PROTEIN ENGINEERING

D. N. Rubingh and R. A. Grayling
Corporate Research Division, The Procter and Gamble Company, USA

Keywords: biocatalysis, bioremediation, biosensor, de novo design, detergency, directed evolution, DNA shuffling, enzyme, fermentation, hybrid enzymes, pharmaceuticals, protease, rational design, starch processing

Contents

Summary

Proteins, either directly or indirectly, are the basis for much of modern biotechnology. Over the past three decades, the ability to modify protein properties for desired benefit has been refined to the point where it is now possible to engineer almost any property for which there exists a predictive screen. In the near future, custom-made enzymes, drugs, and other proteins may be used in every home, hospital, and industry. Both rational protein design approaches, that employ knowledge of protein structure, function, and mechanism, as well as powerful, non-rational approaches termed 'directed

evolution' have been used to generate commercial successes, such as improved enzymes for use in the cleaning and starch-processing industries. Enzymes can be tailored to be more stable, more active, or even perform new functions, and protein drugs can be engineered to be safer and more effective. New protein engineering technologies continue to accelerate the rate of product discovery and development, and many new engineered enzymes and protein drugs are in the pipeline that are poised to have significant commercial impact. It is possible that a significant portion of the twenty-first century global economy will depend on successful application of protein engineering, for continued advances in healthcare, agriculture, and the development of truly sustainable industrial processes.

1. Introduction

The pioneering gene cloning experiments of Herbert Boyer and Stanley Cohen in the early 1970s ushered in an era where proteins could be made and modified as never before. It was not long before the commercial possibilities, as envisioned for example by Robert Swanson and Boyer in 1976, resulted in the application of this technology in companies such as Genentech. The following discussion outlines some of the strategies and commercial applications that have developed in protein engineering over the past 30 year period, and takes a brief look at where the field may be headed in the future.

Protein engineering can be considered a sub-discipline within the broader category of genetic engineering. The distinguishing feature of protein engineering is that the final product is a protein with a modified amino acid sequence, rather than a new (or modified) living organism. Since proteins do not reproduce, many of the concerns (legitimate and otherwise) found in the broader field of genetic engineering (e.g. the current controversy over genetically modified organisms) are not an issue in protein engineering. In this regard, engineered proteins more closely resemble new chemical compounds from non-biological sources, for which concerns around safety and toxicity apply, but which by their very nature are readily biodegradable.

Two strategies have emerged to design proteins to work better or under unusual conditions. The first employs site-directed mutagenesis along with protein structural information to rationally design new or improved function. More recently, this rational approach attempted with some success to create functioning “proteins” *de novo* (“from scratch”). In contrast to this, a non-rational approach called directed evolution employs recombinant DNA techniques to create thousands of possible variants, and then uses high throughput screening methods to rapidly search for the one that offers the best solution. This approach has emerged as a powerful alternative to rational methods, particularly when the relationship between structure and desired function is obscure, but where a representative screen can be developed.

Some of the early commercial successes of protein engineering in improving the performance of enzymes for laundry and the starch processing industry are described in the applications section. Also described are some more recent uses of engineered enzymes and other proteins in pulp bleaching, neutraceuticals and biosensors. In the near future, engineered enzymes are poised to play an important role in chemical processing and bioremediation. These and other emerging applications are discussed in

the final section concerning future directions in protein engineering.

2. Strategies for Protein Engineering\

2.1. Rational Methods

2.1.1. Rational Design

Rational design is a particular strategy in protein engineering, which attempts to create improved protein molecules based on the three-dimensional structure and the relationship between structure and function, which has developed over the years as part of protein science. Because of its roots in protein chemistry, rational design was the earliest approach to protein engineering and still is widely used, either as a stand-alone approach or combined with random mutagenesis or directed evolution when these latter approaches uncover interesting leads.

The earliest discussions of rational design describe an iterative process in which x-ray crystallography provides a three-dimensional structure, which can be represented graphically, and mathematically on a computer. The computer model allows predictions to be made, particularly in the realm of the effect of mutations on structure-based properties.

