Introduction
The assembly of polypeptide chains by protease-mediated semisynthesis can be a valuable complement to site-directed mutagenesis and chemical synthesis (Chaiken, 1981). Proteases that normally catalyze the hydrolysis of the peptide bond also catalyze the ligation of the peptide fragments as demanded by the principle of microscopic reversibility of reactions. The use of the protease action in 'reverse' for synthesis of peptide bond offer several advantages over purely chemical methods. The reactions occur under mild conditions and follow rather strict stereo- and regio-specificity, thus eliminating the need for side-chain protection. Protease-mediated peptide ligation approach provides the ease for incorporation of chemically manipulated protein modules containing non-coded amino acids or spectroscopic probes into proteins. Nonetheless, the protease-catalysed peptide ligations usually occur under special situations in organic co-solvents, and especially for those polypeptide fragments that are held together by non-covalent interactions (interacting fragments) and are bestowed with stereo-chemical proximity of the reacting ends. Thus, there are two major impediments to the peptide ligation by proteases. They are the entropic factor; which manifests as the proximity of the two reacting ends of the substrates, and the enthalpic factor, which emanates from the ionization state of the reacting α-carboxyl group.

Limited proteolysis of many proteins has yielded non-covalently associated complexes that display native-like conformation and, often, function. One remarkable aspect of these complexes, generally referred to as 'fragment complementing system', is that the close proximity of the termini at the breakpoints necessitated by the native fold, can aid the protease to catalyze the reformation of the missing peptide bond(s), in a thermodynamically favorable situation. The aqueous-organic solutions that have low dielectric constants usually promote reverse proteolysis by enhancing the pKa of the carboxyl group. The protease-mediated re-formation of native ribonuclease A (RNase A), triosephosphate isomerase (TIM), staphylococcal nuclease, lysozyme etc. from their respective fragment complementing systems have been demonstrated in the presence of organic co-solvents (Homandberg and Laskowski, 1979; Vogel and Chmielewski, 1994; Taniuchi et al, 1977; Vogel et al, 1996).
Introduction

In those systems that lack the non-covalent interaction or complexation of substrate peptide fragments prior to ligation, the protease-catalyzed splicing of the peptide bond may be accomplished by using product-directed molecular traps. Such traps facilitate the reverse proteolysis by binding to the product, and continuously removing it from the chemical equilibrium (Homandberg et al., 1982; Nyberg, 1988). A classic example of this is the clostripain-mediated synthesis of RNase A\textsubscript{1-15} from RNase A\textsubscript{1-10} and RNase A\textsubscript{11-15} in the presence of S-protein (RNase A\textsubscript{21-124}). The synthesis of RNase A\textsubscript{1-15} proceeds smoothly in the presence of S-protein because the latter binds to the nascent synthetic product (RNase A\textsubscript{1-15}) and thus the scissile peptide bond in the contiguous product is protected from proteolysis leading to the accumulation of the product.

The protease-catalyzed synthesis of non-interacting peptide segments can also be facilitated in the absence of any external molecular trap in situations where the product becomes less susceptible to proteolysis by the virtue of its conformational rigidity. The idea of using the stabilized, protease-resistant structure of the product (conformational trap) that results from the two less
structured peptide reactants, as a mechanism to drive reverse proteolysis has been a focus of several studies in recent years (Kumaran et al., 1997; Roy et al., 1992). Interestingly, the viability of the ‘conformational trap mechanism’ has been demonstrated even in neat and dilute aqueous solution (Kumaran et al., 2000). This raises the intriguing possibility that proteases, under special circumstances, might catalyze the synthesis of peptide bonds \textit{in vivo}.

The notion that proteases may be endowed with the propensity to splice polypeptide segments \textit{in vivo} appears more realistic when one considers the fact that the environment of living cells is crowded, and, therefore, any reaction that is accompanied with an increase of available volume should be favored in the macromolecular crowded milieu. For this reason, macromolecular crowding promotes association or conformational compaction of the polypeptides as they result in the increase of the total available volume. Thus it may be possible that under physiological conditions, and in favorable situations, the energetic barrier to peptide bond synthesis by proteases may be overwhelmed by the structural feature of the product, and the positive influence of volume exclusion effect emanating from the crowded environment of the living cells. The effect of crowding on reverse proteolysis, and other studies related to protease-catalyzed peptide ligation/protein semisynthesis forms the subject matter of this thesis. As a preamble to the work discussed in the thesis, a brief review of the current standing of the synthetic and biosynthetic approaches for peptide/protein ligation is presented below.

