†

From the Industrial Biochemistry Program, Chemical and Environmental Sciences Department, University of Limerick, Limerick City, Ireland

Programs in biochemistry invariably encompass the principles of protein engineering. Students often display increased understanding and enthusiasm when theoretical concepts are underpinned by practical example. Herein are presented five case studies, each focusing upon a commercial protein product engineered to enhance its application-relevant functionality. The case studies may be incorporated into a standard lecture series detailing protein engineering or may form the basis of a class-led post-lecture series discussion.

Keywords: Protein engineering, site directed mutagenesis, biochemistry education.

PROTEIN ENGINEERING BY DIRECTED EVOLUTION

Let us assume a condition where a researcher has to design a protein which binds best with a given receptor or biomolecule. Let us see how directed evolution can help in this designing.

Phage display: George Smith pioneered phage-display technology in 1985. It was reported that a phage displaying the foreign peptides were able to infect bacteria in the same manner as wild type and fusion protein was functional. Phage display is a powerful method for engineering proteins with desired binding specificities. In this method M13, *Escherichia coli*-specific filamentous bacteriophage (a virus infecting bacteria) is used. Gene fragments encoding polypeptides or a peptide library are fused to M13 coat protein genes. This fusion protein becomes part of the capsid and the heterologous protein is displayed on the surfaces of the phage. M13 filamentous phage, contain a circular single stranded DNA genome. The genome encodes 10 proteins, 5 of which are virion structural proteins. The genome is enclosed in a protein coat encoded predominantly by 2700 copies of gene protein VIII (gpVIII) which constitute the major coat protein.

At the 'tail' end of the phage particle 4-5 copies of the gene VII protein (gpVII) and 4-5 copies of the gene IX protein (gpIX) are found and they are involved in initiating phage assembly and maintaining the stability of the viral particle. At the other end of the phage particle are 3-5 copies of the gene III protein (gpIII) which is required for infection of the host cell and 4-5 copies of the gene VI protein (gpVI) involved in the termination of the viral assembly process. Although gpIII is used for making fusion construct but recently gpVII and gpIX have also been shown to tolerate fusions at their amino termini (Fig. 3)

A large library of mutants for a given protein gene is made. In these cases, mutagenesis is frequently done using error-prone PCR. This incorporates random mutation in the gene. The peptide display libraries are cloned directly into the viral genome so that all five copies of gpIII display the peptide fusion. Phage displaying the desired sequence is finally selected (Fig. 4).