The simplest modeling task, but perhaps the most widely used, is to employ the computer to visualize the protein in all its spatial detail (by rotation, zooming in on particular residues, highlighting interaction etc.) so that important "what if" questions can be asked and intelligently answered. For example, if one substitutes a lysine residue at position 123 for an alanine, does it fit? Of course such questions can be posed more quantitatively as well, for example using “Free Energy Perturbation Theory” to calculate the energetic cost of replacing one amino acid for another, but only with considerably more effort.

The final, but essential, component of rational design is the ability to make variants of the native protein by recombinant (r) DNA techniques. The entire iterative scheme is indicated in Figure 1. The rDNA method of choice is commonly site-directed mutagenesis, in which one amino acid at a particular location is replaced with another.

Once made and purified, the new protein is evaluated to see if the desired property is achieved. If the desired property is not achieved, the information obtained in the evaluation is used in a second round of crystallography, modeling, and mutagenesis. The structure of the variant from the first round is determined and used to explain, through molecular modeling, the results of the initial choice and to provide a suggested change for the second round of mutagenesis. For example, if the property targeted for improvement was thermal stability, then the native structure might show unfavorable amino acid interactions, which could be alleviated by site-directed mutagenesis. A crystal structure of the variant would demonstrate whether the intended structural results in the original location were achieved, and perhaps show new interactions, which were introduced as a result of the first mutation.

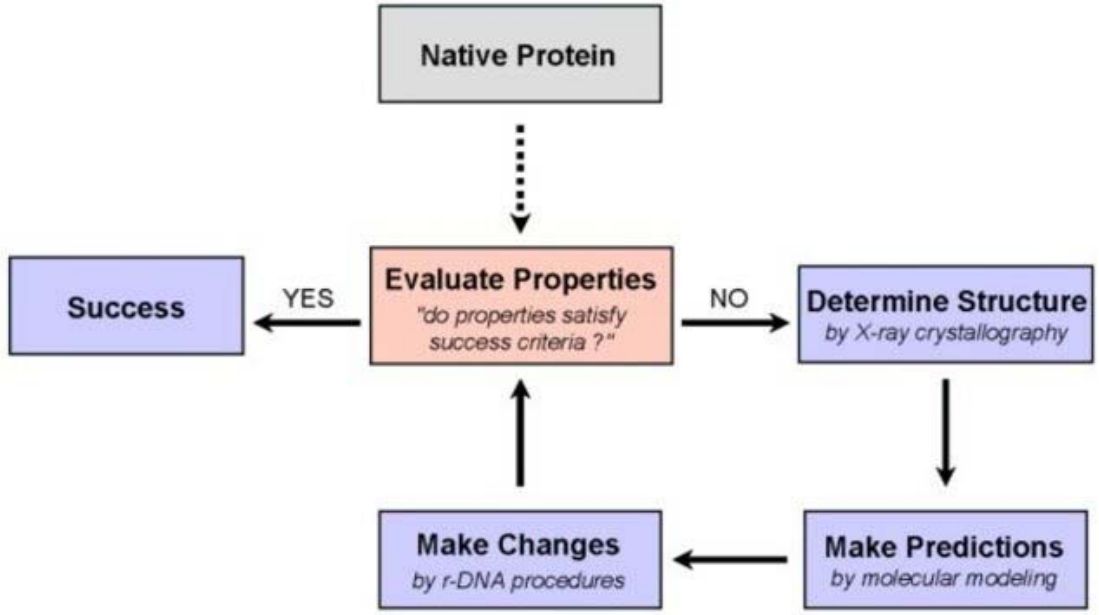

Figure 1. The rational design cycle.

In problems where the relationship between structure and function is straightforward, rational design has worked admirably. In addition to confirming known principles of protein structure, it has been able to refine and extend them. Early successes have included work to stabilize proteases against inactivation by hydrogen peroxide, changing the specificity of a detergent protease, and shifting the optimum in the pH/activity profile. Surprisingly, some of the early attempts where rational design was expected to work, such as improving stability by the introduction of disulfide bonds or salt-bridges, frequently failed, indicating that protein stability was a more complex function of structure than originally thought. In fact, properties such as stability, and those even less obviously related to structure (like detergency), are better achieved using a directed evolution approach (see Section 2.2), provided an efficient screen can be developed.