\textbf{Chemo-enzymatic approaches: Beyond the genetic code}

\textit{Chemical synthesis}

Protein engineering had its origins about twenty years ago, when site-directed mutagenesis, a method that allowed amino acid sequences in proteins to be altered at-will, was introduced. Since then, most studies of the protein action have been carried out by recombinant DNA based expression of proteins in genetically engineered cells. From its introduction, this powerful tool has revolutionized the study of proteins by enabling the production of large amounts of
Introduction

proteins of defined molecular composition, and by allowing the systematic variation of the amino acid sequence of proteins. Expression of proteins in engineered cells is now a mature technology, and its scope and limitations are well understood. The small size proteins (less than 30 kDa) are generally easier to express than large, multi-domain proteins as the folding of the large protein molecules can sometimes be troublesome. Also processing related errors of the nascent polypeptide in the cell may give rise to heterogeneity in the mature product. Expression of proteolytic enzymes can be difficult as accumulation of higher concentrations of proteases can be toxic to the cells. In addition, since cells are used as protein factories, protein engineering is generally limited to the twenty genetically encoded amino acids. Cell free synthesis has been explored to expand the repertoire of ribosomal synthesis to include non-coded amino acids in the expressed proteins (Short et al, 2000). However, these systems are faced with problems of low yields and are limited due to rather poor specificity of aminoacyl-t-RNA synthetases for non-standard amino acids (Balaram, 1992).

Chemical synthesis of proteins is an attractive procedure for protein production. Chemical methods provide the scope for unlimited variation of the covalent structure of a polypeptide for understanding the molecular basis of protein function. There are several examples of proteins synthesized solely by chemical means - prion protein, HIV-1 protease, IL-3, IGF, IL-8, RNase A and phospholipase A2 (Ball et al, 2001; Milton et al, 1992; Clark-Lewis et al, 1986; Li et al, 1983; Clark-Lewis et al, 1991). Notwithstanding the above successes, chemical synthesis methods are generally unsuitable for peptides longer than ~50 amino acids as they result in micro-heterogeneity and decreased solubility, thus leading to low synthetic yields. The peptide length limitation problem, however, could be partly overcome in a fragment condensation strategy (Muir and Kent, 1993).

Though the idea of segment condensation existed very early, the attempts were marred by poor solubility and coupling efficiency because of the maximal protection of the side chain groups. The improvements in chemical ligation methods were brought about by enhancing the chemo-selectivity of the reacting
Introduction

groups for specificity, so that unprotected peptide fragments in aqueous condition and neutral pH can be used to yield products with an unnatural moiety (hydrazone / thioether / thioester) replacing the native peptide bond at the site of ligation (Gaertner et al, 1992). The synthesis of backbone engineered HIV protease is one of the first examples that used this strategy (Schnölzer and Kent, 1992). However, such ester/ether bonds connecting the peptide fragments are generally not stable in extremes of pH and they are stereo-chemically different from the peptide bond. To counter these concerns, a new method of chemical ligation, referred to as the native chemical ligation, was developed. This chemo-enzymatic strategy involved three steps. In the first step, the reacting α-carboxyl group was esterified with a masked glycoaldehyde through trypsin-catalyzed reverse proteolysis reaction, followed by reaction of the amino group of the other peptide reactant with the unmasked aldehyde to form a thiazolidine or oxazolidine ring (Liu and Tam, 1994). Subsequently the formation of the peptide bond occurred through the intermolecular O–N-acyl migration. Dawson et al (1994) introduced a remarkable improvement in the native chemical ligation, where they used thioester activation of the α-carboxyl group of one of the peptide segments and the other containing cysteine residue at the amino terminus. Many moderate sized proteins such as interleukins, bone morphogenetic factor and barnase have been synthesised using this approach (Dawson and Kent, 2000). The speed and efficiency of ligation makes this method a choice for the rapid preparation of small cysteine rich proteins. However, this method is also restricted by polypeptide size and the technically demanding thioester derivatization of the protein fragments.

