Chapter 1

A Brief Introduction to Protein Engineering


INDEX

1.1 Protein sequence-structure relationship............1

1.2 Protein engineering approaches
1.2.1 Pure rational protein engineering....................3
1.2.2 Pure non-rational protein engineering.................4
1.2.3 Hybrid protein engineering............................5

1.3 Focus of present thesis
1.3.1 Basis of protein surface engineering.................6
1.3.2 Relationship of active site to enzyme structure........7
1.3.3 Definition of new perms and phrases..................8

1.4 Significance of the present study......................9

1.5 References................................................10
**Introduction:** Proteins are the building blocks of cells involved in all essential life functions. They are made up of amino acids joined, together by peptide bonds, and fold upon synthesis into a three-dimensional structure possessing some biological activity. Primarily, all proteins are made up of only twenty different amino acids, joined together in various permutations and combinations, but each protein forms a different structure in three-dimensional space, upon chain folding, and this is referred to as the protein's tertiary structure. It is well known that all biological functions of proteins are derived from their tertiary structures, and that the amino acid sequence contains all the relevant information needed to determine the formation of three-dimensional structure by any protein chain. Thus, the three dimensional structure of a protein is dictated by its amino acid sequence (Anfinsen, 1973).

**1.1 Protein sequence-structure relationship:** The precise relationship between protein sequence and structure is highly complex and is still a puzzle. Any change in the sequence of a protein brings about a change in its corresponding three-dimensional structure. These changes can either be subtle, or profound. Subtle alterations in the sequence generally involve only minor alterations in the microstructure of a particular region of the protein, manifesting as changes in the shape of a local cluster of residues (either buried within the protein, or located on its surface) in the neighborhood of the altered residue, without any profound effect on either the protein's overall shape or function, or the trajectory that its peptide backbone takes through its three-dimensional structure. However, when the changes are profound, they alter the entire shape of the protein as well as the trajectory that the backbone takes through the protein's structure, and such changes can even cause a protein to fold incorrectly or not to fold at all. For example, in case of sickle cell anemia, a single amino acid substitution from glutamine to valine at the sixth position of the β-globin polypeptide chain drastically alters the structure of hemoglobin, thus causing a change in the shape of red blood cells. It is not yet possible to accurately predict the effect of alterations made in the sequence of a protein upon its structure. Although nowadays, using *in silico* methods, the effects of very limited sequence changes upon protein structure can be modeled, these require experimental validation because it is yet not possible to model all the parameters involved in determining the effects of these changes on the structure of proteins.
The number of protein sequences known to exist in nature is very large as compared to the total number of protein folds available, which indicates that a number of similar sequences give rise to similar overall folds. Thus, as a general rule, proteins with similar amino acid sequence (with identity of ~40%) tend to adopt similar overall fold/backbone structure, but two proteins from two different organisms that are not evolutionarily related can sometimes be seen to have polypeptide backbones that are almost identical in their overall shape, even though the two proteins have totally different amino acid sequences, with no similarity. The outer shape characteristics of such proteins (i.e., possessing similar, or even superimposable, polypeptide backbones), however, tend to be quite different, and this is on account of the specific 'decoration' of the backbone of each such protein by specific groups of interacting residues (side-chains) that are peculiar to that protein, in a manner determined by its specific amino acid sequence. Conservation of backbone structure thus correlates with broad conservation of function; the precise thermodynamic and kinetic parameters of functionality are influenced almost entirely by the outer shape characteristics of the protein that determine the nature of its interactions with the solvent, and with other interaction partners, and these are determined predominantly by side-chains located upon the protein's surface.

To summarize the above discussion, the precise relationship between amino acid sequence and protein structure is very subtle; all aspects of this relationship are not yet understood, or appreciated, and it is not yet possible to predict the effects of making particular changes in sequence upon a protein’s structure without doing the necessary experimentation, or without reference to a specific structural context. Against this background, any effort to develop a widely-applicable perspective, or scheme, linking sequence changes to changes in structure and/or function would, quite clearly, constitute a significant advance in protein engineering.