Although the early paradigm described above in which a single amino acid is changed in a site-specific manner is still widely employed, more ambitious changes to protein molecules have become increasingly common. Deletion of extended surface loops within protein structures was one of the early multi-amino acid engineering changes. In the detergent protease BPN', a loop participating in the binding of a stabilizing calcium ion was removed, creating a protein which was insensitive to the action of metal chelants (see Section 3.1). In the reverse scenario, metal ion binding sites have been deliberately engineered into enzymes to improve stability or to modulate catalytic activity.

Rational design has been used to create hybrid proteins that are fusions of pre-existing, but unrelated protein domains. Since most protein domains that confer catalytic or other properties can fold independently, they are easy to manipulate, and fusion of proteins with specific binding or catalytic domains has now become a standard tool for protein engineers. This type of rational design has been used to create proteins that are easier to

purify. One technique is to add binding domains to the amino or carboxy terminus of proteins that bind specifically to different affinity chromatographic columns (for example a hexa-histidine tail that allows for purification using a nickel column, cellulose binding domains for cellulose columns, etc.); the binding domains can then easily be removed by cleavage with a specific protease. A number of commercial kits are now available to fuse these binding domains with the protein of interest. A related strategy, which has also been commercialized in kit form, is to fuse the protein of interest with a protein self-splicing element (intein) and a chitin-binding domain. After chitin affinity chromatography, the intein can be made to self-cleave using a pH change, releasing the desired, purified protein in a single step. Such attached inteins can also be modified such that they will join two separate proteins that normally cannot be folded when present in the same polypeptide, in a process termed "expressed protein ligation."

Hybrid enzymes are commonly found in nature (cellulase for example consists of a binding domain that serves to anchor the enzyme to cellulose and a catalytic domain to facilitate chain hydrolysis). Such natural structures have been the inspiration for the rational design of new kinds of hybrid enzymes. For example, the possibility of creating new restriction enzymes by fusing 'zinc finger' proteins that bind to specific DNA sequences to the cleavage domain of an endonuclease has been explored with some success. Improved expression has also been achieved by creating hybrid proteins in which a carrier protein, such as a maltose-binding domain, is coupled to the normally insoluble inclusion body forming enzyme or protein. Finally, hybrid enzymes containing two different catalytic domains to catalyze coupled reactions have demonstrated rate enhancement of two to three-fold over the equivalent, uncoupled enzymes.

2.1.2. De Novo Protein Design

The ultimate dream in protein engineering is to be able to create any functional protein by simply specifying the sequence of amino acids that comprise it. This is not impossible since it is known that the primary sequence of a protein determines its ultimate structure. The problem (known as the folding problem) is that a completely predictive relationship of how amino acid sequences fold into their three dimensional structures has not been determined, although certain elements of the folding problem are better described than others. The types (and sequences) of amino acids resulting in known secondary structural elements such as α-helix or β-sheet are reasonably well understood. Ways to associate these elements, through the use of hydrophobic interactions, disulfide bonds, and coordination of metal ions are also clear. Consequently, much of the current effort attempts to combine known secondary structural elements using hydrophobic interactions, disulfide bonds, or metal ions to yield functional proteins.

In designing proteins "from scratch," one strategy is to employ a "four α-helix bundle" motif – a favorite because of its simplicity and functional diversity. Hydrophobic amino acids, such as leucine are introduced on one side of a α-helix to drive association. By incorporating certain functional groups, such as a heme group, it is possible to attain spectroscopic and electrochemical properties closely resembling those of native heme proteins. Another widely studied structural motif is the "helix-loop-helix." As a dimer,

this arrangement forms the four-helix structure, but it has also been studied on its own. In one example, investigators were able to reduce the three helix immunoglobulin G binding domain of protein A to a helix-loop-helix peptide of about half the size while retaining both the structure and function of the native domain. Finally, the zinc finger proteins, which bind to DNA and are structurally characterized by two β-sheets and an α-helix (ββα architecture), have been significantly downsized and shown to be stable even without their associated zinc ion. A computational design algorithm was successfully used to predict sequences, which yield this architecture without the zinc ion, although the designed proteins did not reproduce the DNA binding function.