The chemical methods for polypeptide assembly alluded to above share mechanistic similarities with naturally occurring protein splicing reactions. Protein splicing is a post-translational process in which a precursor protein undergoes a series of intermolecular rearrangements that result in precise removal of an internal region (intein) and the ligation of the two flanking sequences (exteins). Mechanistically, the first step of protein splicing involves an N→S or N→O acyl shift in which the N-extein unit is transferred to the side chain -SH or -OH group of a conserved Cys/Ser/Thr residue, located in the amino terminus of the intein.
The entire intein is then transferred in a trans-esterification step to a second conserved Cys/Ser/Thr residue at the intein C-terminal boundary. The resulting branched intermediate is then resolved through a cyclisation reaction involving a conserved asparagine residue at the carboxyl terminal of the intein. The intein is, therefore, excised as a carboxy-terminal succinimide derivative. In the final step, an amide bond is formed between the two exteins following an $S\rightarrow N$ or $O\rightarrow N$ acyl shift, a step reminiscent of native chemical ligation (Perler et al, 1997; Gimble, 1998; Holford and Muir, 1998). The mechanistic similarities between native chemical ligation and protein splicing have lead to the development of a new strategy called the expressed protein ligation (Muir et al, 1998). Here, proteins are expressed as amino terminal fusion products with engineered inteins, which are then cleaved by thiols via an intra-molecular trans-esterification reaction to generate a $\alpha$-thioester derivatized recombinant protein. This protein can now react with another protein fragment containing a N-terminal cysteine residue through native chemical ligation. Expressed protein ligation has become the method of choice and a wide range of proteins have been synthesised using this procedure (Xu et al, 1999; Taylor et al, 2001; Valiyaveetil et al, 2002; Evans et al, 1999).

**Enzymatic peptide synthesis**

The enzymatic peptide bond synthesis was initiated by Bergmann (Bergmann and Fraenkel-Conrat, 1938) and his associates in 1938. It has been improved over the years by the developments in the theory of enzyme mechanism. Enzymatic methods offer several advantages over the chemical methods; e.g., reactions can be performed stereo-specifically and reactants do not require side chain protection. Further the need for the expensive protecting groups, organic solvents or hazardous chemicals can be avoided while using the enzyme-mediated approach. Proteases or proteinases (endo-peptidases) can be used for peptide bond formation. Proteases are divided according to their specificity of peptide cleavage and, therefore, can be selected for peptide synthesis on the basis of their selectivity of the amino acid residues on either side of the splitting point (primary and sub-site specificity).
Enzyme catalysed peptide bond synthesis could be carried out in a kinetic controlled or an equilibrium controlled process. In kinetically controlled synthesis, the key element is the rate of the product formation Vs the product hydrolysis. In the catalytic mechanism of a serine or a thiol protease, the first step is the formation of a covalently linked enzyme substrate (peptide) intermediate (acyl-enzyme complex). The hydrolysis of this intermediate is achieved by the nucleophilic attack of water. Likewise, aminolysis of the acyl enzyme intermediate may occur in the presence of an amine nucleophile. Since the peptide bond formation is competitive to hydrolysis, the product formation will depend on the relative rates of the two reactions.

This method has been extensively used in the synthesis of many small bioactive peptides by segment condensation strategy. Peptides such as bovine γ-melanocyte stimulating hormone, Met-enkephalin (Kullman, 1984; Widmer et al, 1981) have solely been synthesised by this methodology. Ester substrates can be very conveniently used as carboxyl components when the synthesis is catalysed by the serine or thiol proteases. In addition the reactions can be performed in the alkaline pH conditions where the secondary hydrolysis of the product is small owing to the minimal activity of peptidases in that range. Tougu et al (1993) demonstrated that free amino acids could be used as nucleophile in chymotrypsin catalysed acyl transfer reaction carried out under frozen conditions.

Although serine proteases are very useful in the reverse proteolysis, because of their esterase activity they favor hydrolysis over the aminolysis. Chymotrypsin and subtilisin showed higher aminolysis/hydrolysis ratio when their active site histidine residue was methylated. Nakatsuka et al (1987) showed that thio-subtilisin, a derivative of subtilisin in which the active site Serine 221 was chemically converted to cysteine, showed thousand fold shift in the preference of aminolysis over hydrolysis. On an improvement over this, Abrahmsen et al, (1991) showed that a subtilisin double mutant with catalytic Ser 221 converted to Cys (S221C) and Pro 225 converted to Ala (P225A) had 10-fold higher peptide ligase activity and at least 100-fold lower amidase activity compared to the single mutant thio-subtilisin. They further showed that the increase in the catalytic efficiency
could be attributed to the relieving of the steric pressure generated by the S221C mutation. Wu and Hilvert (1989) found that selenosubtilisin had a 14,000 fold shift in the preference for aminolysis. Subtiligase, an engineered variant of subtilisin BPN', was used for efficient ligation of esterified peptides in aqueous solution. This enzyme was used in the complete synthesis of ribonuclease A (RNase A) with unnatural catalytic residues. The fully active RNase A was produced from the stepwise ligation of six esterified peptide fragments varying in length from 12 to 30 residues long (Jackson et al, 1994).

