PROTEIN ENGINEERING OF SUBTILISIN BPN'

BY

PAUL GRAHAM THOMAS

Chemistry Department,

Imperial College of Science and Technology,

London, SW7 2AY.

A dissertation submitted to Imperial College of Science and Technology in candidature for the Diploma of Imperial College and to the University of London for the degree of Doctor of Philosophy.

Paul G. Thomas

December 1987

ABSTRACT

Protein engineering involves the use of detailed structural information and recombinant DNA techniques to manipulate protein structure and activity. This thesis describes the development of an experimental system for investigating the magnitude of electrostatic effects on the pH dependence of subtilisin BPN'.

The structural gene encoding subtilisin BPN' from *B. amyloliquefaciens* has been cloned in *E. coli* and expressed in a protease deficient strain of *B. subtilis*. Four mutant subtilisins have been constructed by oligonucleotide-directed mutagenesis of the gene cloned in M13. Targets for mutagenesis were identified from knowledge of the crystal structure of the enzyme and were designed to alter the local electrostatic environment of His-64.

The pK$_a$s of wild-type and mutant subtilisins were determined from the pH dependence of k_{cat}/K_m for the hydrolysis of a peptide substrate. The mutants Lys213→Thr and Asn62→Asp reduce the net positive charge near the active site and result in an increase in the derived pK$_a$ values of about 0.1 and 0.4 units respectively. The mutants Asp99→Ser and Asp36→Gln reduce the net negative charge near the active site and lead to a decrease in pK$_a$ by 0.4 and 0.2 units respectively. The observed pK$_a$ shifts were qualitatively as predicted from a simple electrostatic model where the low pH, protonated and inactive form of His-64 is stabilised giving an increased pK$_a$ or destabilised giving a decreased pK$_a$ as a result of alterations to the local electrostatic environment. The pK$_a$ shifts were maximal at low ionic strength and substantially masked at high ionic strength as expected for electrostatic interactions.

These results suggest that modifying the electrostatic environment of ionisable catalytic groups might be a general method for manipulating the pH dependence of enzymes used in industrial processes.

This experimental system will further our understanding of electrostatic effects in proteins and allow theoreticians to test the applicability and limitations of models of these effects.

ACKNOWLEDGEMENTS

To Mum and Dad

Succeed, Damn You !

CONTENTS

5

CHAPTER 2: MATERIALS AND METHODS

CHAPTER 3: CLONING AND EXPRESSION OF THE SUBTILISIN GENE

CHAPTER 8: ANALYSIS 2 - pK_a DETERMINATION

CHAPTER 9: FURTHER APPLICATIONS OF ENGINEERED SUBTILISINS

CHAPTER 10: SUMMARY AND PERSPECTIVES

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: GENERAL INTRODUCTION

1.1 Protein Engineering

1.1.1 A Definition of Protein Engineering

"Protein Engineering" is a term used to describe a new approach to the functional analysis of proteins, which has arisen from recent advances in DNA manipulative methods. Recombinant DNA technology has had a significant impact on protein biochemistry. The gene encoding essentially any protein can be cloned from the genome of a microbe, plant or animal, and when that gene is placed under the control of the appropriate regulatory signals and harboured in a suitable host, the gene product can be expressed at high levels. Proteins which previously could only be obtained in very small amounts, can now be purified in sufficient quantity for detailed study. Furthermore, site-specific mutagenesis (see chapter 4) allows any codon or codons to be substituted with any other to generate novel forms of the protein. This opens the way to direct experimentation in the relationship of protein structure to function and activity.

A useful working definition of protein engineering is: the use of detailed structural information and recombinant DNA methods to manipulate protein structure and activity systematically.

1.1.2 Applications of protein engineering

Protein engineering has had considerable influence on the study of enzyme function and mechanism. The traditional approaches to investigating structure/function relationships in enzymes have employed data from chemical modification studies and structural information derived from X-ray

16

crystallography to deduce the role of a particular amino acid residue in enzyme catalysis.

The implicit assumptions with these approaches are that: because modification of a specific residue affects (usually destroys) activity, then that residue must be functionally essential; or alternatively, because a specific residue is at the active-site then it must be important. Conclusions based on such studies can only be tested if selected structural changes can be made to the enzyme. Naturally occurring mutants or mutants derived from random mutagenesis procedures provide a set of variant proteins for analysis, but the number and type of such mutants is normally rather limited.

The weakness of chemical modification studies is that they are rarely absolutely specific for the residue of interest, and some degree of modification at other positions is likely. The nature of the modification may also be rather drastic and lead to concern that gross structural perturbations, caused by bulky derivatives, are responsible for the changed activity. Furthermore, generally only those amino acid side chains which are exposed to the solvent or are highly reactive are amenable to modification. These problems can be largely avoided with protein engineering since conservative or even isosteric substitutions are possible at any position in the protein molecule, thereby reducing the risk of structural perturbations. Furthermore, the enzyme is not exposed to the drastic conditions usually involved in chemical modification.

Thus, protein engineering provides the enzymologist with the ability to construct any number or type of mutations necessary to dissect various aspects of enzyme function such as the binding of substrate or regulatory molecules or the mechanism of catalysis. This flexibility makes protein

17

engineering a very much more delicate probe of enzyme function and this new approach will render chemical modification studies largely obsolete.

Protein engineering has already contributed some significant advances to the understanding of enzyme function. Tyrosyl-tRNA synthetase (YTS) from *Bacillus stearothermophilus* is a dimeric enzyme which catalyses the aminoacylation of tRNA^{Tyr} by a two step reaction. This involves the formation of enzyme bound tyrosyl adenylate from the substrates tyrosine and ATP and the subsequent transfer of the tyrosyl adenylate to tRNA^{Tyr}. Extensive kinetic studies on wild-type and mutant forms of the enzyme, constructed using protein engineering, have allowed the interaction energy between the enzyme and its substrate to be measured along the entire reaction pathway. The results of these experiments show directly that catalysis of the first step of the reaction is achieved by stabilisation of the transition state (Fersht *et al.*, 1986). The apparent strengths of hydrogen bonds and salt-bridges have also been determined (Fersht, 1987; Ward *et al.*, 1987).

