ABSTRACT

A major obstacle to the understanding of the relationship between structure and function/activity in any protein is the inability to apply the principles of chemistry to macromolecules. This gap was most acutely felt in the field of enzymology. Although the fields of enzyme catalysis and X-ray crystallography have emerged to throw light on structure-function relationships, the need to modify and change the structures of the macromolecules continues to be strongly felt, the idea being to modify functional groups present at enzyme active sites, or at locations critical to protein folding, and assembly, and examine their effects on activity. The advent of the field of "protein engineering" opened up many avenues for protein chemists to study the mechanisms of enzyme catalysis. It also contributed many insights into catalysis, stability, specificity, and folding of proteins. Within a span of 25-30 years, protein engineering has made progress - by applying either rational or combinatorial approaches - in changing proteins in respect of small or big alterations (from one to two residues to entire loops and even domains) to make new variants, or even completely new proteins.

Proteins are the most fascinating of life's macromolecules, performing roles in each and every domain of function within living cells and organisms. Since they are so important, they are studied intensely in respect of their structures, functions and folding behavior. Unveiling the secrets behind the evolution of proteins and the diversification of their functions has long been a goal of protein scientists. At the present time, there exists a diverse array of protein folds and structures, which adds to the complexity of our understanding of protein folding. Protein engineering continues to play an essential role in enhancing our basic understanding of how naturally-occurring proteins fold and function, paving the way for the generation of new structural/chemical scaffolds, or activities, and enhancing our understanding of their behaviour outside their natural biological framework and to build novel proteins that do not exist in nature, purely on the basis of design based on principles gleaned thus far.

Nature itself acts as an opportunistic modifier or engineer, making new scaffolds by mixing and matching them, generating new activities in those scaffolds (enzyme promiscuity) by rounds of mutations and natural selection. Enzyme promiscuity is a recently-coined term in enzymology and deals with the broad substrate specificities (non-

specific nature) of enzymes. It also points towards the evolution of the enzyme to become more specific.

To look into what must have happened during the course of evolution, one needs a protein or a structural fold which is not intricate, displays a symmetrical and ordered arrangement of secondary structures, and can be dissected into parts - as and when required - as per the need of the experiment. TIM barrel or the $(\beta/\alpha)_8$ barrel proteins top the list in this regard and are considered as the favourites of protein engineers, due to their ordered structures, regular arrangements and high structural symmetry. They also account for 10 % of enzyme crystal structures solved, significantly exceeding other structural folds of enzyme in number and functional diversity. A lot of engineering trials have been done on this fold. Gene duplication and fusion is one of the strategies adopted by proteins to diversify. This is best exemplified in TIM barrel enzymes where symmetry and orderliness of the secondary structural elements appears to have been exploited by evolution in ways that are possible for us to potentially mimic, through protein engineering. In particular, TIM barrels are thought to have arisen through gene duplication and mutations, resulting in the assembly of two half barrel domains into a full beta/alpha barrel domain. In principle, extending the process that appears to have occurred during evolution, half barrel domains can be selected and shuffled, in modular fashion, to form full beta beta/alpha barrels of half domains sourced from different enzymes. Attempts to engineer beta/alpha barrels through the creation of numerous novel chimeric constructs combining half-barrel domains of different beta/alpha barrels of the endoglucanase/cellulase varieties are described in Chapter 3. In Chapter 4, we describe studies pertaining to the surprise discovery and detailed structural-biochemical investigation of a novel promiscuous function in a beta/alpha barrel enzyme, namely an endoglucanase function apparently derived from a 'second site' with allosteric interactions with the primary site, in a triosephosphate isomerase from *P. furiosus*.

Along a different vein, considering the important roles played by the molecules (e.g, antibodies) in which it exists, the immunoglobulin (Ig) fold too - like the beta/alpha barrel, or TIM fold - has also been extensively engineered. In antibodies, it has been engineered for increasing affinity, avidity improving half life, *et cetera*. Numerous engineered variants of antibodies are present in the market, which are produced in different hosts, i.e. mammalian cell lines, yeast and *E.coli*. The production of single chain antibodies possessing both neutralization and effector function is a recent trend. There are

many 'gap' areas in the field of antibody engineering which can be explored. One of these is the building of novel genetic constructs involving antibodies. Another is disulfide bond engineering to obviate the need for formation of disulfide bonds within antibodies, to the extent possible, and facilitate their production in prokaryotes, in the cytoplasm. Attempts along these lines have been made in the present thesis and is presented in chapter 5.