The restrictive substrate specificity of enzymes and the risk of undesired proteolytic cleavage of substrates or products in segment condensation reactions were overcome by using substrate mimetics. The 4-guanidinophenyl esters of acylated aminoacids were found to be good substrates in trypsin catalysed peptide synthesis (Bordusa et al, 1997). The 4-guanidinophenyl ester (OGp) leaving group of these substrates binds to the active site of trypsin in a way that mimics the trypsin specific side chains of arginine. As a result, peptide coupling occurs irreversibly and independent of the Arg / Lys specificity. Bordusa and co-workers (Xu et al, 2001) further showed that the trypsin mutant Asp189Glu exhibited significantly higher preference for substrate mimetic peptides than peptides with normal Arg/Lys.

In the equilibrium controlled process, the peptide bond synthesis is the reverse of hydrolysis. Here peptide ligation utilizes conditions that shift the aminolysis/hydrolysis equilibrium towards aminolysis. The proteases catalyze forward (aminolysis) and reverse (hydrolysis) reactions and thus act only to hasten attainment of, but does not alter the final equilibrium. As discussed earlier, within the pH range at which most proteases are active (pH 5-9) in aqueous solutions, the thermodynamic barrier to synthesis is the ionisation of the carboxylic acid and amino groups to form the unreactive carboxylate and ammonium forms (Bongers and Heimer, 1994). Thus, the overall extent of synthesis is the product of the competing ionisation and aminolysis equilibria. Therefore, appropriate pH of the reaction is very important.
There are number of ways by which the synthetic yield of the reaction can be improved e.g., precipitation of the product, use of immiscible two-solvent systems etc. The thermolysin-catalysed synthesis of aspartame is a classical example of the utility of precipitation as a method for driving protease mediated ligations; the aspartame precursor is obtained in nearly quantitative yields as it precipitates almost totally as an insoluble salt. (Oyama and Kihara, 1984). In the two-phase methodology, water and a water immiscible solvent, such as ethyl acetate or chloroform is used. The two phase systems usually contain 2-5% of water and could be regarded as an emulsion of an enzyme aqueous solution in water-immiscible organic solvent. Here the enzyme in the aqueous phase will react with the substrate that has diffused from the organic phase and the ligation product formed by enzymatic reaction in the aqueous phase diffuses back in to the organic phase. This kind of partitioning system is suitable for water insoluble products. Additionally, the pKa increases and the effective pKb falls in the biphasic systems. Thus, at neutral pH, the relative amounts of the uncharged species of both amino and the carboxyl reacting groups will increase favoring proteo-synthesis. Martinek and Semenov (1981) successfully utilized biphasic systems for chymotrypsin-catalysed synthesis of dipeptides.

However, in general, both precipitation and biphasic systems cannot be used for the synthesis of water-soluble products such as short peptides containing basic amino acid residues or in the semisynthesis of higher molecular mass peptides and proteins. In these situations, the use of organic cosolvents is generally fruitful. Homandberg et al (1978) were the first to unequivocally establish the role of organic co-solvents in peptide ligation. They suggested that the main cause for the facilitation of ligation reaction was the increase in the pKa of the carboxyl group of the reactants or more appropriately, the decrease in the equilibrium constant for the transfer of a proton from the \( \alpha \)-carboxyl group to the \( \alpha \)- amino group. Thus the enthalpic barrier of peptide bond synthesis could be overcome by the use of organic solvents in the reaction milieu. Laskowski and coworkers (Homandberg et al, 1978) were the first to exploit the principle of shifting the equilibria from proteolysis to synthesis by the use of organic co-solvents to the
semisynthesis of proteins. They found that the protease-catalysed splicing of peptide bond in organic co-solvent was facilitated in protein/peptides systems that were held together by S-S bonds or non-covalent interactions. The enzymatic re-synthesis of proteolytically nicked proteinase inhibitors and RNase A splicing reactions (Mattis and Laskowski, 1973; Jering and Tschesche, 1976; Homandberg and Laskowski, 1979) in the presence of organic co-solvents were the early examples of the success of this approach (Homandberg and Laskowski, 1979). Subsequently, re-synthesis of fragment complementing systems of staphylococcal nuclease and somatotrophin (Komoriya et al, 1980; Graf and Li, 1981) was demonstrated. A very recent example in this category of reaction is the subtilisin-mediated regeneration of native triosephosphate isomerase from six or more proteolytic fragments in the presence of 60% acetonitrile (Vogel and Chmielewski et al, 1994). The concept of non-covalent reconstitution of two or more fragments to achieve the entropic activation and high effective concentration of both the reacting termini has been the core of the ligation of “fragment complementing” protein systems.