1.2 Protein engineering approaches: The two main approaches followed in the field of protein engineering are: (a) rational engineering approach, and (b) non-rational engineering approach. The rational engineering approach uses the detailed information about the known structure-function relationship in proteins to introduce mutations at specific sites whereas the non-rational (combinatorics-based and directed evolution-based) approach
involves either random mutagenesis such as error-prone PCR or recombination of gene fragments such as gene shuffling or phage display, followed by high-throughput screening of mutants with desired characteristics. A comparison of the two approaches is shown in Fig 1.1.

The rational engineering approach was the earliest approach used to modify the specificity of enzymes, but it suffered from the viewpoint of general applicability in protein engineering due to lack of complete knowledge about the relationship between sequences, structures and functional mechanisms of proteins. The difficulties involved in rational engineering thus led to a switching over to other protein engineering-based approaches like 'directed evolution'. However, even here, practical problems like screening large libraries of mutants for desired traits, and the problems associated with developing a high-throughput screening methodology for identification of desired functional properties, made the application of directed evolution infeasible in relation to many proteins. Therefore, protein engineering today uses a combination of rational and combinatorial methods, with the expectation that this could open up new vistas in the design of stable and efficient proteins/enzymes for biotechnological applications.

1.2.1 Pure rational protein engineering: Rational engineering has been successfully used to improve enzyme properties like stability, substrate specificity, cofactor specificity, catalytic activity etc., and some examples of these are mentioned below. Fersht and coworkers improved the specificity of subtilisin BPN’ towards substrates with large hydrophobic residues by single amino acid replacement at two positions (Rheinnecker et al., 1993). Feussner and coworkers converted cucumber linoleate 13-lipoxygenase to 9-lipoxygenase by a single substitution of a residue presumed to be responsible for the positional specificity of the substrate (Hornung et al., 1999). Mobashery’s group enhanced the substrate affinity of class A TEM_{PEC19} β-lactamase against expanded-spectrum β-lactams such as cephalosporins, which were earlier exceedingly poor substrates for this class of lactamases, by introducing two site-directed mutations in the active site of the enzyme (Vakulenko et al., 1999). Similarly, Perham and coworkers altered the coenzyme specificity of glutathione reductase from NADP^{+} to NAD^{+} by introducing mutations at specific sites without altering its substrate specificity (Scrutton et al., 1990). Interesting and successful
Fig 1.1 Detailed comparisons of the two routinely used techniques in the field of protein engineering i.e. rational protein design and directed evolution.

(Bornscheur, U.T. and Pohl, M., Curr Opin Chem Biol, 2001)
attempts have also been made to convert a protein scaffold devoid of enzymatic activity into an enzyme. Benkovic and coworkers rationally redesigned a scytalone dehydratase-like enzyme using a structurally homologous protein scaffold of (NTF2) nuclear transport factor 2 (Nixon et al., 1999). Caradonna and coworkers generated a catalytically active iron superoxide dismutase by introducing the active site of a nonheme iron superoxide dismutase into the hydrophobic core of thioredoxin, a protein with no metal binding site (Pinto et al., 1997). This approach has also been used to achieve structural stabilization of specific proteins through the introduction e.g., of specific electrostatic interactions, or other additional bonds such as disulfide bonds. Such attempts have been based on the knowledge that surface salt bridges (Anderson et al., 1990) as well as disulfide bonds provide additional stability to proteins. However, such rational attempts have met with little success as exemplified by the work of Perham and coworkers (Scrutton et al., 1988) who introduced a disulfide bond into glutathione reductase by design, to try and improve its stability, and produced an active enzyme that formed the intended disulfide bond but showed no additional structural stability.