Overall, this thesis aims to understand structure-function relationships in protein structures of the beta/alpha and all-beta varieties, the strategy followed by different enzymes to adapt to different environments, folding mechanisms in fusion proteins (if sequences are sourced from different organisms), plasticity of interactions forming the core of a beta barrel, and stabilization of the structure by various forces and factors. In this work, the rational approach of protein engineering has been applied to do engineering trials, involving two common protein folds, viz., the alpha/beta barrel fold and the immunoglobulin fold. Besides, as already mentioned, the thesis also describes the discovery and characterization of a novel endoglucanase activity found in triosephosphate isomerase from an archaeon, a glycolytic pathway enzyme.

The thesis is divided into six chapters or sections, details of which are mentioned below:

(I) Chapter/Section I provides a general introduction to protein structure, folds and domains and, in particular, the TIM barrel (alpha/beta-based) and immunoglobulin (beta sheet based) domains, the engineering strategies used and their outcomes, etc. This chapter also provides a detailed account of enzyme promiscuity (a relatively recently-coined term in enzymology); its types, development, *et cetera* in enzymes.

(II) Chapter/Section II introduces the general materials and methodology followed, i.e., it gives a general outline of the materials used and methodology followed to perform the experiments reported in the thesis. The first half of this chapter explains cloning, expression and purification. The second half explains the spectroscopic, spectrometric and calorimetric techniques/handles used to characterize studied proteins and answer various questions.

(III) Chapter/ section III titled "**Endoglucanase activity in chimeric beta/alpha barrel cellulases assembled through half barrel gene shuffling**" explains the concept, design, cloning, expression, purification and characterization of 'chimeric cellulases' formed by fusion of half-domain alpha/beta barrel structures. A rational approach was applied for

engineering trials for chimera genesis by half-domain fusions of beta/alpha barrel proteins (mainly cellulases) sourced from different organisms. The background to the use of half-barrels has been explained in the general introductory chapter and also in the introduction to this chapter. The chimeras produced by fused half domains (utilising SOE-PCR approach) were analysed for folding, assembly, structure formation, stability and activity. Although, all the chimeras ended up being present in inclusion bodies, they were purified under denaturing conditions, refolded by various protocols and then analysed. Despite low sequence similarity, the feasibility of combining two half barrels contributed by two different enzymes to generate chimeras with substantial activity is insightful enough to be explored further.

(IV) Chapter/section IV titled **"Endoglucanase activity arising from a second site in triosephosphate isomerase from *Pyrococcus furiosus* : Compensation for a metabolic handicap ?"**explains a serendipitous finding relating to an alpha/beta barrel protein, i.e., the presence of an adventitious (second-site based) endoglucanase activity in *P. furiosus* Triose phosphate isomerase (PfuTIM), a glycolytic pathway enzyme of a hyperthermophile of archaeal origin – which was selected originally to function as a control in the experiments in Chapter/Section III. It explains that by being capable of playing a dual role as an isomerase and an endoglucanase, PfuTIM appears to represent an example of divergent evolution and of a survival strategy in which a TIM enzyme doubles as a generator of glucose. The activity not only peaks at the expected high temperature (for an enzyme encoded by the genome of a hyperthemophile archaeon), but also at the lower temperature of ~40 °C. We have also shown beyond doubt that the activity does not owe to the standard active site responsible for the interconversion of glyceraldehyde-3-phosphate (G3P) and dihydoxyacetone phosphate (DHAP), using the twin strategies of examining the actions of known TIM inhibitors, and effects of mutations of standard active residues.

(V) Chapter/section V titled **"Experiments in antibody engineering: light chain, heavy chain fusions and disulfide engineering"** describes the design, expression, purification and initial characterization of a single chain whole antibody (SCWA) made by fusion of the heavy and light chain of an antibody of human origin produced in *E.coli* and *Pichia pastoris.* Apart from that disulfide mutants (in which the cysteine residues being replaced with the residues of naturally occurring immunoglobulin domains which lack disulfide bind) were also cloned.