Komoriya et al, (1980) reported that enzymatic resynthesis could be carried out efficiently between peptide fragments even when these fragments were not non-covalently bound together. The N-terminal RNase A peptide 1-15 can be produced from the non interacting 1-10 and 11-15 when incubated with clostripain in the presence of S- protein (21-124). The S- protein fragment acted as a kinetic trap for the re-synthesized peptide (1-15), by binding to it to produce RNase S complex. The product accumulated as the 10-11 peptide bond was rendered resistant to proteolysis in this complex. In systems that lacked such product specific binding proteins, the use of product specific antibodies was suggested (Chaiken, 1981). The enzymatic synthesis of substance P was achieved using the product directed antibodies as molecular traps (Nyberg, 1988). An interesting variation of the product directed “molecular trap” concept is the “conformational trap” proposed to explain the ligation of the non-interacting segments of α-globin. The helix inducing organic solvents like 1-propanol, 2-propanol and trifluoroethanol were found to be effective in the V8 protease-catalysed semi-
Introduction

synthesis of \( \alpha \)-globin (Acharya et al, 1985; Sahni et al, 1989). The helix formation in the enzymically ligated products in the presence of helix enhancing co-solvents was attributed as the key driving force in the reverse proteolytic splicing of the unprotected and un-activated complementary fragments (17-30 and 31-40) derived from \( \alpha \)-globin (Roy et al, 1992). The viability of using conformational trap as a mechanism to drive proteolysis in reverse was amply demonstrated by using peptide fragments derived from the helical regions of thermolysin and ribonuclease (Kumaran et al, 1997; 2000).

Notwithstanding the mechanisms, reverse proteolysis could be exploited to engineer non-standard amino acids, fixed elements of three dimensional structure and spectroscopic labels in a protein. It turns out that the repertoire of 20 naturally occurring amino acids cannot provide complete molecular details of enzymatic reactions and macromolecular interactions. For example, based on a leucine to serine mutation in the case of nicotine acetylcholine receptor (nAChR) gating, it was hypothesized that the 10-fold increase in the opening of the ion channel was a consequence of increased polarity (Corringer et al, 2000; Arias, 2000). This interpretation appeared to be ambiguous since leucine to serine mutation brought about many structural changes in addition to the desired polarity change. The situation became clearer with the introduction of unnatural amino acids (for e.g., O-methylthreonine and isoleucine; allo-O methylthreonine and allo-isoleucine) which were considered isosteric to leucine and were capable of altering the polarity of the micro-environment. It was unambiguously demonstrated that it was the polarity change that caused enhanced opening of the channel (Kearney et al, 1996). In another study related to Ach – nAChR interaction, substitution of unnatural amino acid at the site in question established the involvement of an aromatic residue of the ACh in the binding with nAChR. Photoaffinity labeling studies had implicated the involvement of nine aromatic residues (tyrosine and tryptophan) near the agonist-binding site, but each of their individual roles was unknown. One of the possibilities was that the quaternary ammonium group of ACh was binding to one or more aromatic residues through a cation-\( \pi \) interaction (Dougherty, 1996; Ma and Dougherty, 1997). The incorporation of tyrosine and
phenylalanine derivatives (Nowak et al, 1995) at specific locations suggested that Trp 149 of the α-subunit was indeed binding to ACh through cation-π interaction (Zhong et al, 1998).

