Abstract

Polypeptide chains, following their synthesis on the ribosome, fold into biologically active three-dimensional structures known as proteins. A protein molecule shows extraordinary efficiency and incredible rate of folding from an unfolded structure to its native folded form. The astronomically large number of possible conformations, and the very small fraction of time a polypeptide chain takes to fold, clearly indicate that high fidelity mechanisms control protein folding. Accordingly, it is conceivable that the search for the native state is not random (thermodynamic control) but rather follows a precise scheme (kinetic control) involving definite pathways. Since, many proteins can unfold and refold with high fidelity even *in vitro* (i.e. even in the absence of cellular chaperone machinery) the amino acid sequence must contain the information required for the formation of a final, biologically active three-dimensional structure.

In this work, we have attempted to understand aspects of the relationship between amino acid sequence and protein folding. My work, describes the biophysical and biochemical characterization of some 'non-natural' proteins derived from the sequence of naturally occurring proteins through the introduction of profound mutations. To elucidate the exact nature of the fold adopted by profoundly mutated sequences, structure at atomic resolution is required to be solved. But both the methods of structure determination *viz.* NMR and X-ray crystallography demand solubility of proteins at high concentration. Unfortunately, most of our redesigned proteins showed a tendency to aggregate at the very high concentrations required for NMR or X-ray crystallography, although, we attempted crystallization with some of the proteins with the highest possible concentration and managed to obtain some initial needle like crystals. Further study is currently underway towards solving high resolution structure both by X-ray crystallography and NMR.

Nevertheless, the present study in which we have mainly studied the secondary and quaternary structural status, folding/unfolding behavior in response to heat and denaturant etc. and overall population behavior of the redesigned proteins address a number of fundamental question in a variety of ways. In particular, we have examined (a) the influence of amino acid 'nature' (as opposed to amino acid 'identity') on folding, (b) the influence of backbone direction, or polarity, on folding, and in addition, we also examined (c) the excisability and folding behavior of certain locally determined small elements of structure.

The entire thesis is divided into five chapters with references included at the end of each chapter. In **chapter 1,** a general brief introduction to the sequence-structure relationship in proteins has been discussed, followed by a detailed discussion of the most commonly used materials and methods in **chapter 2**. The remaining three chapters describe the three major studies done as mentioned above. **Chapter 3** consists of studies querying the influence of *'nature'* of amino acids in relation to chemical *'identity'*, on protein folding behavior. **Chapter 4** describes the biophysical and biochemical characterization of a backbone reversed form of the *E. coli* co-chaperone GroES, (retro)GroES. **Chapter 5** describes the creation of a small all-β mini-protein and its biophysical and biochemical characterization. Each chapter starts with a brief 'introduction to the chapter' followed by an introduction of the protein used in the study. A brief discussion of the 'experimental design and strategies' specific to each chapter has also been introduced, which contains mainly the gene-synthesis strategy with the list of primers and the details of PCR conditions that actually worked after trial and errors with the primers for a particular gene. This section is followed by 'Results' 'Discussion' and 'References' sections for each chapter. The general methods applicable to chapters 3-5 are described in chapter 2, as already mentioned.

Residue *'nature'* versus residue *'identity'*:

In **chapter 3**, we have tried to examine the role of amino acid 'nature' over its 'identity'. Mutations involving replacement of amino acid residues in proteins by other residues of different chemical identity, but of similar physical and chemical nature are known as conservative mutations. The occurrence of such mutations, frequently seen in nature and also in course of evolution, suggests that a certain degree of plasticity might apply to the packing of side-chains within protein interiors, facilitated by the freedom of rotation available around chemical bonds. In this part of the work, we explore the conformational behavior of three globally-mutated (GM) variants of a model 65 residues-long protein, wild-type chymotrypsin inhibitor 2 (WTCI2). Two of these mutants, GMCI2-1 and GMCI2-2, incorporate 55 global conservative residue substitutions mutations out of 65 residues, according to a defined algorithm. The third one GMCI2-3, is derived to incorporate activity to the non-active GMCI2-1 and GMCI2-2. To begin with, WTCI2 and the two globally, conservatively mutated variants, GMCI2-1 and GMCI2-2, were created *in silico* as well as *in vitro*, and examined both computationally and experimentally for signs of WTCI2-like structure-forming behavior.

The GMCI2-1 and GMCI2-2 variants differed only in respect of the replacement of 'Val' by 'Ala' (GMCI2-1) or 'Ile' (GMCI2-2). Energy minimization and molecular dynamics simulations established that both GMCI2-1 and GMCI2-2 can adopt and retain a well-optimized WTCI2-like structure, starting from a non-optimized structure with the same backbone coordinates as WTCI2. Independently, structural-biochemical studies established that both the recombinant variants can adopt a compact, spherical, folded and thermally stable conformation through folding. However, GMCI2-1 was found to be soluble over a wide range of pH and was monomeric and soluble up to a concentration of ~2.0 mg/ml at pH 7.0 while GMCI2-2 was soluble only up to ~0.5 mg/ml at pH 4.0 buffer and showed a marked tendency to form amorphous globular aggregates of ~25 nm diameter at higher concentrations. As evident from CD spectroscopy, GMCI2-1 folded to a structure dominated by β_{II}-type structure very similar to that of WTCI2, with extreme thermal stability indicating retention of the bulk of WTCI2's structural-biochemical characteristics while GMCI2-2 folded to a structure with non-WTCI2-like secondary structural content, appearing to be dominated by α-helical structure. On probing the presence of hydrophobic patches by ANS binding, no surface hydrophobicity was noticed for GMCI2-1, like with WTCI2, but GMCI2-2 displayed marked surface hydrophobicity. Though the computational and experimental results indicate that a CI2-like conformation for both GMCI2-1 and GMCI2-2 is feasible and adopted, biochemical and biophysical characterization of the recombinant expressed protein suggests, presumably a WTCI2-like behavior/fold for GMCI2-1 but not for GMCI2-2. As both the forms, GMCI2-1 and GMCI2-2, failed to inhibit subtilisin, a chymotrypsin-analogous protease, a function readily shown by WTCI2, we created a third variant GMCI2-3. GMCI2-3 is a reactive loop restored form of GMCI2-1. The hope was to impart WTCI2-like protease inhibitory activity toGMCI2-1. However, GMCI2-3 also failed to show any activity. Moreover, its biophysical and biochemical characterization showed similarity with GMCI2-2 rather than GMCI2-1.

