International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 301-307
Published Online June 2020 in IJEAST (http://www.ijeast.com)

A REVIEW ON - APPROACHES, CHALLENGES, APPLICATIONS, CURRENT SCENARIO OF PROTEIN RATIONAL DESIGNING

Mansi Garg
Department of Pharmacy
Indore Institute of Pharmacy opp. to IIM Indore,
M.P., India

Yashasvi Sharma
Department of pharmacy
Indore Institute of Pharmacy opp. to IIM Indore,
M.P., India

Abstract— **Protein engineering is a very helpful approach in the branch of medicinal science and proteins are the complex biomolecules which carry out a variety of functions in our body. Protein engineering is useful tool for researchers because it modifies the protein through replacement, addition and removal of amino acid. Rational protein designing and directed evolution are two important strategies of protein engineering. Rational protein designing is a significant approach in protein engineering which require previous information about three-dimensional structure as well as the biophysical records and the function of the protein. Directed evolution is also recognized as irrational protein design that mimics the procedure of natural selection to modify protein structure or nucleic acids to create better results over rational design. In this review article, we are discussing about various approaches of rational protein designing, challenges in rational protein designing and applications of protein engineering. the approaches of rational protein design we are discuss here are: fusion protein, split protein, chemical modification, bioinformatics, active site modification, disulphide bonds, cyclization, and computer modeling.**

Keywords— **protein, protein engineering, rational protein designing, directed evolution, approaches of rational protein designing, and challenges in rational protein designing.**

I. INTRODUCTION

A. What is protein?

Proteins are the richest organic molecules that is present in all the living organisms. In 1839, Dutch chemist **G.J. Mulder** was first to describe proteins. The term protein is derived from a Greek word **"proteions"** meaning first place.

Proteins are define as the large, complex biomolecules, or macromolecules, which are made up of amino acids sequence. In proteins there are one or more, long or complex sequence of amino acids and the amino acids are bind each other by a peptide bond. They do the majority of the work in cells.

Amino acids are organic compounds. It contains mainly two functional groups as its name indicates **amino, amino** means**.** [-NH2] group is basic in nature and the acid means carboxyl group [-COOH] is acidic in nature. About 300 amino acids are occur in nature but out of 300 only 20 of them present in proteins.

Functions of proteins are:

- Maintenance, growth and repairing of body tissues and cells
- Energy source for growth of tissue and cells
- Catalyze the metabolic reaction in body
- Replicate the DNA
- Responding to stimuli
- Store other substances in the organism.
- Providing structure of cells and organisms.
- Transport one or more molecules from one place to other.

Protein structures

Proteins contains n-number of amino acids. Hence, it is also known as polymers of amino acids. Individual amino acids are bind together by a bond known as peptide bond to form the linear polypeptide chain. This linear polypeptide chains folded into specific structural conformations or simply structure. A protein can have up to four levels of structural conformations. They are designated as:

1. Primary structure
2. Secondary structure
3. Tertiary structure
4. Quaternary structure

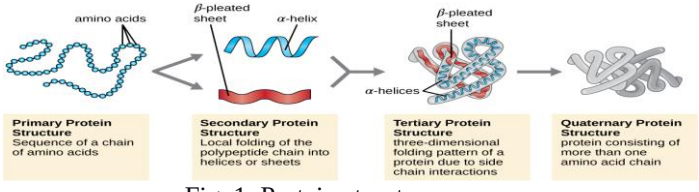

Fig. 1. Protein structures

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 301-307
Published Online June 2020 in IJEAST (http://www.ijeast.com)

***B.* What is protein engineering?**

Protein engineering is the process which is used by a researcher for modifying a amino acid sequence present in protein through replacement, addition, or removal of amino acids sequence, intending to achieve modified protein that is more appropriate for a specific function or reason than the unchanged protein. Protein engineering main focus on application sets protein engineering which is separate from the term "targeted mutagenesis". Targeted mutagenesis, is a method in which a particular position was targeted within a gene sequence is changed. Such modification can be performed for engineering purposes, as in protein engineering, or for examining the outcome of a specific alteration in a gene.

Objective of protein engineering:

- To produce a better quality of enzyme to catalyze the manufacturing of high-value definite chemicals.
- To produce an enzyme in larger quantities.
- To produce biological compounds (include synthetic peptide, storage protein, and synthetic drugs) superior to a natural one.