Introduction of fluorescent groups can allow static and dynamic investigations of protein function. The unnatural amino acid nitrobenzoxadiazole (NBD) fluorophore, was incorporated into the neurokinin-2 receptor (a G-protein coupled receptor that binds to tachykinin). The use of fluorescent labeled tachykinin enabled the fluorescence resonance energy transfer between the fluorescent agonist and the labeled receptor and provided valuable distance information relating the location of the agonist binding site to selected residues in the receptor (Trucatti et al, 1996). Hecht and coworkers (Arslan et al, 1997; Karginov et al, 1997) incorporated several analogs of phosphorylated amino acids into firefly luciferase and dihydrofolate reductase to study the importance of such post-translational events. Schultz and coworkers (Cook et al, 1995) incorporated caged serine at a crucial site in a self-splicing, intein system. Photo-decaging launched the splicing mechanism, establishing a key mechanistic role of the serine residue.

The introduction of a α-hydroxy residue in the backbone of the peptide is interesting as it can be used to probe the secondary structure, because this replacement will produce a backbone ester instead of the usual amide bond, thereby displacing the peptide NH group that is so crucial in establishing both the α-helix and β-sheet structures. A backbone ester cannot form α-hydrogen bonds of the sort found in a α-helix. In addition to this, it also makes the carbonyl a poorer hydrogen bond acceptor. Schultz and co-workers (Koh et al, 1997) were able to evaluate the relative importance of both effects on the stability of the helix.

The structural and thermodynamic characterization of 5-fluorotryptophan (5F-Trp) -substituted immunoglobulin binding domain B1 of streptococcal protein G (GB1) carried out by NMR and CD spectroscopy has shown that the introduction of fluorine does not perturb the global and local architecture and has no influence on the thermodynamic stability (Campos-Olivas et al, 2002). The substitution of trifluoroleucine for leucine in mellitin enhanced self-association.
and membrane binding activities (Niemz and Tirrell, 2001). The thermal stability of the leucine zipper coiled coil peptides was seen to be elevated by 13 degrees when the d-positions (abcdefg, heptad motif) of the peptides were replaced with trifluoroleucine (Niemz and Tirrell, 2001). Further, these peptides were shown to exhibit resistance to chaotropic denaturants, suggesting that fluorination of substructures in peptides and proteins may provide new means of increasing protein stability, enhancing protein assembly and strengthening receptor-ligand interactions (Tang et al, 2001). The enhanced stability in the coiled coil peptides derived upon the introduction of fluorinated amino acids presumably emanate from sequestration of the more hydrophobic trifluoromethyl groups from the aqueous solvent (Bilgicer et al, 2001).

**Macromolecular crowding: Concept and ramifications**

Crowding is characterized by the presence of high concentrations of macromolecular solutes in the medium. The macromolecules need not necessarily be of a single molecular species but all the molecules put together can constitute a crowded system. In such crowded environments, the common principle of molecules to occupy certain space and prevent other molecules from occupying the same space, termed volume exclusion (Steric repulsion), is highly magnified. Because of the high incidence of steric repulsion in a crowded media, the volume available for each molecule is limited. Thus any externally introduced molecule will also face a similar constraint of occupying least volume, or rather excluding least volume to other molecules. In such a scenario, if a molecular species that has a propensity to assume an organized state with lower volume occupancy were introduced, then the system (crowded media) would stabilise the compact organization of the introduced molecule. Physiological media is generally characterized by the presence of high concentration of macromolecules like proteins, lipids and carbohydrates. The typical concentration of proteins and RNA inside Escherichia coli is in the range of 300-400 g/l (Zimmerman and Trach, 1991). Taken together they occupy about 20-30% of the total volume. This volume is, thus, physically unavailable to the other molecules. The steric exclusion-
Introduction

generated volume exclusion effect has considerable energetic consequences that are not normally observed in a dilute solution. The thermodynamic and kinetic consequence of a crowded environment has been known for the last four decades (Laurent, 1963).

A) Excluded (pink and black) and available (blue) volume in a solution of spherical background macromolecules. The volume available for a test molecule of size comparable with background molecules. B) Excluded (pink) and available (blue) volume in a pore of square cross-section. volume available to a test molecule of size comparable with pore dimensions.

The equilibrium constants of reactions are ideally expressed in terms of effective concentrations (or thermodynamic activities) rather than in terms of actual concentrations. The ratio of effective concentration to actual concentration is termed as the activity coefficient. The dependence of activity coefficient of hemoglobin on actual concentration of hemoglobin exhibits a non-linearity. The effective concentration of hemoglobin was found to exceed its actual concentration by 10 fold at 200 mg/ml and by 100 fold at 300 mg/ml. Quantitative studies with cell extracts show that the high concentration of macromolecules within E.Coli cells produce large increases in values of activity coefficient (Zimmerman and Trach, 1991; Zimmerman, 1993). The activity coefficient of macromolecules inside such cells increases dramatically with its molecular mass. The activity
coefficient triples as the molecular mass approaches 3000 Da but exceeds two orders of magnitude as the molecular mass passes 50,000 Da (Ellis, 2001a).