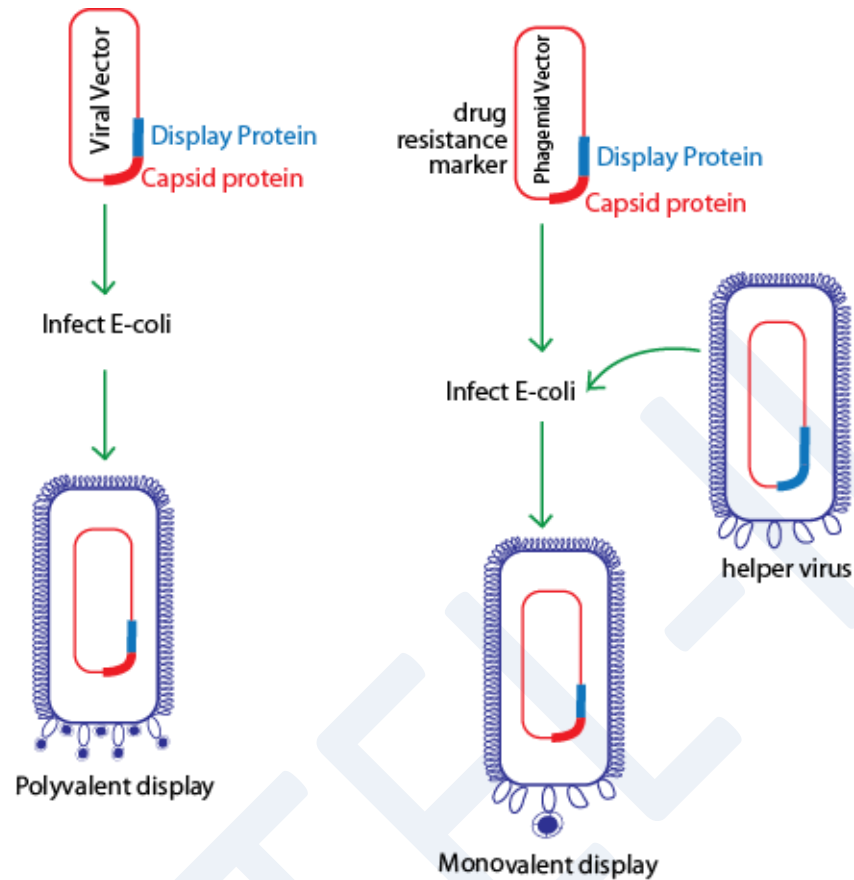


Figure 3: Proteins fused to capsid proteins can be displayed in either of the two formats. Protein can be cloned directly into a viral vector as a fusion to a capsid protein resulting in every copy of the capsid protein displaying the fusion (polyvalency). In case of polyvalent display all gpIII are conjugated with peptide and this results in occasional loss of infectivity. Alternatively, the protein fusion can be constructed in a phagemid vector that carries a copy of the viral capsid gene. In this case, helper phages are mixed at the time of infection. As capsid protein also comes from helper phage (during viral assembly in *E.coli*), infectivity of phage displaying peptide is not compromised. However, valency of display of peptide is decreased. This may have some disadvantages at the time of selection. The valency of display depends on ratio of phage containing phagemid vector and helper phage at the time of infection. Since the helper phage do not carry a drug resistance marker, they are lost upon subsequent selection for drug resistance.

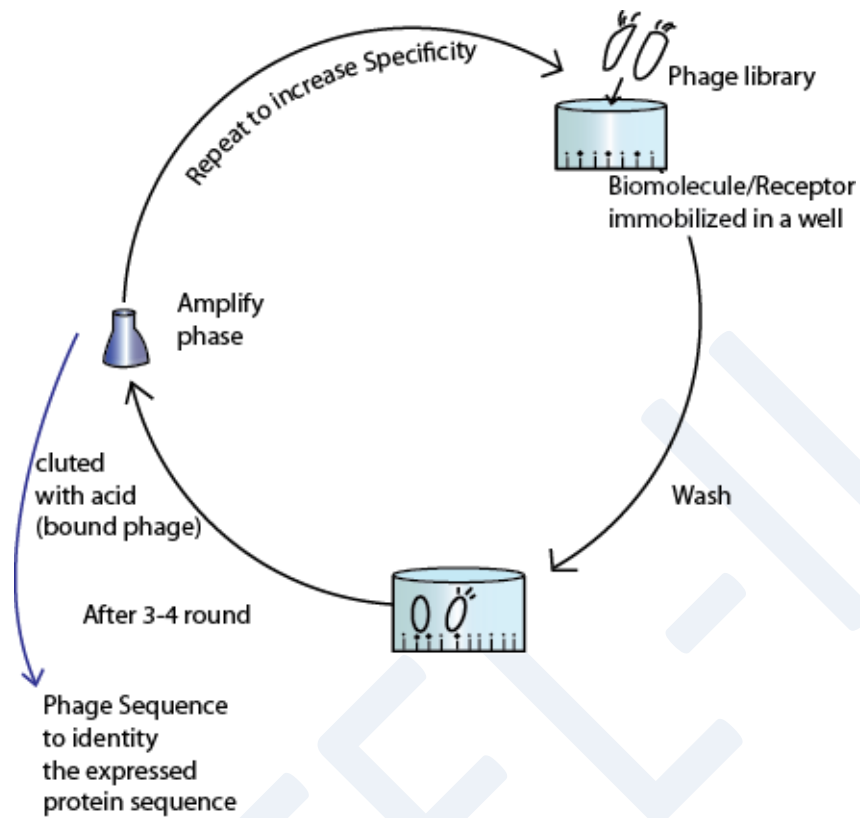


Figure 4. Selection of desired protein (Biopanning). Bimolecule /receptor against which a protein/peptide is being designed was immobilized in the wells of a microtitre plate and probed with the phage peptide library. Unbound phages were removed by washing and then the bound phage were eluted with acid and amplified in liquid culture. The process was repeated 3-4 times and finally the phage was sequenced to identify the expressed protein sequence with high affinity. The process by which phage displaying desired protein sequence is enriched is called biopanning