Protein engineering is mainly performed by two important strategies. They are:

1. Rational protein designing, and
2. Directed evolution

Directed evolution: In directed evolution, the random mutagenesis is implement to protein which alters the properties of an enzyme or a protein; a selection process is used to select a variants that have the characteristics which are required. Further, the rounds of mutation and the procedure of selection has been done.

There are various libraries of variants which are used in in the processing of directed evolution. The libraries of variants avoids the difficulty of choosing variants of desired property.

Directed evolution is also known as irrational protein design that mimics the process of natural selection to evolve proteins or nucleic acids to. Its main focuse on specific molecular properties and it does not create a new molecule.

Rational protein design: it is a very useful strategy of protein engineering, which is used to make better-quality protein molecules based on the 3-D arrangement and the correlation between structure and function, which has developed over the past years as a branch of protein science.

Firstly, the rational protein designing describes about the 3-D structure by using x-ray crystallography process which also provide us graphical and mathematical representation on a computer. Through computer representation it is easy to made predictions, preferentially in the region of the result of mutations on structure-based properties.

It is a easy modeling task but probably the most largely used to visualize the protein structure in all its spatial detail.

The final but most important part of rational design is the capacity to make variants of the native protein by (r) DNA techniques. The whole interactive scheme is represented in the below figure. The rDNA technique which is most commonly used is site-directed mutagenesis, where one amino acid is replaced by other at a particular place.

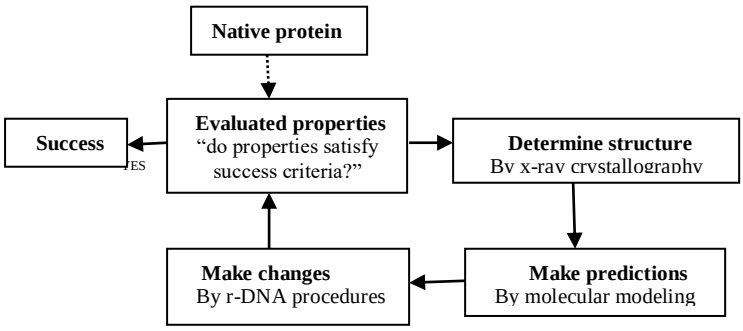

FIG. 2. RATIONAL DESIGN CYCLE

Once the modified protein was made completely then it will be purified and evaluated to visualize the required property is achieved. If the required property is not achieved, then the information which we achieved in the evaluation process is used in a second round of crystallography, modeling, and mutagenesis.

From the first round the structure of variant was determined and through molecular modeling it is ued to explain the results of initial choice and also give information about recommended change for the second round of mutagenesis. For example, if the targeted property was developed in thermal stability, then the native structure may demonstrate undesirable amino acid interactions, which could be decreased by site-directed mutagenesis. A variant crystal structure would show whether the intended structure outcome in the original site were achieved and possibly illustrate new interactions, which were introduced as an outcome of the first mutation.

Rational design was mostly used to generate hybrid proteins that are a mixture of pre-existing, but distinct protein domain. Since the majority of protein domains that confer catalytic or other properties can fold independently.

II. APPROACHES TO RATIONAL DESIGNING OF PROTEIN

A. ***Fused protein:*** **-** In a fused protein a protein is consisting of at least two domain which is enclosed by separate genes which are joined together. So they can be transcribed and translated as a single protein unit.

According to one of the studies, leprosy was registered worldwide as a fatal disease. In leprosy mycobacterium leprae the causative agent is transmitted without being detected. But after studies, several recombinant antigens are identified specifically in a leprosy patient. And then chimeric fusion proteins are used which possess the antibody binding

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 301-307
Published Online June 2020 in IJEAST (http://www.ijeast.com)

properties and this enhances the test development and enhances the production cost.

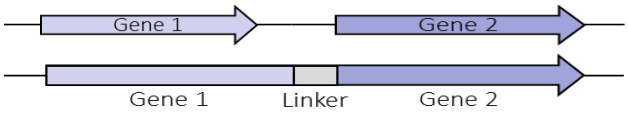

Fig. 3.

B. ***Split protein:*** - Many proteins are split through rational deigning and that the proteins are reassembled by covalent binding which then called functional protein.