The mechanism of tRNA charging has been studied in a similar fashion and it has been shown that each dimer of YTS uses the same active site repeatedly. The second active site shows no activity and the enzyme functions asymmetrically in solution. This contrasts with the structure of the crystalline enzyme where the dimeric enzyme is symmetrical. These results have provided considerable insight into the mechanism of half-sites activity (Ward and Fersht, 1988). The direct study of an enzyme in this detail was not possible before the development of protein engineering.

The results and general principles which are emerging from the widespread application of protein engineering to other enzyme systems have been reviewed (Leatherbarrow and Fersht, 1986; Shaw, 1987; Knowles, 1987).

The commercial potential of protein engineering has been noted (Ulmer, 1983; Rastetter, 1983). Several enzymes are used in industrial bioprocesses, for example α-amylase and glucose isomerase, which are used in the production of high fructose syrups from corn starch, are the two largest volume industrial enzymes. Evolution has optimised enzymes for their natural *in vivo* roles and not for *in vitro* applications invented by man. The process conditions under which these enzymes are required to function are decidedly non-physiological and may range from extremes of *p*H and temperature or involve exposure to chemicals. Such conditions clearly need quite robust enzymes. Unfortunately many of those enzymes in use are fairly labile and the costs of the process are therefore quite considerable. The ability to modify these enzymes in order to improve their stability under process conditions or perhaps allow them to utilise novel substrates would clearly have an enormous commercial impact.

Several proteins have biomedical applications, some are even used directly as pharmaceuticals. Examples include insulin, growth hormone, interferons, factor VIII etc. Antibodies are also useful as diagnostic tools. It has been speculated that protein engineering may be used in the development of second generation pharmaceutical or diagnostic products based on these molecules.

1.1.3   Prerequisites for protein engineering

Protein engineering is a multidisciplinary approach to the study and manipulation of protein function. It is generally considered to be a fusion of three areas of science; molecular biology, protein biochemistry and enzymology, and X-ray crystallography.

Accordingly there are three prerequisites for any protein engineering study.

19

Firstly, the gene encoding the protein of interest must be cloned and expressed in a suitable host/vector system to allow the construction of mutants by site-specific mutagenesis, and their subsequent expression so that sufficient material can be purified for analysis.

Sensitive methods for the quantitative analysis of the effects of the mutation on the catalytic or functional properties of the molecule must also be available, for example enzyme kinetics or direct physical methods such as NMR.

Protein engineering also requires detailed structural information, such as a crystal structure derived from X-ray diffraction studies. Ideally this would be at high resolution to allow individual residues to be identified (ie 2.5 Å or better). This information is essential to provide a basis for the design of mutations and also in the interpretataion of experimental data. It is sometimes argued that the structure of the mutant proteins should also be determined, to establish whether any structural changes have occurred as a result of mutation which might be responsible for the altered properties of the molecule.

Based on the requirements described above, the subtilisins have been selected as a suitable model system for an extended protein engineering study. The subtilisins are an attractive model since the crystal structures of several subtilisins have been solved to high resolution. This family of enzymes has been studied for many years by conventional approaches such as chemical modification and kinetics, and methods are available for the quantitative analysis of mutants constructed by site-specific mutagenesis. The only missing factor is the gene encoding a subtilisin. Recombinant DNA technology is now sufficiently advanced to predict that cloning and expression of the gene should not present any insurmountable problems.

1.2 The subtilisins

1.2.1 Introduction

The subtilisins (E.C. 3.4.4.16) are a family of extracellular serine proteases produced by species of *Bacillus* prior to sporulation. Subtilisins are synthesised towards the end of exponential phase growth and accumulate in the culture medium during stationary phase together with a variety of other hydrolytic enzymes including amylases, nucleases, proteases and esterases (Priest, 1977). Most of these exoenzymes, including the subtilisins, are believed to have a nutritional role. Their expression is induced in response to depletion of available nutrients. Production of a battery of hydrolytic enzymes allows the cells to scavenge nutrients from the environment (Priest, 1977).

All the subtilisins exhibit maximal activity at high pH (8-11), hence their alternative name, the alkaline proteases. Since their discovery in 1947 (Linderstrom-Lang and Ottesen), this group of enzymes has been the subject of intense study and an extensive body of literature has been published. Table 1.1 lists all the subtilisins identified to date and the *Bacillus* strains from which they are isolated.

The chemical and physical properties, three-dimensional structure, kinetics and mechanism of action of the subtilisins have been reviewed (Ottesen and Svendsen, 1970; Markland and Smith, 1971; Kraut, 1971 and 1977; Philipp and Bender, 1983). An overview of our current understanding of this family of enzymes is presented below.

Table 1.1: Species of *Bacillus* producing subtilisins of known amino acid sequence

Bacillus Species	Subtilisin	Reference
B. amyloliquefaciens	BPN'/Novo	Olaitan *et al.*, (1968)[1]
		Wells *et al.*, (1983)[2]
B. licheniformis	Carlsberg	Smith *et al.*, (1968)[1]
B. amylosacchariticus	amylosacchariticus	Kurihara *et al.*, (1972)[1]
B. subtilis DY	DY	Nedkov *et al.*, (1983)[1]
B. subtilis 168	168	Stahl and Ferrari (1984)[2]
B. mesentericus	mesentericus	Svendsen *et al.*, (1986)[1]

Notes to Table 1.1:
(1) Determined by direct protein sequencing.
(2) Predicted from the nucleotide sequence of the cloned gene.

1.2.2 Primary structure and homologies

The primary structures of six members of the family have been determined either by direct protein sequencing or by deduction from the nucleotide sequence of the cognate gene. The subtilisins from *B. amyloliquefaciens*, *B. amylosacchariticus*, *B. subtilis and B. mesentericus* consist of a single chain of 275 amino acids which contains no cysteine residues so no disulphide bonds can form in these enzymes. The enzymes exhibit quite marked temperature stability and resistance to denaturation by urea or organic solvents (Markland and Smith, 1971) and detergents (Richelli *et al.*, 1982).

The subtilisins from *B. licheniformis* and *B. subtilis* DY consist of 274 amino acids, the conserved Asn residue at position 56 in the other enzymes having been deleted.

The enzyme family is highly homologous with identical amino acids appearing at 162 positions in all members of the family (i.e. 59% amino-acid identity). A very large proportion of the non-identical positions contain conservative amino acid substitutions. There is however absolutely no detectable amino acid homology with the mammalian pancreatic serine proteases of the trypsin/chymotrypsin family (Smith *et al.*, 1966).