Further reading and highly recommended articles

Combinatorial Chemistry & High Throughput Screening, 2002, 5, 1-14 1

Random Sequence Libraries Displayed on Phage: Identification of Biologically Important Molecules

Christopher G. Adda^{1,3}, Robin F. Anders^{1,3}, Leann Tilley^{1,2} and Michael Foley^{1,2*}

Department of Biochemistry, La Trobe University¹, Bundoora, 3083, Victoria, Australia, the Cooperative Research Centre for Diagnostic Technologies², and the Cooperative Research Centre for Vaccine Technology³

Abstract: Phage display has become a widely used tool for the identification of proteins or peptides with affinity for a variety of biomolecules. The versatility, simplicity and cost effectiveness of this application has pervaded a wide variety of research areas. Although not without its limitations, phage display has provided a convenient methodology for obtaining ligands to study the function, structure and diagnostic or therapeutic potential of various macromolecules. This review highlights some recent research employing this technology that serves to illustrate its utility in various research and clinical applications.

Chem. Rev. **2001**, 101, 3205–3218 3205

Protein Design and Phage Display


Ronald H. Hoess*

Dupont Pharmaceuticals Company, Experimental Station E336/205, Wilmington, Delaware 19880

Received February 2, 2001

Contents

I. Introduction	3205
II. Phage Display	3206
A. Biology of Filamentous Phage	3206
B. Basic Display Formats	3206
C. Library Construction	3209
D. Affinity Selection Methodology	3209
III. Selection Based on Binding	3209
A. Affinity Maturation	3209
B. DNA-Binding Proteins	3210
C. Scaffold Proteins	3211
IV. Phage Display and Folded Proteins	3211
A. Understanding Protein Folding Using Phage Display	3211



Ronald H. Hoess was born in Philadelphia, Pennsylvania, in 1948. He

Other Methods of Directed Evolution: The other methods of directed evolution are error-prone PCR and DNA shuffling. The gene of protein is mutated using these two methods and **MUTANT WITH DESIRED PROPERTIES IS SELECTED AND AMPLIFIED** using appropriate methods as we have seen in the case of phage display technology.

Error Prone PCR: We shall discuss about polymerase chain reaction (PCR) in detailed during module 7 of the course. Error prone PCR is a variant of PCR. Error prone PCR is a technique used to generate randomized genomic libraries. It allows the initiation of DNA amplification, starting with tiny amounts of parent molecule, and produces considerable amounts of the mutated gene. The working principle of this technique is based on the ability of Taq polymerase to anneal incompatible base-pairs to each other, during amplification, under imperfect PCR conditions. Under these conditions, the polymerase makes mistakes in the base pairing during DNA synthesis which inculcates errors in the newly synthesized complementary DNA strand. The frequency and number of errors introduced into the sequence can be regulated by carefully controlling the buffer composition. For proper functioning of this technique, it is important to use a Taq DNA polymerase which lacks proof-reading ability, because use of a proof-reading DNA polymerase in an error prone PCR reaction will result in the automatic correction of the mismatched nucleotides, therefore any mutations that were introduced during the reaction will be lost. This method has proven useful not only for generation of randomized libraries of nucleotide sequences, but also for the introduction of mutations during the expression and screening process in the mutagenesis step.

PCR



dCTP, dTTP
dGTP, dATP
 Mg^{2+}

ERROR PRONE PCR



dCTP, dTTP ↑
dGTP, dATP ↓
 Mg^{2+} ↑
 Mn^{2+}

Contrast of External Conditions
between PCR and error prone PCR

DNA shuffling: DNA shuffling is also known as sexual PCR. Before beginning with DNA shuffling, a pool of closely related molecules, with different point mutations, is prepared either through error-prone PCR or other mutation techniques such as oligonucleotide-directed mutagenesis.