In recent studies, Cas9 which is an RNA guided DNA endonuclease found in clustered regulated interspaced short palindromic repeats (CRISPR) is involved. According to this studies Streptococcus pyrogens Cas9 alone or with single targeted DNA and guide RNA revealed a bilobed protein that produces a change in binded RNA and DNA to study the molecular determinants of rearrangement of the molecule. Cas9 is split in alpha-helical lobe and nucleus polypeptide and it is recruited by SgRNA which recapitulates in single length Cas9 which catalyzes site-specific DNA change in Cas9.

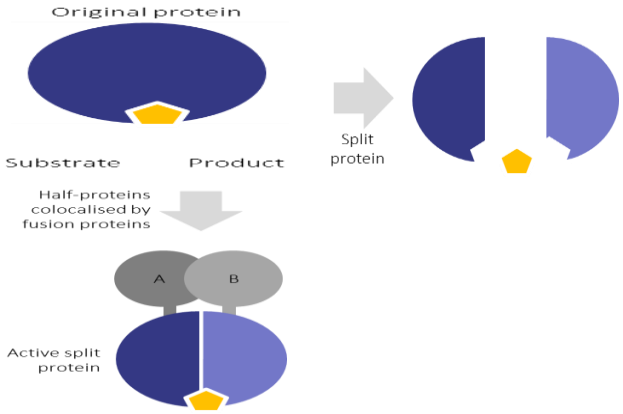

Fig. 4.

C. ***Chemical modification:***-Modification of proteins are problem-solving techniques in research and technology. Generally, this modification is in the reactive chains of protein and is predominantly reduction, oxidation, nucleophilic and electrophilic substitution. Also involves deterioration like beta elimination, peptide bond scission, etc. they are modified to get new improved product and the modified product are more specific and potent to their work.

For example - one of the protein modifications is the acetylation of a lysine residue. It is a co-translational and posttranslational process.

Histone acetylation and DE acetylation had an important role in gene regulation. The acetylation and DE acetylation on lysine residue of histone protein at Nitrogen - terminals. In one of the study increase acetylation of the cytoskeleton protein, especially microtubule protein in response to reactive oxygen .thus, occur clampdown of SIRT2 aggravate the mitochondrial dysfunctional.

D. ***Bioinformatics:*** - as the traditional method of rational designing stays that random protein engineering as opposed to traditional engineered. The bioinformatics approach is used to identify specific sequence changes that alter the functional property of an enzyme within a broad class.

For example:-specifically Bayesian sequence – based-algorithms PROBE and classifier to identify a strand -turn-strand motif that increases the thermophillicity in the member of a serine protease. 16 amino acid was replaced in mesophilic subtilisin E with bioinformatics generated thermophilic model sequence the M.P. of subtilin E increased 13 degrees C and wild type subtilin was inactive at 90 degrees C.

E. ***Active site modification:*** - the active site is the region of any molecule where the substrate is bind. In the case of protein enzyme is the molecule and active site of the enzyme is where the substrate is a bind for undergoing a chemical reaction.

The general objective of modification of active site

(1) Identification of amino acid and ligand-binding site of regulatory protein along with the identification of immunoglobulin.

(2) Introduction of the physiochemical receptor.

(3) Labeling to isolate from a multi-component system.

F. ***Disulfide bonds:*** - Some proteins do not have disulfide bonds. So enhancing one particular activity in the protein the disulfide bond is induced protein by rational designing which enhances the stability.

For example 1, 3-1,4 beta-glucanase is one of the significant biocatalysts in the fermenting industry and animal food industry. Which cuts thermostability and enhances the bid.

The stability of the beta-glucanase is enhanced by introducing disulfide in protein structure. Protein spatial configuration was analyzed and exclude residue pair which undesirably conflicted with structure.

G. ***Cyclization:*** - cyclization of protein is done to establish the stability of the protein.

In the rational deigning of protein. In carcinogenesis, a hippo pathway is involved in this YAP regulatory protein is central regulation and this YAP pathway has a function to interact with the TEA domain transcription factor. And in cancer, suppression of oncogenicity is by the pharmacological disruption of YAP – TEA complex and it is a potential therapeutic strategy.

In this the crystal structure is examined and two hotspots are examined rationally in this complex interface and this hotspot termed as PS-1 and PS-2 and they are self-inhibitory peptide.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 301-307
Published Online June 2020 in IJEAST (http://www.ijeast.com)

From dynamic stimulation, energetic analysis, and flurosense polarization the intrinsic disorder and the binding affinity is defined of these two hotspots and then cyclization is done of this hotspot via sulfide bond.