1.2.2 Pure non-rational (combinatorial) protein engineering: This technique does not require in-depth knowledge of structure-function relationships and is nowadays in routine use to improve enzymes for industrial and biotechnological applications (Tao and Cornish., 2002). Arnold et al used random mutagenesis to improve the activity of a protease, subtilisin E, in organic co-solvents (Chen and Arnold., 1993). Ellington et al, used random mutagenesis and shuffling to improve the thermal stability of beta-glucuronidase (Flores and Ellington., 2002). Similarly, Murakami and co-workers improved the usefulness of firefly luciferase enzyme in genetic reporter assays by using a combination of gene chimerization and random mutagenesis approaches (Hirokawa et al., 2002). Examples of modifying enzyme selectivity using directed evolution are also available in literature. For example, Kagamiyama and co-workers re-engineered aspartate aminotransferase, a \( \beta \)-branched aminotransferase, using random mutagenesis and a cell-based selection for complementation of an engineered deficiency in the endogenous beta-aminotransferase (Yano et al., 1998). Zhao et al created a novel corticosterone activity in human estrogen receptor alpha ligand binding domain, by an in vitro coevolution method (Chen and Zhao., 2005). Another notable
example of the use of a purely random engineering approach is that of Romesberg and co-workers, who converted a DNA polymerase into a RNA polymerase using co-expression of the enzyme and substrate on the phage surface (Xia et al., 2002).

1.2.3 Hybrid (rational-combinatorial) protein engineering: An impressive example of the combination of rational site-directed mutagenesis and random error prone PCR approaches is the work of Pedersen and co-workers, who produced a mutant of heme peroxidase from Copernius cinerius suitable for use as a dye-transfer inhibitor in laundry detergent (Cherry et al., 1999). Oshima and co-workers also successfully combined both random and site-directed mutagenesis to improve the thermal stability of a 3-isopropylmalate dehydrogenase variant from Bacillus subtilis (Akanuma et al., 1999). Similarly, using a combination of both mutagenesis strategies, Glick and co-workers, reduced the maturation half-life of Discosoma red fluorescent protein (DsRed) from 11 h to less than 1 h and made it suitable for live cell imaging (Bevis and Glick, 2002). Successful attempts to create proteins with novel catalytic activities have also been made by some researchers. For example, Kim and coworkers evolved a new catalytic activity (β-lactamase activity) on the αβ/βα metallohydrolase scaffold of glyoxalase II (Park et al., 2006). Another successful example is that of Segovia and Peimbert who evolved a beta lactamase activity on a D-Ala D-Ala transpeptidase fold (Peimbert and Segovia, 2003). Nowadays, the combination of both rational mutagenesis and directed evolution is the approach of choice to create novel enzymes of commercial importance.

Comparable rates of success or failure have characterized all protein engineering approaches used till date. Most approaches, as already discussed above, have focused on altering the chemical nature of enzymatic activity, e.g., the identity of the preferred substrate, while some have also focused on altering physical aspects of activity such as substrate turnover number, or structural stability.

Notably, however, virtually no efforts have been made until now to rationally engineer the temperature regimes of enzyme activity, although such changes have inadvertently been obtained as side-effects of alterations in global thermal stability, e.g., in the form of a broadening, or narrowing, of the regime of activity. A survey of the available literature reveals four instances of the making of chimeric proteins with altered global
thermal stability, or altered temperatures of optimal function, using swapping of whole catalytic domains between homologous proteins, to recombine the structural features of one protein with functional features of another protein, usually sourced from a different domain of life. The first two instances of such chimeric proteins involved beta-glucosidases from the work of Hayashi and coworkers (Singh and Hayashi, 1995; Goyal et al., 2001), a third instance involved citrate synthase from the work of Danson and coworkers (Arnott et al., 2000) and the last instance involved avidin (Hytonen et al., 2007). In the first instance, chimeras of homologous β-glucosidases from Agrobacterium tumefaciens and Cellvibrio gilvus (~37% sequence identity; 40% sequence similarity) were made (Singh and Hayashi, 1995). In the second instance, chimeras of homologous β-glucosidases from Agrobacterium tumefaciens and Thermotoga maritima were made (Goyal et al., 2001). In the third instance, chimeras of homologous citrate synthases from Thermoplasma acidophilum and Pyrococcus furiosus were made (Arnott et al., 2000). In all three instances, the intention was to obtain chimeras with enzymatic properties of improved enzymatic stability and altered temperature and pH of optimal function. In a fourth instance, which was found in the patent literature but not in the primary journal literature (Hytonen et al., 2007), the thermal stability of a chicken avidin protein was improved by replacing one of its structural domains, named beta 4, with the entire beta 4 domain of a different avidin-related (AVR) protein.