There are 85 amino acid differences between subtilisin BPN'/Novo and subtilisin Carlsberg including the deletion at position 56. Comparison of the three-dimensional structures of subtilisins BPN' and Carlsberg shows that all but one of the differences occur in exterior chain segments (Kraut, 1971). The only internal substitution is highly conservative, Ile at position 31 in Subtilisin BPN' being replaced by Leu in Subtilisin Carlsberg. This suggested that the three-dimensional structures of both these enzymes and indeed all the members of this family must be nearly identical, and this

23

has been subsequently comfirmed by X-ray crystallography (see later).

1.2.3   Crystal structure of the subtilisins

The crystal structure of Subtilisin BPN' (various inhibited forms) has been determined and refined to 2.5 Å resolution (Drenth and Hol, 1967; Wright et al., 1969; Drenth et al., 1972), and more recently to 1.8 Å resolution (R. Bott, personal communication).   The structure of Subtilisin BPN' complexed with streptomyces subtilisin inhibitor (SSI) from *Streptomyces* species has been determined to 2.6 Å resolution (Hirono et al., 1984), and with chymotrypsin inhibitor-2 (CI-2) from barley to 2 Å resolution (McPhalen et al., 1985 and 1987).   The structure of Subtilisin Carlsberg complexed with the leech inhibitor, eglin C, has been solved and refined to 1.2 Å resolution (Bode et al., 1986).

The structures of Subtilisin BPN' and Subtilisin Carlsberg are almost identical, a least squares imposition of the 274 structurally equivalent α-carbon atoms gives a root-mean squared deviation of only 0.53 Å (McPhalen et al., 1985).   The high degree of sequence conservation between subtilisins suggests that the three-dimensional structures of the other enzymes in the family will be very similar to that of Subtilisins BPN' and Carlsberg.

The subtilisin molecule is approximately spherical with a diameter of about 42 Å.   The active site and substrate binding groups are located in a very shallow cleft which extends about 18 Å across the surface of the enzyme (Morihara et al., 1969).   The molecule contains 8 segments of right handed α-helix, all running approximately parallel to one another ($\pm$ 15° to the common direction) and all in the same N to C-terminal orientation (Kraut, 1971).   Approximately 30% of the amino acids are involved in α-helix

24
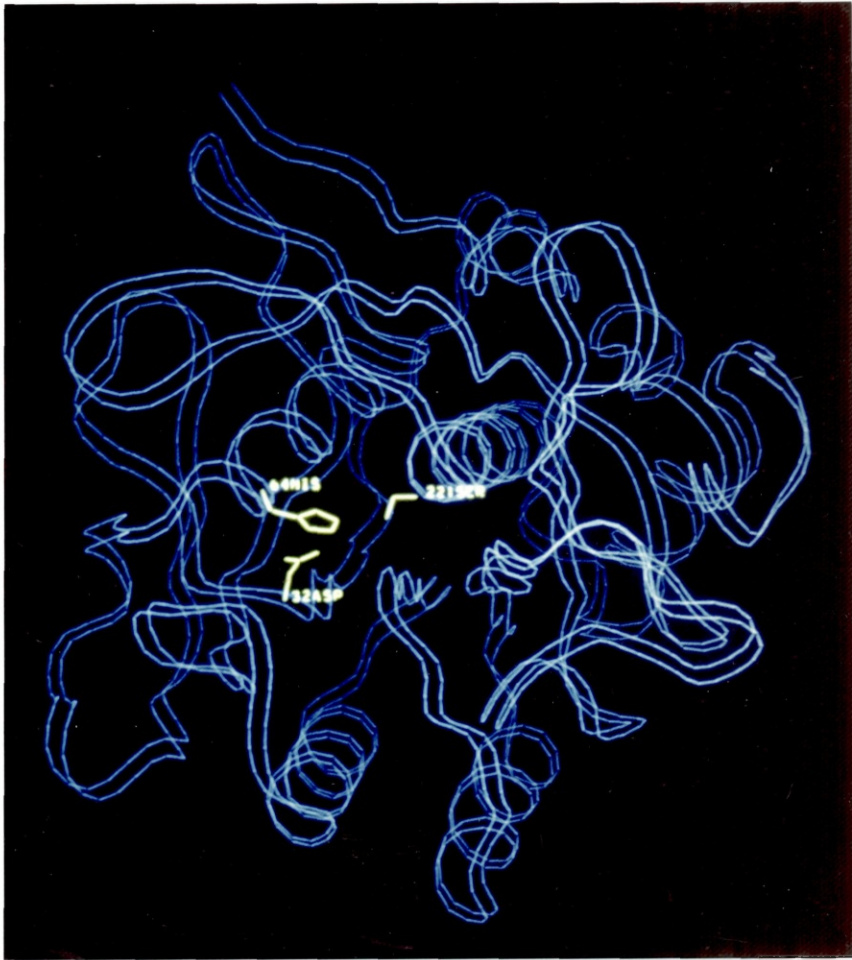
Figure 1.1   The α-carbon backbone of Subtilisin BPN'.

formation. Figure 1.1 shows the α-carbon back-bone structure of Subtilisin BPN' viewed along the common axis of the α-helices. The active-site residues are indicated (see Section 1.2.5) and the substrate binding cleft runs from top left to bottom right across the face of the enzyme in view.

The three-dimensional structure of the subtilisins is completely different from the three-dimensional structure of the mammalian serine proteases of the trypsin family (Wright *et al.*, 1969). However, the catalytically functional groups in Subtilisin BPN' (i.e. Ser 221, His 64, and Asp 32) are arranged in a similar geometry to the catalytically functional groups in the active site of trypsin and related serine proteases (Alden *et al.*, 1970).

1.2.4 Substrate binding

The three-dimensional structure of the complexes formed between subtilisins and small, naturally occuring protein inhibitors give a detailed picture of the interactions between the enzyme and substrate in the Michaelis complex (Hirono *et al.*, 1984; Bode *et al.*, 1986; McPhalen *et al.*, 1985).