A second major strategy in *de novo* design uses known protein structures as natural scaffolds to present a new property such as catalysis, inhibition, or metal binding sites. Examples of this approach include the introduction of an elastase-binding loop into the structure of interleukin-1β. The engineered protein acts as an elastase inhibitor suggesting that loops can be used to introduce new functionality. Metal ion binding proteins designed on the α/β scaffold of charydotoxin have been made in which the metal binding sites of carbonic anhydrase B were inserted, resulting in an engineered protein to which copper and zinc metal ions bind with high affinity. Similarly, superoxide dismutase activity (conversion of metabolically-generated superoxide to hydrogen peroxide) was designed into thioredoxin by introducing an iron atom in the hydrophobic core through the addition of a $(His)_3Asp$ site.

The inability to predict how a protein will fold has led to the use of molecular templates to provide some control in the position of different structural regions. In this approach peptide structural elements are covalently linked to chemical templates with precisely located reactive groups. Some templates include porphyrins, calix [4] arenes, and well-characterized surfaces. Again the four-helix bundle has been a favored motif, in one case even allowing the creation of a proton channel. Such template assembled synthetic proteins (TASPs) are serving as useful tools in the elucidation of the factors which control ion channel conductivity, since they allow quite precise control of the number of peptides comprising the channel, their secondary structures, and the nature of the amino acids facing the central pore. The TASP approach has also been involved in early attempts to develop redox proteins *de novo*. In one case, for example, a template allowed the assembly of a four-helix bundle resembling the binding core of the transmembrane protein cytochrome b, a key protein involved in cellular energy generation. In a further example of their flexibility, TASPs on surfaces have been made as biosensors for metal ions and antibodies (see Section 3.4).

2.2. Directed Evolution

2.2.1. Introduction and Background

Although rational design approaches have resulted in a number of successes, commercial and otherwise (see Sections 2.1, 3), the naive current state of knowledge regarding protein structure and mechanism has largely limited its predictive utility to well-studied protein families, such as the serine proteases. An alternative approach to protein engineering that dispenses with the requirement for such knowledge has emerged in the past decade, that promises to be perhaps the most significant and

powerful yet developed.