1.3 Focus of the present work: In this thesis, our focus is on re-engineering protein surfaces to design novel proteins that combine the structural features of one protein with the functional features of another homologous protein sourced from a different organism.

1.3.1 Basis of our novel protein surface engineering approach: Protein structures with predominantly beta-sheet secondary structure were chosen to attempt surface engineering because of the predictability of the effect of making changes in β-sheets as compared to α-helices, present on a protein surface. This can be explained by the fact that in β-sheets, the side chain of any residue in a strand faces away from the sheet in a direction exactly opposite to that of its two immediate neighbors on either side, whereas, in case of α-helices, next-neighbor residues are present on the same side of the helix; thus, mutations
performed in helices without making accompanying (compensating) mutations in the immediately neighboring residues can greatly upset structure-formation and stability, but this need not be the case in sheets. Alternating sets of residues in a β-sheet face away in opposite directions and can evolve entirely independent packing schemes for their side-chains, with little scope for any mutual influence of the two faces on each other. In perfectly planar sheets, residues on the same face can sometimes be too far apart to interact effectively to create a surface; thus, in multi-sheet structures, residues achieve rigid conformations through the stacking together of sheets which allows facing sets of side-chains to interact through inter-digitation. Beta sheets made of long strands are, however, rarely perfectly planar, with most curving naturally. Although such curved sheets can also stack, residues on the concave, solvent-exposed face of the top sheet in a set of stacked sheets essentially interact mutually to create an autonomous packing scheme, creating a surface that is effectively insulated away from the other sheets in the stack. This creates possibilities for the selective remodeling of the concave face of the top sheet in such β-sheet-rich proteins. Fig 1.2 shows in detail, the arrangement of residues in beta-sheet structures and their side-chain interactions to form a surface.

1.3.2 Relationship of active site to enzyme structure: By definition, the active site of an enzyme comprises key residues directly involved in catalysis but in this thesis, we are introducing a new term “active surface” for the first time. We define “active surface” as the entire surface of an enzyme comprising all the residues involved in binding substrate as well as the key residues known to have a direct role in catalysis. Active sites, due to their role in catalysis, form the most important part of an enzyme molecule. It has been shown by Tsou that, enzyme inactivation occurs before the unfolding of the global structure of the molecule, suggesting that active sites are more “fragile” and more easily “perturbed” than the rest of the molecule and are, therefore, more conformationally flexible than the rest of the molecule. (Tsou, 1993; Tsou, 1995). A detailed comparison of conformation and activity changes of creatine kinase during denaturation by guanidinium hydrochloride and urea resulted in an initial phase of rapid inactivation and kinetically, the inactivation rates were over three orders of magnitude faster than the rates of global molecular conformational changes under the same conditions. Similar results were obtained for other enzymes like
Fig 1.2 Alternating geometrical disposition of residues within strands participating in beta-sheet formation, and the formation of a surface by side-chains from adjoining strands in a sheet.