H. ***Computer modeling:*** - De novo synthesis of protein method which is only a few decades old but it gives very exciting results and development and it gives a very significant impact in the field of biotechnology and chemical biology. In De novo synthesis we designed a completely new model or structure of the protein which has no relation to any old known structure.

So in this computational designing method or computational modeling is very helpful to predict the structure and success of structure and this is very effective in the implementation of an idea. Deigned – based paradigm and alteration – the based paradigm is used to predict the success of structure.

Fig. 5.

III. CHALLENGES IN RATIONAL DEGINING OF PROTIEN

Part of Atomic energies to designing protein structurally
For any of the amino acid sequences, it is necessary to estimate the free energies of protein or the composite of protein and ligand. It is done by using a physical model of atomic interaction and empirical data. There are 3 types of models are defined (a) statically effective energy functions or based on the information that is mined from protein, its structure, and sequence databank. (b) Empirical effective energy functions it includes an accounting of stability of protein obtained empirically. In this, each term is optimized for protein design by using a set of empirical data collection (c) Physical effective energy roles built on the first- principle picture of interatomic interaction in protein. The above three models use the same parameterization irrespective of the target assembly and provide a procedural advance in force field parameterization. The physical principles governs both protein properties and small molecules. Models of equally folded and unfolded conditions are needed as the free energy takes into the explanation of energetic competition between folded and unfolded conditions.

[1] In terms of energy - The potential energy functions in computational protein design are used as it provides applicable and exact to detention the characteristics of atomic structure linked to protein structure and its function and also have a rapid computational implementation.

According to Gordon et al. [1999] & Harbury [2007] energies that are related to bonding energies and covalent bonding are not used in deigning. Internal coordinates like bond angles, bond length which remain unchanged in the folding process

[2] *Folded state modeling* - According to the algorithms of protein design 3-D structures are used as input and analyze the sequence which has to be used. By this process, we can avoid the complexity of structure space by using a motionless sequence. Folded state modeling faces the complexity of side-chain conformational space. So we solve this by using rotamer libraries in which side chain is well-defined by the value of inner dihedral angles.

[3] *Unfolded protein modeling*- unfolded protein demonstrating wants since the folding free energy takes into explanation the energetic competition among folded and unfolded conditions

For example -on assuming that the denatured state depends on the composition of amino acid and maintains fixed amino acid composition.

[4] *Entropic contribution* - Folding free energy includes the contribution of the entropy of unfolded and folded states. This term could rise from a vibrational degree from bonding interactions, solvation entropy, and conformational entropy. This type of contribution is an extensively disputed topic and this is neglected in close packaging of the residue in the protein.

IV. APPLICATIONS

[1] Food Industry: The kind of food processing enzymes are used in food industries, for example: amylases and lipase, through using (r)DNA technology and protein engineering methodologies the belongings of enzymes was enhanced. The elimination of innate genes encoding extracellular proteases, for example, amplified enzyme manufacture yields of microbial masses.

a number of huge groups of enzymes like proteases, amylases, and lipases are imperative for both nutrients and detergent trades, as they have an extensive series of industrial applications.

• Proteases are used in numerous applications of the food industry concerning low allergenic infants formulas, milk clotting, and flavors. They are also play an important role in the detergent industry by removing protein stains.

• Amylases are also show a central part in together nutrients and detergent industries. They are used in the progression of starch liquefaction that renovates starch into syrup and saccharification which involves manufacture of glucose and maltose of starch as well as in alteration of flour and bread smoothness and size in baking. In the detergent industry, they are cast-off for the exclusion of starch stains.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 301-307
Published Online June 2020 in IJEAST (http://www.ijeast.com)

• Lipases are used in nutriment industry for the strength and acclimatizing of an emulsifier, and also use for giving flavored to cheese. In detergent industry, lipases are used in the exclusion of lipid stains.

[2] Environmental applications: it is also another important field of protein engineering.

•recombinant DNA technology and protein engineering methods are used for creating new microorganisms two which are used to remove environmental toxins and to regulate gene expression to give great catalytic activity under environmental stress conditions, such as the presence of a toxic compound, rational changes were also introduced in rigid protein that will manage catabolic activities and also create new metabolic routes and combinations.

•it is an also useful application for detoxification of organic pollutants by using enzymatic oxidation.

•Petroleum biorefining applications such for example fuel biodisulferization, denitrogenation of fuel, exclusion of heavy metal etc. by using biocatalyst.