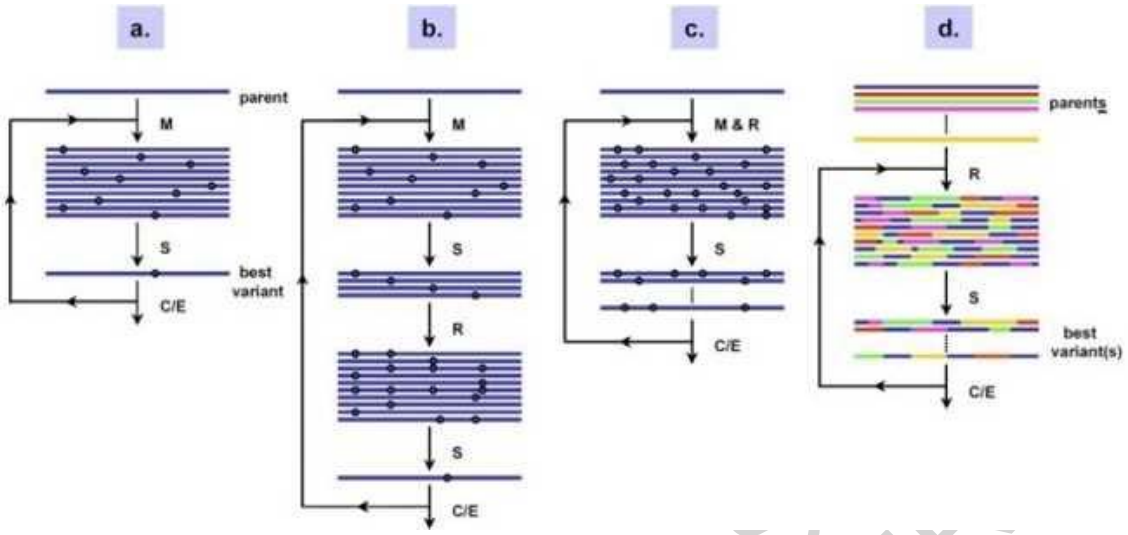

Figure 2. Directed evolution formats. Four possible DE formats that operate on DNA are in current use ("a"-"d"). Usually the formats begin with individual gene(s), and consist of combinations of steps that include mutagenesis (M), selection or screening (S), recombination (R), and repeating the cycle of steps (cycling) or reaching an endpoint (C/E). In formats "c" and "d" more than one variant can be chosen for the next cycle.

Directed evolution (DE; also termed *in vitro* evolution, directed molecular evolution, and accelerated evolution) encompasses several molecular techniques that mimic the processes of Darwinian evolution *in vitro*, by combining random mutagenesis and/or recombination of DNA (and possibly RNA) with high-throughput screening or selection for corresponding protein variants that have incorporated the desired properties (Figure 2). Because this approach is not based on any "rational" structure-function knowledge of the protein of interest, other than its functional properties and its corresponding DNA sequence, DE is sometimes referred to as "irrational" or "semi-rational" design. However, three billion years of natural evolution have provided convincing evidence that the evolutionary approach is anything but irrational; rather it has been extremely powerful, having given rise to the incredible molecular diversity seen on the planet today. Recent advances in molecular biology, genetics, and screening and selection techniques have allowed for compression of natural evolutionary timescales, measured in millions of years, to weeks in the lab.

-
-
-

Bibliography

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.]

Arnold F. H., and Volkov A. A. (1999). Directed evolution of biocatalysts. *Current Opinion in Chemical Biology* **3**, 54–59. [A good review of how directed evolution is being used to improve important enzymes.]

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.]

Chen W., Bruhlmann F., Richins R. D., and Mulchandani A. (1999). Engineering of improved microbes and enzymes for bioremediation. *Current Opinion in Biotechnology* **10**, 137–141. [Provides a readable introduction to the area of bioremediation with many references to the current literature.]

Crabb D. W. and Shetty J. K. (1999). Commodity scale production of sugars from starches. *Current Opinion in Microbiology* **2**, 252–256. [A good start on the literature of protein engineering of enzymes important in starch processing.]

Deber C. M., ed. (1999). *Biopolymers Peptide Science* **47**(1). [This particular volume contains a number of informative articles devoted to *de novo* design of proteins.]

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.]

Nixon A. E., Ostermeier M., and Benkovic S. J. (1998). Hybrid enzymes: manipulating enzyme design. *Trends in Biotechnology* **16**, 258–264. [This article provides an excellent introduction to hybrid enzymes.]

Petit C. W. (1998). Touched by nature: putting evolution to work on the assembly line. *US News and World Report*, 07/27/1998 edition; science section. [A very interesting, general account of how biological evolutionary principles are being applied in a range of science and engineering disciplines.]

Wall R. J. (1999). Biotechnology for the production of modified and innovative animal products: transgenic livestock bioreactors. *Livestock Production Science* **59**, 243–255. [Provides a snapshot of pharmaceuticals produced in transgenic cows and issues associated with "gene pharming".]

Biographical Sketches

Donn N. Rubingh is Research Fellow and Section Head in the Corporate Research Division of The Procter and Gamble Company. Dr. Rubingh obtained his Ph.D. degree in physical chemistry from the University of Wisconsin at Madison in 1973. Since then, he has worked as a scientist and research manager at P&G. His research interests include polymers, surface and colloid chemistry, surfactants, and more recently, biotechnology. Dr. Rubingh has managed a protein engineering group involved in discovering and improving industrially important enzymes since 1986.

Rowan A. Grayling received a Ph.D. in microbiology from The Ohio State University in 1995, and subsequently joined the Corporate Research Division of The Procter and Gamble Company, where he is currently Senior Scientist in Protein Engineering. At P&G, Dr. Grayling is responsible for fermentation and directed evolution efforts. His current research interests include protein engineering for industrial applications, and protein expression in and the fermentation physiology of *Bacillus* and *Pichia* species.