The major effect of crowding on biochemical equilibria is to favor the association of macromolecules. Equilibrium constants for such reactions can be increased by as much as two to three orders of magnitude, depending on the relative size and shape of the reactants and products and on those of the background macromolecules. For example, the equilibrium constant for the association of a spherical monomer of molecular mass 40 kDa in to a homodimer is estimated to be 8 to 40 fold higher (depending on the specific volume of the protein) if the protein is expressed in crowded cells as E. Coli., compared to its value in a dilute solution. For a tetramer, the shift in equilibrium towards tetramerization is in the range of $10^3$-$10^5$. These striking effects arise from the reduction in the volume obtained when macromolecules bind to one another. As the number and size of the molecules in a solution increase, the less randomly they can be distributed. The configurational entropy of each of the macromolecular species becomes smaller, and its contribution to the total free energy of the solution increases with increasing concentration of macromolecules. The binding event is favored because the reduction of excluded volume or the concomitant increase in available volume decreases the total free energy of the solution. Thus, the most favored state is that which excludes the least volume to all the other macromolecules present.

Macromolecular crowding affects all those biochemical processes in which a change of excluded volume occurs; the collapse of a newly synthesised polypeptide chain in to compact functional protein, the unfolding of proteins induced by stress, the formation of oligomeric structures like fibrin, collagen and, multi-enzyme complexes in metabolic pathways, and the association of proteins into nonfunctional aggregates such as bacterial inclusion bodies, and the plaques in human amyloid diseases. The effects of crowding on the reactivity of the macromolecules is also implicated as one of the ways in which cells sense and respond to changes in their overall volume that is induced by osmotic alterations (Minton et al, 1992; Lang et al, 1998; Burg, 2000).
One of the consequences of crowding that is easy to grasp is the reduction in molecular diffusion rates (Zimmerman and Minton, 1993; Minton, 1983). The reduction applies to both large and small molecules, but is greater for large molecules. The measurement of diffusion coefficients (D) show that diffusion is reduced by three to tenfold in the crowded environments of the cells as compared to that in water (Ellis, 2001b; Luby-Phelps, 2000; Elowitz et al, 1999). The average time a molecule takes to diffuse a certain distance varies with 1/D so, if D is reduced tenfold, the molecule will take ten times longer to move this distance. Thus, the rate of any biochemical process that is diffusion-limited will be reduced by crowding, whether this process involves large molecules, small molecules or both types of molecule.

The effects of crowding on biochemical reaction rates are complex because although crowding reduces diffusion, it increases thermodynamic activities. The net result of these opposing effects depends upon the precise nature of each reaction. For a simple reaction of form A + B $\rightarrow$ AB*, crowding will reduce the overall rate by reducing diffusion. But, if the reaction rate is limited by the activity of AB*, crowding will increase the rate because crowding increases activity. The maximum rate possible for any multimolecular reaction is ultimately set by the encounter rate of the components, so even for transition-state-limited reactions, the rate will eventually fall as crowding increases. The effect of crowding on reaction rate thus depends upon the nature of the reaction and on the degree of crowding and, for this reason, it is necessary to measure the quantitative effects of crowding in all studies of macromolecular interactions if these studies are to be regarded as physiologically relevant (Minton, 1997).