[3] Medical application: medical application is also an important and diverse field of protein engineering for researchers.

•The major area of curiosity is cancer treatment studies. The usage of protein engineering in cancer management is enhanced the use of pretargeted radio immunotherapy.

•The another application of protein engineering in cancer management is to change antibodies to target cancer cells.

•Protein engineering also improver pharmacokinetic properties of antibodies and also produce various sizes of antibody variants.

[4] **Biopolymer production:** protein engineering is very useful applications used for biopolymer production. Biopolymer also known as peptide-based biomaterials.

Examples pf biopolymer which is produced by protein engineering are rubber, elstin, lignin, etc.

• Protein engineering technique also used to produce protein domains which can be utilize for the production of new biomaterial for medical and engineering applications.

•The example of protein domain which are made by protein engineering are leucine zipper coiled coil domain, EF-bond domain, etc

[5] Nanobiotechnology: nanobiotechnology applications are also becoming increasingly important in the field of protein engineering.

• proteins, carbohydrates, and lipids are biological macromolecules which are helpful in the synthesis of genetic tissues in aqueous environments and slight physiological conditions, where this biosynthetic process under genetic regulation.

•The additional exciting application of protein engineering in nanobiotechnology is nanowire construction by the use of amyloid fibrils. The amyloid fibrils used as a structural templates for nanowire creation.

[6] Redox proteins and enzymes: the application of protein engineering is likewise useful to generate redox protein and enzyme which are used in biofuel cells, chemical synthesis, and biosensors.

•Protein engineering of redox proteins pointed out the binary region of protein engineering of redox proteins and enymes: novel nucleic acid-based catalyst construction, and intramolecular electron transfer network makeover.

• The another bid of protein engineering is to produce cytochrome P450 biocatalysts which is beneficial for medical, and biotechnological applications.

[7] Other new applications: "Insertional protein engineering" applications are a new technique becoming more considerable, mostly for biosensor studies.

•"Zinc finger protein engineering" is another new strategy that has been developed for gene regulation application and to control gene expression.

•"Virus engineering" is likewise a one of the greatest used field in which the virus units are adapted by protein engineering. Viruses have many important applications in medicine, biotechnology, and nanotechnology

rtional pro They could be used as innovative vaccines, molecular imaging agents, and building blocks for electronic nanodevices or nanomaterials construction.

•"Protein cysteine modifications" are also protein engineering applications in which cysteine alteration in protein has been completed which basis varieties in protein roles, cysteine thiol chemistry has been used for the production of in vitro glycoprotein

V. CURRENT SCENARIO OF PROTEIN RATIONAL DESIGNING

In the past few, there is much advancement in the protein structure prediction and design but along with this there are many challenges that remain the same such as the balance between non-polar and polar interaction, solvation, etc. and this result in the lowering of the success rates of the model. Hybrid approaches are also there in which the water molecule has interacted with the protein but it is still a challenge due to the computational cost and there is also a difficulty in the interaction between antigen and antibody because in this the interface energetics have to be compounded accurately. Currently, there is a need for new approaches for the prediction and designing of flexibility and motion. There is a need for the advancement of machine learning and pattern recognition. If the protein designing and prediction will

continuously grow then it is very beneficial for the field of medicines and biology.
There are many pieces of research and experiments that are going on currently. Like in July 2019 an article was published by IISER Pune. In IISER they are working on the self - assembly of protein in superstructures which has a variety of applications like in gene delivery, vaccine design, and diagnostic. IISER Pune developed a chemical method for designing protein to form a micelle.

VI. CONCLUSION

In this article, we studied the approaches, challenges, and application of protein rational designing. In today's world where the medicines are evolving as a narrow spectrum, instant relief, targeted mode of action, and many other benefits. In this era of modern efforts designing protein rationally is very beneficial and gives positive results in the field of medical research and development hence there are many approaches and challenges of protein rational designing. The main challenge of rational protein designing is the complexity of the structure of the proteins and it made the designing process very difficult. But the challenges are decreased by computational rational designing. Rational protein designing is important for research and development.