The notation of Schechter and Berger (1967) is applicable to the subtilisins. The amino acids in the acylating segment of the substrate (see Section 1.2.5), that is those residues extending from the scissile bond towards the amino terminus, are designated P_1, P_2, P_3 etc. The residues in the leaving group segment extending from the cleaved bond towards the carboxy terminus, are designated P_1', P_2', P_3' etc. The corresponding sub-sites on the enzyme which interact with the substrate are designated S_1, S_2, S_3 and S_1', S_2', S_3' according to the segment of substrate with which they interact. Figure 1.2 shows this schematically.

The acylating portion of the substrate forms an anti-parallel β-sheet with a

26
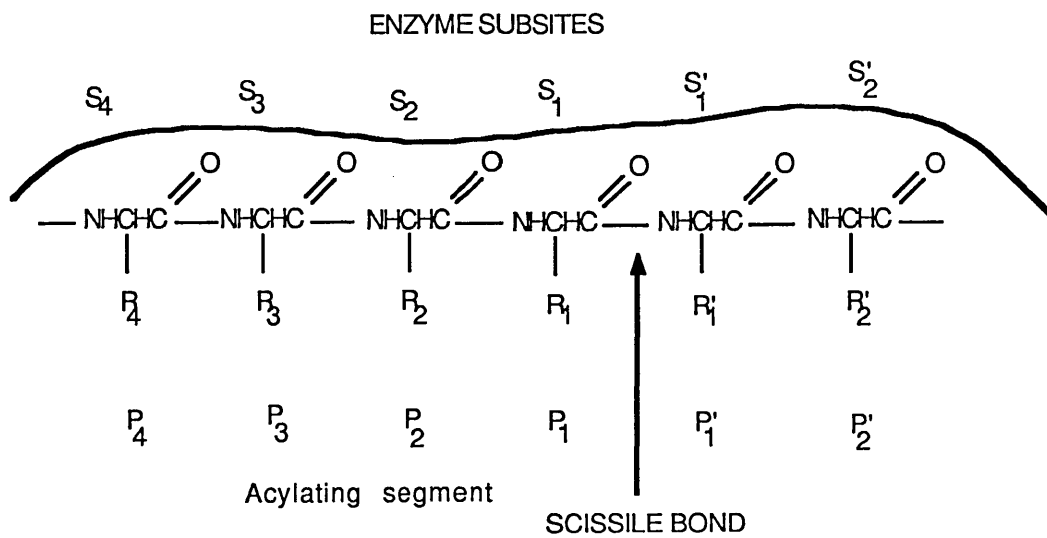
Figure 1.2:
Schematic diagram of the interaction between subtilisin and a peptide substrate

segment of the subtilisin molecule consisting of Ser 125-Leu 126-Gly 127, which constitutes at least part of the sub-sites S_1, S_2 and S_3 on one side of the substrate binding crevice. Bode *et al.,* (1986) and Hirono *et al.,* (1984) have shown that the interaction could extend further still since the segment containing residues Gly 101 to Tyr 103 is able to extend the anti-parallel β-sheet with the substrate main chain to form a 3 stranded intermolecular β-sheet. This region has been designated as the S_4-S_6 sub-sites (Hirono *et al.,* 1984). Thus, one side of the substrate binding crevice is composed of the planar peptide links in the segment containing residues 125-127 and constitutes sub-sites S_1-S_3. The other side of the substrate binding crevice is an irregular surface composed of the side chains and main chains of the segment Ala 152-Ala 153-Gly 154. The substrate binding site is quite large and therefore can accomodate most amino acid side chains. This explains the rather broad side-chain specificity at P_1, typical of the subtilisins, which although they prefer aromatic or other hydrophobic side-chains at this position will also accept polar side chains. In comparison, the S_1 crevice in chymotrypsin is well defined and best accepts planar aromatic side-chains at the P_1 position (Matthews *et al.,* 1967; Kraut, 1977).

Subtilisin appears to have an additional side-chain binding crevice at S_4 which can accommodate an aromatic residue by movement of the side-chain of Tyr 104. This may account for the ability of subtilisin to hydrolyse substrates with an abnormal (i.e. non-hydrophobic) P_1 residue, due to the over-riding specificity for the aromatic residue at P_4 (Kraut, 1977).

Difference fourier analysis of Subtilisin BPN' complexed with various polypeptides and model building studies have shown that, on the leaving group side, the substrate main chain appears to form a short stretch of anti-parallel β-sheet with the main chain segment of the enzyme at

28

Asn 218-Gly 219. The substrate side-chains make contact with the enzyme in the region of Phe 189 and Tyr 217 (Robertus *et al.*, 1972). This presumably constitutes the S_1' binding site, however there is no evidence to suggest that interactions between the enzyme and substrate on this side of the scissile bond can affect substrate specificity. Examination of the structure of subtilisin complexed with the various naturally occuring inhibitors should reveal whether or not the substrate makes further contacts with the enzyme.

1.2.5   Catalytic mechanism

Subtilisins catalyse the hydrolysis of peptide and synthetic ester substrates. The reaction is an acyl transfer and proceeds via an acyl-enzyme intermediate. The catalytically functional groups in all subtilisins are the conserved residues Ser 221, His 64 and Asp 32. Histidine 64 functions as a general base during catalysis, accepting a proton from Ser 221 as it forms an ester linkage with the substrate carbonyl carbon. Catalysis is achieved by the specific binding and stabilisation of the tetrahedral transition state characteristic of acyl-transfer reactions (Robertus et al, 1972). The reaction pathway for a peptide substrate is shown in Figure 1.3.

The substrate is specifically bound to the enzyme, as described above, to form the Michaelis complex. Formation of the acyl-enzyme intermediate proceeds via a tetrahedral transition state complex during which a covalent bond is formed between the substrate carbonyl carbon and the reactive serine Oγ. The proton from the serine hydroxyl is transferred to the imidazole Nϵ2 atom of His 64, and then to the leaving group as the transition state collapses to generate the acyl-enzyme intermediate. Deacylation of the enzyme is essentially the reverse of acylation, the acylating portion of the substrate being transferred to a water molecule,