Panel A: Alternating residues in a strand that is a part of a beta sheet face away from the plane of the sheet in opposite directions. Panel B: Side-chains from adjacent strands in a beta sheet facing away from the sheet in the same direction lie adjacent to each other. Panel C: These atoms of side-chains shown in panel B interact with each other. Panel D: The atoms of side-chains shown in panel C interact well enough to form a surface.
ribonucleaseA (Liu and Tsou., 1987; Yang and Tsou., 1995), D-glyceraldehyde-3-phosphate dehydrogenase (Xie and Tsou., 1987), citrate synthase (West et al., 1990), glutamate dehydrogenase (West and Price., 1988), and myosin subfragment-1 (Nozais et al., 1992), and many more such examples are known. This phenomenon has also been found in proteins other than enzymes, for example, Deves and Krupka reported that the inhibition of choline transport in erythrocytes by n-alkanols, involves unfolding/disruption of a small but essential part of the carrier protein (Deves and Krupka et al., 1990). All of these results prove that enzyme active sites are relatively autonomous of enzyme structure i.e., (i) active site structure and function is dependent on the retention of overall enzyme structure, but not a necessary consequence of overall structure retention by an enzyme; and (ii) active sites can behave autonomously of the overall three-dimensional structure, in respect of their micro-conformational flexibility.

Therefore, we decided to explore the feasibility of re-engineering of the active surfaces of enzymes, incorporating the active site (i.e., including the catalytic and substrate/ligand-binding regions).

1.3.3 Definition of new terms and phrases: We have defined below, some new terms and phrases relevant to our novel approach of engineering enzymes. The phrases 'SURFACE REMODELING' or 'SURFACE REENGINEERING' be taken to mean the changing of either a part, or the whole, of the surface of a protein. Where the change is made to try and imitate the surface of another protein, the phrase 'SURFACE TRANSPLANTATION' may be used. Where the change is made to try and imitate only the parts of the surface of another protein that are involved in interactions with one or more (macromolecule or small molecule) substrates, the more descriptive phrase, 'ACTIVE SURFACE TRANSPLANTATION', may be used, along with 'ACTIVE SURFACE GRAFTING', 'PROTEIN SURFACE GRAFTING' or 'PROTEIN SKIN GRAFTING'. It may be noted that the phrase 'ACTIVE SURFACE' includes the conventional phrase, or term, 'ACTIVE SITE'; thus, whereas 'ACTIVE SURFACE' defines regions of a protein's surface involved both in substrate/ligand binding, and in catalysis (where catalysis is involved), 'ACTIVE SITE' defines only the specific residues involved in a catalytic role. Therefore, an 'ACTIVE SURFACE' may be defined both for enzymes, and for proteins that
are not enzymes but which engage in intermolecular interactions. In contrast, the phrase ‘ACTIVE SITE’ has, by convention, come to describe catalysis.

1.4 Significance of the present study: The major challenge in the field of protein engineering is to create tailor-made biocatalysts, and other proteins of desired functionality and/or interaction characteristics, and very limited success has so far been achieved in this direction. A number of attempts have already been made to engineer the chemical characteristics of enzymes, like changing substrate specificity, enhancing catalytic turnover or thermo-stability, using the available techniques in the field of engineering, as already mentioned above in section 1.2, but no efforts have actually been made to engineer the physical characteristics of enzymes like altering the temperature of optimal activity or stability of an enzyme surface. Therefore, the challenge lies in designing novel and rational strategies to engineer protein surfaces from the perspective of altering physical characteristics like the temperature of optimal activity or stability. This led to us to frame the broad objective of our study i.e. to examine the effect of replacing/altering the active surface of an enzyme on its physical characteristics.

We demonstrate in this thesis a novel and rational strategy, which can be widely applied to other beta-sheet based protein structures, to create novel enzymes, recombining structural stability characteristics of one protein with the surface and functional characteristics of another homologous protein, sourced from two different domains of life. Also, the successful demonstration of this surface re-modeling approach confirms the autonomous behavior of active surfaces of enzymes, and thus, opens up new opportunities in the field of enzyme engineering to design novel enzymes by replacing one active surface with another from a different enzyme, using the principles as already mentioned in section 1.3.1 and 1.3.2.

Therefore, the surface remodeling approach offers the advantage of greater applicability to beta-sheet based protein/enzyme structures, as compared to the existing engineering strategies, for designing novel biocatalysts of industrial importance.
1.5 References