VII. REFERENCES

[1] Kuhlman Brian, Bradley Phillip,(2019) " advanced protein structure prediction and design" Published by technologies and techniques nature reviews, molecular cell biology, volume 20, November 2019, page no. 681-697

[2] Korendovych V. Ivan(2018), “Rational and semi-rational protein design”, published by a method in molecular biology (Clifton, N.J.), methods in the boil. 2018; 1685:15-23

[3] Riggs p et.al. (2013), “fusion protein”, published in Brenner’s encyclopedia of genetics, second edition,

[4] Duthie S Malcolm, Hay N Marah et.al.(2009), “rational design and evaluation of multiepitope chimeric fusion protein with the potential for leprosy diagnosis”, published by clinical and vaccine immunology, Dec 19, DOI: 10.1128/CVI.00400-09

[5] Britto S sandanraj et.al.(2018), "Rational design of supramolecular dynamic protein assemblies by using a micelle- assisted activity-based protein labeling technology ", published by European chemical societies publishing, volume 24, issue 60, 12 August 2018.

[6] Ditursi kate Mary, Kwon Seok-Kwon(2006, bioinformatics- driven, rational engineering of protein thermostability", volume 19, issue 11, November 2006, pages 517-524.

[7] Muller M. manuel (2018) , “post- translational modification of protein backbones: unique functions mechanisms, and challenges, 2018 Jan 16; 57(2) : 177-185, DOI:10.1021/acs.biochem.7b00861

[8] Tao liu et.al.(2016), "enhancing protein stability with extended disulfide bonds, PNAS may 24 2016 113(21)5901-5915

[9] Tiwari manish kumar et.al(2012),computational approaches for rational design of proteins with novel functionalities, volume 2 , issue 3, September 2012, https://doi.org/10.5936/csbj.201209002

[10] Maria suarez and Jaramillo Alfonso (2009), "challenges in the computational design of proteins", published by the royal society publishing, 2009 mar 11 doi no. 10.1098/rsif.2008.0508.focus (page no. 477-491).

[11] W.V.Shaw (1987), review on protein engineering: the design, synthesis, and characterization of factitious proteins, published in biochemical jourana.1987 aug 15, doi no. 10.1042/bj246001 (page no. 1-17).

[12] Philip N. Bryan, a research review paper on protein engineering, published by biotechnology advances, volume 5, issue 2, 1987, page221-224.

[13] Martin K.M. Engqvist, Kersten S. Rabe, application of protein engineering and directed evolution in plant research, published by PMC, plant Physiol. 2019Mar; 179(3): 907-917.

[14] Bornscheuer U.T, Poul M., improved biocatalyst by directed evolution and rational protein design, published by pub med, 2001 Apr;5(2):137-43. Doi 10.1016/s1367-5931(00)00182-4.

[15] Peter H. Tobin, David H. Richards et.al, protein engineering: a new frontier for biological therapeutics, published by Curr Drug Metab. 2014; 15(7): 743-756.

[16] Kaur Jaspreet, Sundar Durai et.al, computational protein engineering approaches for effective design of new molecules, in the encyclopedia of bioinformatics and computational biology.

[17] K.M. Helena Nevalainen, Valentino S. Jnr.To, protein engineering in fungal genomics, in applied mycology and biotechnology, 2003.

[18] Marks C.B, Vasser M.N, P. Henzel,.W and Anderson. S, (1986) J Biol Chem. 261, 7115-7118.

[19] Courtney. M, Tallat S, Tessier L.H, Benavente A, Crystal R.G, and Leeoeq J.P (1985) Nature 313, 149-151.

[20] Liao. H, Mckenzie. T and Hageman. R, (1986) Proc Natl Acad Sci USA 83, 576-580.

[21] Chothia C and Lesk A.M., (1986) EMBO J 5, 823-826.

[22] Roberto A. Chica, protein engineering in the 21st century, published by protein science (protein sci.) 2015 Apr; 24(4): 431-433, DOI: 10.1002/pro.2656.

[23] Affholter J. and Arnold F. H. (1999). Engineering a revolution. Chemtech 29(9), 34–39. [An excellent review of directed evolution, and how it is giving rise to a new generation of catalysts for chemical processes; written for those with limited experience in the biological sciences.]

[24] Bull A. T., Bunch A. W., and Robinson G. K. (1999). Biocatalysts for clean industrial products and processes. Current Opinion in Microbiology 2, 246–251. [A good review of biocatalyst opportunities in clean industrial processes.]

[25] Minshull J. and Stemmer W. P. C. (1999). Protein evolution by molecular breeding. Current Opinion in Chemical Biology 3, 284–290. [A good review of DNA shuffling and a broad range of applications.]