Figure 1.3: Reaction scheme for the hydrolysis of peptide substrates by subtilisin
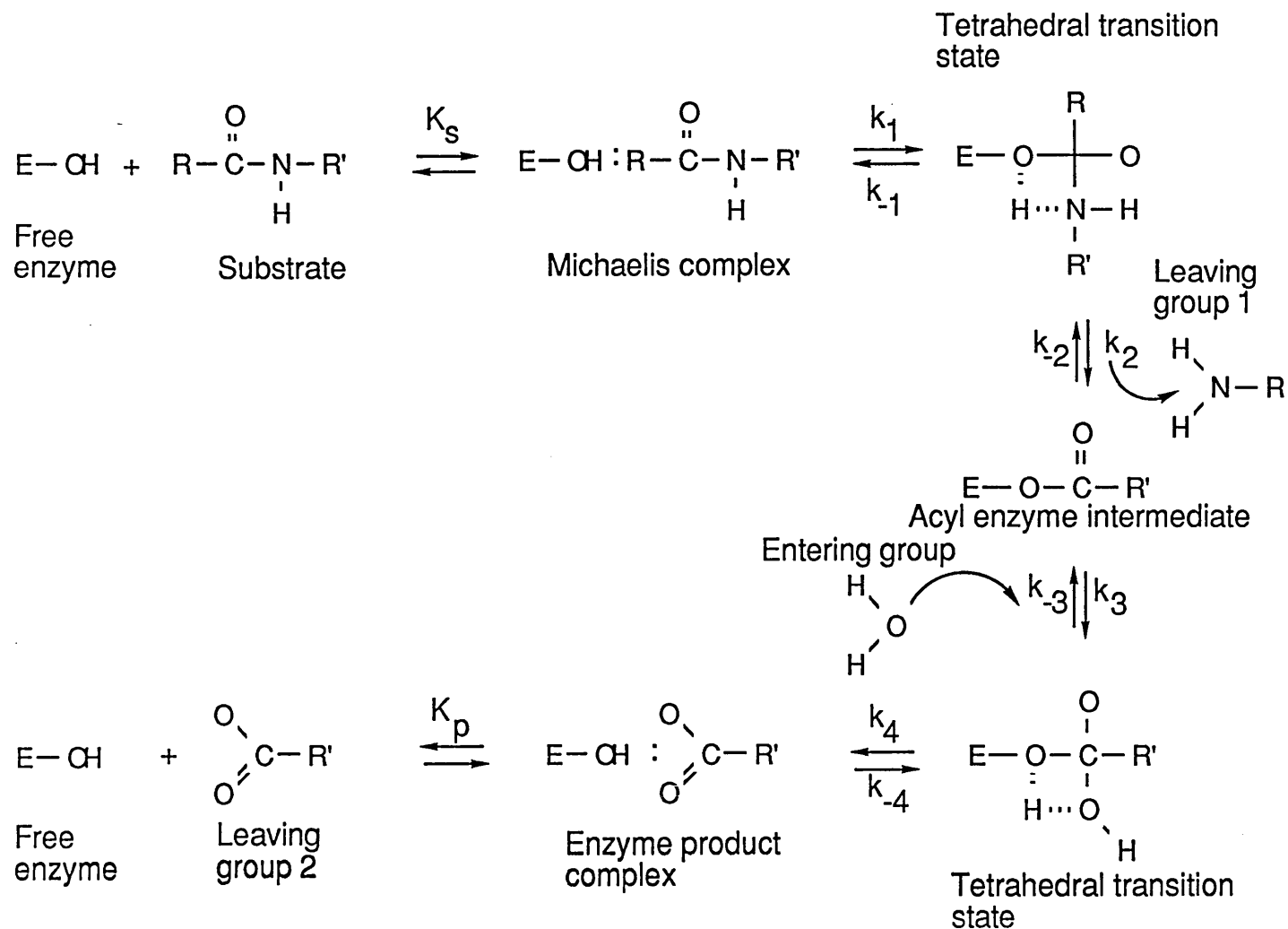
again via a tetrahedral transition state.

Stabilisation of the charged transition state is facilitated by the ability of the negatively charged oxyanion to form good hydrogen bonds with the back-bone -NH of Ser 221 and with the side-chain amino group of Asn 155. These bonds can only form in the tetrahedral configuration and do not occur in the Michaelis complex or the acyl-enzyme intermediate where the carbonyl oxygen is in a planar configuration (Robertus *et al.*, 1972). The site of oxyanion binding is refered to as the oxyanion hole.

The activity of the enzyme is dependent on the charge state of the active site histidine. Catalysis can only occur at high pH when the histidine is uncharged and able to receive a proton from Ser 221. At lower pH values, His 64 becomes protonated and, as it can not accept a further proton from Ser 221, is inactive. Thus, the catalytic activity of subtilisin varies with pH, reflecting the ionisation of this residue. The pK_a of the active site histidine can be determined accurately by kinetics.

Earlier ideas that the triad of catalytic residues in serine proteases constituted a "charge relay mechanism" (Blow *et al.*, 1970), which functioned to enhance the nucleophilicity of the reactive serine, have been superceded. The His-Asp couple is now regarded as providing a binding site for the proton which is in the process of being transferred between the serine and the leaving or entering group (Kraut, 1977). The role of the buried aspartate of which the $O\delta2$ is hydrogen bonded to the $N\delta1$ of the histidine imidazole, seems to be to maintain the correct geometry of the histidine with respect to the serine. Low resolution structural data suggested that the $N\epsilon2$ of His 64 is H-bonded to the Ser 221 hydroxyl, however the distance between these two side-chains is probably too great for a real bond to exist (Kraut, 1977).

The primary and three-dimensional stuctures of the subtilisins bears little relationship with those of the trypsin family of serine proteases. The two enzyme families are clearly the products of two evolutionarily unrelated genes. Surprisingly both enzyme families share the same catalytic triad of Ser, His and Asp residues. Furthermore, these are arranged in an almost identical geometry and the mechanism of catalysis, by proton transfer and stabilisation of the transition state oxyanion, is the same in both enzymes. In addition the interactions between the enzyme and the acylating portion of the substrate via a short stretch of anti-parallel β-sheet is very similar in both enzyme families. It has been suggested that the similarity of both substrate binding and the mechanism of catalysis between the subtilisins and the trypsin family of serine proteases is an example of convergent evolution. That is, two apparently unrelated gene products have evolved to use the same mechanism of catalysis (see Fersht, 1985).

1.2.6   Kinetics of the subtilisins

The steady state kinetics of the subtilisins have been extensively reviewed (Philipp and Bender, 1983). Aspects relevant to this study are discussed later (see Section 7.2).