Thus the present upcoming understanding that entirely different sequence can adopt the same fold if residues of like nature are placed at corresponding chain locations, is still poorly understood, at least, experimentally. Our results provide an ambivalent answer to the question of whether polypeptides incorporating residues of identical nature at equivalent chain locations fold with similar characteristics; even so, the remarkable WTCI2-like folding of GMCI2-1 suggests that more studies may be merited.

Backbone reversed (retro) Proteins:

In **chapter 4** we have studied the structural consequences of polypeptide backbone direction reversal ('retro' modification). To examine the effects of 'retro' modification on all-β sheet structures, our lab have previously studied the recombinant 'retro' forms of two all-β sheet proteins; (retro)HSP 12.6, which folds to adopt a suitable multimeric structure (Shukla and Guptasarma, 2003; *J. Biol. Chem.*) and (retro)CspA, which folds poorly and transforms rapidly into amyloid fibers through intermolecular association (Shukla and Guptasarma, 2003; *Protein Engg.*). In this part of the work, we have studied the behaviour of a third 'retro' protein, the retro-form of the 97 residues-long GroES co-chaperonin of *Escherichia coli,* which appears to be the best-behaved retro-protein studied in our lab till now, with solubility ranging from 5.0-5.5 mg/ml in pH 8.0 buffer. (retro)GroES is found to adopt a β_{II}-type structure with a far-UV CD spectrum entirely like that of the parent *E. coli* GroES, suggesting a β sheet-dominated structure consisting of short β-strands interspersed with poly(Pro)II-type (PPII) structures for both polypeptides. FTIR spectroscopy also indicates the formation of a β sheet-based structure, while gel-filtration chromatography reveals that the folded chain adopts a trimeric or pentameric quaternary structures depending on the buffer ionic strength. The protein undergoes cooperative dissociation and non-cooperative unfolding by guanidine hydrochloride and urea, as expected for a molecule rich in PPII structure. The protein refolds readily from denaturant, and is unfolded by 7.0 M of guanidine hydrochloride (GdnCl). Mutational substitution-insertion of tryptophan as spectroscopic probe at three different sites in the sequence of (retro)GroES causes no perturbation in either the fold and its CD spectral characteristics, or the protein's quaternary structural status. Very interestingly, unlike with GroES, (retro)GroES is not unfolded by heat. Instead there occurs a reversible structural transition involving conversion of PPII structure to β-sheet structure; however, with no attendant aggregation at temperatures as high as 90°C. The protein displays a reversible consolidation of structure upon heating to a temperature as high as 90°C, rather than any structural melting. Needle-like crystals have been obtained, suggestive of homogeneity in structure adoption and attempts are on to refine crystallization conditions and determine the structure of (retro)GroES.

The structure-forming characteristics of GroES thus appear to be partly preserved through the backbone reversal process, although the differential conformational behavior upon heating indicates that fundamental differences remain.

Creation of an excised engineered all-β mini protein:

Chapter 5 discusses the creation of a soluble all-β mini-protein through engineering of a β-hairpin excised from α-hemolysin. A short β-finger like structure is present in α-hemolysin, a heptameric pore-forming protein of *S. aureus*. Each monomer of this protein has a stretch of 40 amino acid residues (110-150), which forms an anti-parallel β-stranded hairpin (β-finger). During assembly of monomers of α-hemolysin into heptamers, seven such β-fingers associate into a 14-stranded β-barrel. In this part of our study we have made an attempt to fashion a globular protein out of two conjoined β-hairpin structural motif(s). We created a gene encoding, in tandem, two copies of the 40 residues-long transmembrane β-hairpin tongue (BHT) motif of α-hemolysin of *Staphylococcus aureus,* with mutations to replace selected hydrophobic residues. Seven selected hydrophobic residues on each copy of the BHT motif's lipid-facing surface (residues of the β-finger that faced outward, that is towards the membrane of the pore in the pore forming protein) were mutated to hydrophilic residues, to prevent or reduce any non-specific aggregation based on hydrophobic interactions. We examined the excisibility and autonomy of folding of the small β-finger to study the ability of a peptide sequence to form structure independently of the rest of the α-hemolysin structure. We studied the CD, FTIR and NMR spectroscopy to obtain structural information of the sequence. Tandem BHT turned out to be expressed as a soluble polypeptide which could be raised to concentrations of ~2 mg/ml. It displayed several characteristics of a folded mini-protein, including: (i) the presence of β-sheet and polyproline type II secondary structure, (ii) dispersion of signals in its 1-dimensional proton NMR spectrum, (iii) evidence of the occurrence of unfolding and dissociation transitions in the presence of denaturants, and (iv) the ability to transform from sheet to helical structure through a biphasic structural transition in the presence of the cosolvent, trifluorethanol.