Volume exclusion in the crowded media preferentially destabilizes either the reactants or products such that the most favored state of the system excludes the least volume to the predominant background species, thus affecting the equilibrium of many reactions (Minton, 1997). A compact globular macromolecular conformation excludes less volume to other macro-solutes than
extended or swollen conformations with the same solvent-excluding volume. And as oligomeric or polymeric aggregates of macromolecular subunits exclude less volume to other macro-solutes than equal number of isolated subunits, macromolecular crowding generally provides a non-specific force for compaction and association in crowded solutions (Minton, 1997). Ralston and colleagues have documented the enhancement of spectrin self-association in the presence of polyethylene glycol (Cole and Ralston, 1994) and dextran (Lidner and Ralston, 1995) by kinetically trapping oligomers and separating them from a stable monomer via native gel electrophoresis. It was recently demonstrated that monomeric fibrinogen can be induced to form dimers and possibly larger oligomeric species in the presence of sufficiently large amounts of BSA (> 40 g/l) without significant effect of the hetero-association of fibrinogen and BSA (Rivas et al, 1999). The same group has also found that under conditions inhibiting the formation of microtubules, tubilin may be induced to form soluble oligomers by the addition of the polysaccharide dextran T20 at sufficiently high concentration. The equilibrium constant for dimer formation was reported to increase by an order of magnitude in the presence of 20% (w/v) dextran. Kosk-Kosicka and coworkers (1995) used measurements of fluorescent energy transfer, catalytic activity and sedimentation equilibrium to establish that under condition normally favoring the formation of a less catalytically active monomeric state of detergent-solubilized membrane Ca\(^{2+}\) - ATPase, a more catalytically active dimer is formed in the presence of 10 % dextran (Kosk-Kosicka et al, 1995; Sackett and Kosk-Kosicka, 1996). \(^{13}\)C spin lattice relaxation measurements carried out on \(\gamma\)-crystallin over a wide range of concentrations have indicated that the equilibrium constant for the addition of a monomer to a growing oligomer of crystallin is enhanced by about an order of magnitude at the highest protein concentration (350 g/l) relative to the equilibrium constant at lower concentration (80 g/l) (Stevens et al, 1995). Zimmerman and Murphy have studied the condensation of Escherichia coli DNA as a function of the concentrations of the E. coli DNA-binding protein (HU) and polyethylene glycol (PEG) (Murphy and Zimmerman, 1994) or HU and extracts of E. Coli non-binding proteins (Murphy and Zimmerman, 1995). They found that
HU and PEG or concentrated non-binding protein act synergistically in promoting DNA condensation, and proposed that the presence of both HU and concentrated non-binding proteins mandate the condensation of DNA in intact bacterial cells (Zimmerman and Murphy, 1996). The enhancement of protein-protein and protein-DNA associations by volume exclusion effect has recently been used to clarify functional relationships in the T4 DNA replication and transcription complexes, both of which share three accessory proteins, gp45, gp44 and gp62. These three proteins appear to form a shuttle complex that is associated with DNA and can move along the helix in association with either the DNA polymerase (for replication) or the RNA polymerase (for transcription). The addition of 7.5% PEG12000 is reported to enhance interaction between gp45 and gp43 (DNA polymerase) by 10-20 fold, with a concomitant increase in the processivity of the enzyme even in the absence of gp44/gp62 (Reddy et al, 1993). Similarly, addition of 10% PEG 8000 has been reported to eliminate the requirements for 44/62 complex and an enhancer site in assembly of the translation complex (Sanders et al, 1994). Crowding agents like PEG 6000, have been shown to stabilise the structure of G-actin in the native state by increasing the denaturation temperature by 5°C (Tellam et al, 1983).

Folding events in proteins are also influenced by the presence of crowding agents. In GroEL, the molecular chaperon, that assembles from 56 kDa ATPase subunits into two stacked rings of seven subunits, each functional decamer re-assembles to 20% recovery after complete denaturation in urea. This recovery increases to 70% when Ficoll 70 (300 g/l) is added to the refolding buffer (Galan et al, 2001). The in vitro and in vivo binding of variety of soluble enzymes to elements of the cytoskeleton could be shown only in the presence of sufficient concentrations of inert polymers. Addition of permeabilised hepatocytes has been found to retard the dissociation of phosphoglucoisomerase, lactate dehydrogenase and aldolase (Agius, 1996). It was inferred from these findings that the binding of the above three enzymes to the cyto-skeletal elements of the cytoplasm is enhanced in the presence of PEG. It has been proposed that crowding will markedly slow the release of partially folded substrate from a molecular chaperone. 
following the ATP hydrolysis thereby providing additional opportunity for the partially folded substrate to acquire a native conformation before the chance encounter with other partially folded intermediates that might aggregate (Martin and Hartl, 1997; Ellis, 1997).

The present thesis

This thesis deals with structural and mechanistic aspects of the peptide ligation and protein semisynthesis catalyzed by proteases. The thesis is divided into three chapters. The influence of macromolecular crowding on protease-catalyzed synthesis of peptide bond has been examined with model peptide and protein systems in Chapter I. The role of substrate sub-site specificity in the context of “conformational trap” mechanism, alluded to earlier, is dealt with in Chapter II. Studies on semisynthetic analogs of ribonuclease A (RNase A) containing nonstandard fluorophenylalanine residues is described in Chapter III.