1.3 Subtilisin: A protein engineering paradigm

1.3.1 Introduction

The availability of very high resolution crystal structures combined with a wealth of literature regarding the physical, chemical and kinetic properties of this enzyme, makes subtilisin an attractive model system for protein engineering. The structural gene encoding subtilisin BPN' from *Bacillus amyloliquefaciens* has proven relatively easy to clone and express and this

32

has been achieved in at least three different laboratories. All the prerequisites given above are available to allow an extensive investigation of this enzyme by protein engineering.

Several speculative articles were published during the early days of protein engineering in which the authors speculated on the potential analytical and commercial applications of this approach (Ulmer,1983; Rastetter,1983). Properties of proteins identified as suitable targets for protein engineering include:

kinetic parameters, (k_{cat} and K_m)

substrate specificity

thermostability

chemical stability

pH optimum

allosteric properties

size and sub-unit structure

cofactor requirements

stability/activity in non-aqueous systems

suitability for immobilisation

Many of these targets have been addressed through studies on several model systems. An extensive protein engineering exercise on subtilisin, is already underway in several laboratories.

1.3.2 An oxidation resistant subtilisin

Subtilisin is remarkably resistant towards denaturing conditions, however it is susceptible to inactivation by chemical oxidants, for example H_2O_2 (Svedson, 1976). Peptide mapping studies (Stauffer and Etson, 1969),

demonstrated that loss of activity is due to oxidation of the invariant Met residue at position 222 (adjacent to the catalytic Ser-221), to give the methionine sulphoxide. Estell *et al.*, (1985) anticipated that replacement of Met-222 should produce a subtilisin variant that would be more resistant to oxidative inactivation. Since it was not possible to predict the optimal replacement for Met-222 to achieve this end, all 19 possible replacements were constructed using "cassette mutagenesis" (Wells *et al.*, 1985). Mutant enzymes were produced in a protease deficient host to avoid contamination with endogenous subtilisin.

Mutants containing Leu, Val or Ile, which are normally considered functionally homologous (ie. hydrophobic) with Met, were non-optimal for the maintenance of enzyme activity. The specific activity of these mutants being reduced to around 2-12% of wild-type. The most active mutants contained small amino acids (Gly, Ser, Ala) at position 222. These retained 30-50% of the wild-type activity. Charged or bulky replacements also gave very low activities. With regard to oxidative stability the Ser, Ala and Leu mutants were completely resistant to 1 M H_2O_2 over a period of 1 hr whereas the wild-type enzyme was completely inactivated under the same conditions. Interestingly the Cys-222 mutant had a greater specific activity than wild-type due to an increase in k_{cat}. However this mutant was still sensitive to oxidation and was rapidly inactivated. The Ala-222 mutant retained the highest activity of the mutants (50% of wild-type) and had dramatically improved oxidative stability. Variant subtilisins of this type may be useful in washing powder formulations containing bleaching agents.

The Met at position 356 in α-1 anti-trypsin is also sensitive to oxidation and has been replaced with a Val to produce a variant with similarly improved oxidation resistance and which may have therapeutic value in the treatment of emphysema (Rosenberg *et al.*, 1984; Courtney *et al.*, 1985).

This approach seems to be generally useful in improving the oxidative stability of proteins.

1.3.3 Subtilisins with altered substrate specificity

The subtilisins have a broad substrate specificity but do exhibit a preference for bulky hydrophobic groups such as Tyr and Phe at the P_1 position in the substrate (i.e. to the N-terminal side of the scissile bond). The side-chain of the P_1 amino acid extends into a hydrophobic cleft (S_1 subsite) which consists of three segments of the main chain. The amino acids Ser-125, Leu-126 and Gly-127 form one wall of the cleft and Ala-152, Ala-153 and Gly-154 the other wall. The amino acids Val-165, Gly-166, Tyr-167 and Pro-168 form the back of the cleft. In the wild-type enzyme the side chain of Gly-166 extends into the bottom of this cleft.

Estell *et al.*, (1986) have investigated the effect of amino acid substitutions at position 166, on the P_1 side-chain specificity of the enzyme. All 19 amino-acid substitutions at this position were constructed by "cassette mutagenesis" (Wells *et al.*, 1985), and the mutant enzymes expressed and purified. The substrate specificity of the various mutants was probed using a series of synthetic peptide substrates of the form succinyl-L-ala-L-ala-L-pro-L-X-*para*-nitroanilide, which differ only in their P_1 residue (X). The kinetic parameters, k_{cat} and K_m for several substrate and mutant combinations were derived from steady-state experiments.

The clearest trend emerged from analyses of mutants with non-ionic substitutions at position 166. Mutations which alter the side-chain volume and hydrophobicity of the residue at position 166 have a significant effect on enzyme specificity through hydrophobic and steric effects.

Increasing the volume of the side-chain at position 166 results in a shift in P_1 specificity from large to small side-chains. Furthermore, the catalytic efficiency towards substrates with small hydrophobic groups at the P_1 position were increased up to 16 fold by increasing the side chain volume of the hydrophobic substitutions. For example, where Ala is the P_1 residue in the substrate, k_{cat}/K_m increases approximately 10 fold on mutation from Gly-166 through Ala, Met, Phe, Trp, Ser, Thr, Val, Leu and Ile. There is also a 10 fold increase in k_{cat}/K_m for P_1 = Met across the series of mutants from Gly to Leu. k_{cat}/K_m for the Leu-166 mutant with P_1 = Tyr is as good as the wild-type enzyme (Gly-166) with its best substrate where P_1 = Tyr.

Although changes in substrate specificity due to hydrophobic effects are substantial, they are considerably outweighed by steric effects since in cases where the combined volume of the replacement and P_1 side-chains exceeds the volume of the cleft (approximately 160 Å^3), then a dramatic decrease in k_{cat}/K_m is observed, for example with Ile-166 and P_1 = Tyr catalytic efficiency is about 1000 fold less than wild-type for the same substrate. The specificity of the Ile-166 mutant resembles elastase rather than chymotrypsin.

Mutations in the P_1 binding cleft of subtilisin produce large changes in specificity for the P_1 amino-acid. Some mutant enzymes have increased catalytic efficiencies towards particular substrates. For some optimal enzyme/substrate pairs (eg. Ile-166 and P_1 = Met), k_{cat}/K_m exceeds that of wild-type enzyme with its best substrate (P_1 = Tyr).