Various steps involved in DNA shuffling are as follows:

- Step 1: Breakage of molecules into random fragments using DNase I which can randomly create nicks along each strand of a DNA molecule.
- Step2: Sampling of fragments of lengths within a certain range.
- Step 3: Further these sampled fragments go through PCR *without added primers*. There are three steps in the PCR process without primers.
 - Step 3A: Denaturing of double-stranded fragments by increasing the temperature, so that double stranded DNA are separated completely into single- stranded ones.
 - Step 3B: Annealing by lowering the temperature, so that single-stranded fragments anneal to other fragments overlapping by a certain number of bases that are complementary at the overlapping region. Once annealing step is over, homologous templates prime each other to form 5' and 3' overhangs.
 - Step 3C: Polymerase extension by increasing the temperature optimum for DNA polymerase to extend the 5' overhangs using the other annealed strand as template. The 3' overhangs are not changed as DNA polymerase can only extend from a 5' end.
 - The three steps of denaturation, annealing and polymerase extension are repeated for multiple cycles.

Diagrammatic representation of DNA shuffling is shown in Figure 5.

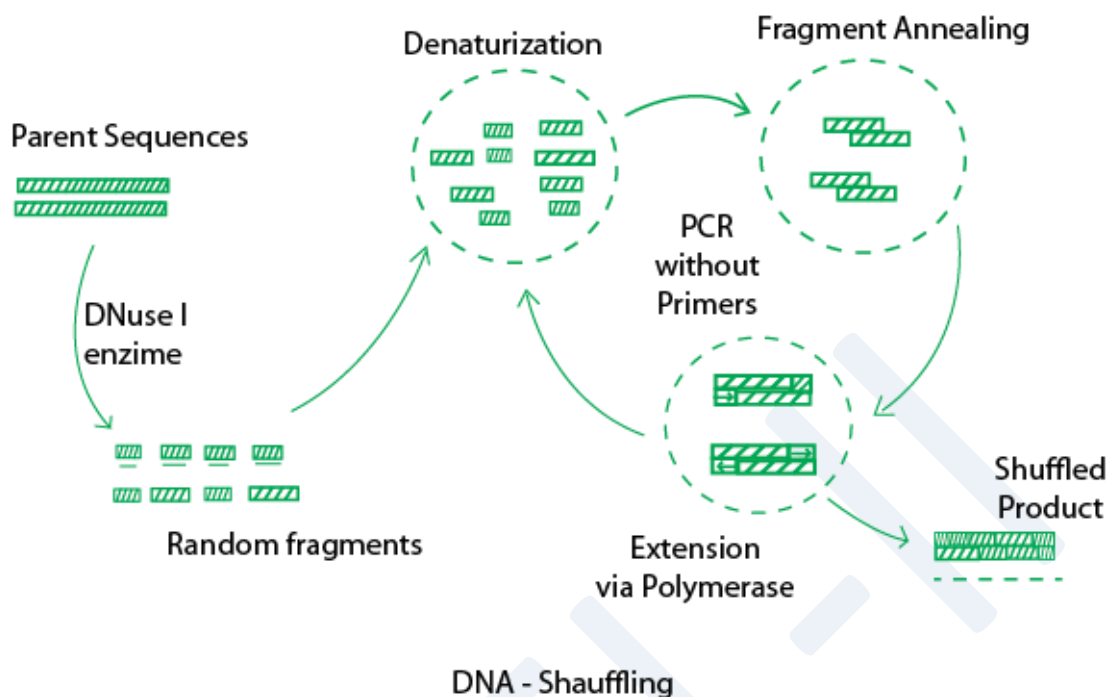


Figure 5: Diagrammatic representation of DNA shuffling

In each PCR cycle repeat the average fragment length increases. After many cycles of PCR without adding primers, molecules of original size are expected. Recombination occurs when a template from one molecule primes a fragment from another molecule.

The purpose of DNA shuffling is to recombine beneficial mutations from different molecules, and obtain molecules with even increased function. Further DNA shuffling improves the search of local fold space by means of a random yet correlated combination of homologous coding fragments that contain limited numbers of beneficial amino acid substitutions.