Substitutions at position 166 involving ionic amino acids also have significant effects on substrate specificity, but are less straight forward to interpret. Replacing Gly-166 with Glu or Asp causes a reduction in

k_{cat}/K_m for su-AAPF-pN and a dramatic increase in k_{cat}/K_m for bz-VGR-pN, whereas values for the isosteric substitutions Asn and Gln are similar to wild-type (Wells *et al.*, 1987). Interpretation of these results is complicated by the fact that suAAPFpN is negatively charged and will clearly experience electrostatic repulsion by acidic groups at position 166 in addition to effects localised to the substrate binding cleft.

1.3.4 Thermostability: The effect of engineered disulphide bonds

Most enzymes are inactivated rapidly above 55 °C , but enzymes from thermophilic organisms retain their normal activity for prolonged periods at elevated temperatures. It would be immensely useful to be able to engineer enhanced thermostability into commercially important proteins.

Disulphide bonds are important in the conformational stability of proteins (reviewed by Anfinsen and Sheraga, 1971; Thornton, 1981). One strategy which has been employed to achieve increased thermostability is the introduction of additional disulphide bonds.

Disulphide bonds have been introduced into T4 lysozyme (Perry and Wetzel, 1984) and Dihydrofolate reductase (Villafranca *et al.*, 1983). In both cases the mutant enzymes exhibited improved thermostability compared with the wild-type enzymes. The mutants were constructed by engineering single site mutations which introduced a Cys residue close to a pre-existing Cys. Since the mutants were produced in *E. coli* the disulphide bonds could only be formed by chemical oxidation *in vitro*.

Members of the subtilisin family contain no Cys residues and therefore constitute a good model for engineering disulphide bonds into a protein. Additionally, since the enzyme is secreted, the bonds should form *in vivo*.

Disulphide bonds have been introduced in subtilisin at a position which is distant from the functional regions of the enzyme and such that the geometry of the bond should conform well to naturally occuring disulphides and not cause distortion of the main chain.

The double mutant enzyme Ser24→Cys/Ser87→Cys (SC24/SC87), is secreted normally and has the same kinetic parameters as wild-type subtilisin. Polyacrylamide gel electrophoresis in the presence and absence of reducing agents confirms that the bond does form *in vivo* (Wells and Powers, 1986). Formation of the disulphide bond has also been confirmed by X-ray crystallography. This mutant appears to be marginally more stable to heat denaturation and consequent autolysis than does wild-type subtilisin. Differential scanning calorimetry of the same mutant constructed elsewhere (Pantoliano *et al.*, 1987) reveals that the T_m for the transition between the folded and unfolded states is 62.0 °C compared with 58.9 °C for wild-type. On reduction the T_m drops to 56.2 °C. Furthermore, the rate of thermal inactivation due to autolysis is 1.5-2.0 times slower than wild-type under a variety of conditions. The introduced disulphide bond clearly improves the thermostability of the C24/C87 mutant and consequently enhances its resistance to autolysis. The apparent contribution of the introduced bond to the free energy of unfolding relative to the reduced form, has been estimated at about 1.3 kcal mol^{-1} (Pantoliano *et al.*, 1987).

Wells and Powers, (1986) have constructed the TC22/SC87 mutant, but this variant seems less stable than wild-type. The reasons for this are unclear, but in this case the mutant contains an additional substitution (Tyr-21 to Ala) which was introduced as part of an unrelated study. It is possible that the additional mutation may destabilise the mutant to a degree that is beyond recovery by the formation of a single disulphide bond. Future attempts to enhance thermostability by introducing disulphide bonds should

consider avoiding the disruption of interactions which already stabilize the molecule in addition to ensuring that the bond has good geometry. The possibility that in some cases an introduced disulphide bond might prevent protein mobility essential to the mechanism of catalysis and therefore decrease catalytic activity must be also be considered.

Random mutagenesis of the subtilisin BPN' gene followed by selection for thermostable variants has produced single site mutants with enhanced thermostability and resistance to autolysis (Bryan *et al.*, 1986). One of the mutants isolated was identified as Ser→Asn at position 218. Differential scanning calorimetry showed this mutant to have a T_m 2.4-3.9 °C higher than wild-type. Furthermore, the rate of thermal inactivation due to autolysis was reduced to 25% that of wild-type. Comparison of the crystal structures of wild-type subtilisin and the SN218 mutant shows an improvement in hydrogen bond parameters in the vicinity of the mutation. Thus, increased hydrogen bonding can also yield enzymes with enhanced thermostability.

Volkin and Klibanov (1987) have questioned the wisdom of attempting to achieve increased thermostability by engineering additional disulphide bonds into proteins. The mechanism of irreversible thermal inactivation for many proteins is heat induced disulphide exchange or the destruction of disulphide bonds by β-elimination. They therefore consider that it is unwise to engineer new disulphide bonds into proteins required to function at high temperatures for prolonged periods. A comparison of proteins found in nature gives no indication that enzymes from thermophiles utilize additional disulphide bonds to stabilize their structures (Hochachka and Somero, 1984), rather additional hydrogen bonding and salt bridges (Perutz, 1978). Perhaps a more attractive approach to enhancing thermostability of enzymes by protein engineering is to increase the number of hydrogen bonds and/or salt

bridges.

1.3.5 Transition-state stabilisation: Role of the oxyanion hole

Subtilisin achieves catalysis by stabilising the oxyanion which develops on the carbonyl oxygen on formation of the tetrahedral transition state (Robertus *et al.*, 1972). Structural studies have suggested that hydrogen bonds form between the negatively charged carbonyl oxygen (oxyanion) and the amide side-chain of Asn-155 and also the main-chain amide of the catalytic Ser-221 (Robertus *et al.*, 1972). Other H-bonds between the enzyme and substrate remain unchanged in both the enzyme-substrate and the transition-state complexes.

Mutant subtilisins have been constructed with substitutions which remove the hydrogen bonding interaction at position 155. Asn-155 has been replaced by Thr, Gln, Asp and His (Wells *et al.*, 1986) and also by Leu (Bryan *et al.*, 1986). In all cases k_{cat} for the hydrolysis of the peptide substrate suAAPFpN (see Section 5.5.1) was reduced dramatically (200-4000 fold), whilst K_m remained unaffected or only slightly reduced (<7 fold). The large decrease in k_{cat} confirms the importance of H-bond formation in transition-state stabilisation in this enzyme and that this interaction confers a significant energetic incentive to the formation of the tetrahedral intermediate. Wells *et al.*, (1986) have estimated that the removal of this interaction results in a loss of transition-state stabilisation energy of 9.2-20 kJ mol^{-1} , which is in close agreement with a value of 15 kJ mol^{-1} estimated by Bryan *et al.*, (1986).

1.4 Electrostatic effects in proteins

All forces between atoms and molecules are electrostatic in nature.

However, the term electrostatics is frequently used to describe just those interactions which occur between atoms or molecules bearing permanent or induced dipoles.

The electrostatic interaction energy (E) between two point charges (e_1, e_2), separated by a distance (r), in a medium of dielectric constant D is given by:

$$E = (e_1 \ e_2)/D \ r$$

Dielectric constant is defined as the measure of a materials ability to attenuate the electrostatic interaction of two point charges separated by that material. In simple cases, where D is a continuum, such as ions in homogenous solution, the electrostatic forces are relatively simple to calculate. However, in proteins the dielectric constant is non-uniform and the magnitude of electrostatic effects is therefore very difficult to calculate. Historically the value of D for the interior of proteins has been taken to be low (4-10) compared with a value of 78 for pure water. The dielectric constant varies depending on the micro-environment in the protein; this can be very polar in the vicinity of ionic side chains and peptide bond dipoles or non-polar in regions consisting mainly of hydrocarbon side-chains. Rogers (1986) likened proteins to Swiss cheese in that the generally hydrophobic interior of globular proteins is permeated by hydrophilic channels. It is clearly an oversimplification to assume that the dielectric constant in proteins is a continuum, it will vary from a low value in non-polar regions to a value close to that of the solvent at the protein/solvent interface. It is likely that in highly polar regions the dielectric might exceed that of bulk water.

Electrostatic effects dominate virtually every aspect of protein structure and

function (Perutz, 1978). The structure and overall stability of a protein molecule can be considered as the sum of all electrostatic interactions: hydrophobicity; hydrogen bonding; ion-pairs (salt bridges); van der Waals interactions.

Electrostatic effects also have a profound role in enzyme catalysis (see Fersht, 1985). Many enzyme catalysed reactions proceed via charged transition states, stabilisation of these charged species by interactions with charged groups on the enzyme is a key factor in catalysis. The pH dependence of enzyme catalysis depends on the ionisation of catalytic groups which are themselves sensitive to the local electrostatic environment.

A variety of theoretical approaches to calculating electrostatic effects in proteins have been applied in recent years (reviewed by Matthew, 1985). Whilst an estimate of the distance separating charged groups can be obtained from high resolution crystallographic data, the direct evaluation of effective dielectric constant is considerably more difficult. Many such calculations founder on inconsistencies such as whether the effective dielectric constant in the region of interest is low or high and also on the effect of the boundary between the high dielectric of the solvent and the lower, probably heterogenous values for the protein.

1.5 Effect of surface charge on pH dependence

Chemical modification of surface charged residues has a marked effect on the pH dependence of chymotrypsin (Valuenzuela and Bender, 1971). The surface amino groups of chymotrypsinogen A were reacted with succinic anhydride to give a derivative from which succinylated chymotrypsin was isolated. The apparent pK_a of the ionising group on which enzyme activity depends (His 57) was shifted from a value of 7.0 in the unmodified

enzyme to 8.0 in the succinylated form which has an increase in surface negative charge. A further chemically modified derivative with an increase in surface positive charge was produced on amidation of 12 of the 14 surface carboxyl groups using ethylene diamine in the presence of aqueous carbodiimide. In this case, the apparent pK_a of His 57 is reduced from 7.0 to about 6.1 on modification. These studies clearly show that the active-site histidine is sensitive to changes in the electrostatic environment. Furthermore, they suggest that altering the electrostatic environment of ionic catalytic groups may be a generally applicable approach to modifying the pH dependence of enzyme catalysis by protein engineering (Thomas *et al.*, 1985). The extensive charge changes that these chemical modifications generate may produce structural perturbations in the protein. For example, expansion of the molecule as a result of the high charge density.

Little is known of the effect of modifying single charged groups in a protein. Selective modification of individual surface lysine residues in cytochrome C results in changes in the redox potential relative to the unmodified protein (Rees, 1980). The use of these redox shifts to calculate the effective dielectric constant in that region has been criticised by Rogers *et al.*, (1985), since the modifying agent used to neutralise the amino groups was *m* - trifluoromethylphenyl-carbamoyl, which is bulky and may cause structural perturbations.

Protein engineering could provide a much more delicate probe of electrostatic effects in proteins and would avoid the necessary caveats placed on conclusions drawn from chemical modification studies.

In subtilisin, enzyme activity depends on the ionisation of His-64. The enzyme is only active at alkaline pH where His-64 is unprotonated and catalytic activity varies with pH following the ionisation of this residue.

Single amino acid substitutions which either remove or introduce a charged residue may stabilise or destabilise the protonated (i.e. charged) form of His-64 and result in a shift in the pK_a for the ionisation of His-64. So, removal of a negative charge in the vicinity of the active site is expected to destabilise the positively charged form of His-64 and so lower its pK_a. Conversely removal of a positive charge may be expected to stabilise the protonated form of His-64 and consequently raise its pK_a. The magnitude of any pK_a shifts so generated can not yet be predicted, but will probably depend on the distance and environment separating the mutated residue and the active-site histidine.

1.6 Aim of this study

This study is aimed at developing an experimental system which will allow the magnitude of electrostatic effects in proteins to be measured accurately. Subtilisin BPN' has been selected as a model system for the reasons discussed earlier. Mutant subtilisins will be constructed such that ionic amino acids are removed from, or introduced at, the enzyme surface. The effect of these mutations on the pH dependence of enzyme catalysis will be determined using enzyme kinetics.

The data will allow theoreticians to test the applicability and limitations of current models of electrostatic interactions in proteins. A detailed understanding of electrostatic interactions in proteins is a prerequisite for the rational design of novel proteins.

The electrostatic effects of surface charges on the ionisation constants of catalytically active groups in enzymes are interesting from a practical point of view, since modification of charge may provide a general means of tailoring the pH dependence of enzyme catalysis (Thomas et al., 1985).

CHAPTER 2: MATERIALS AND METHODS

2.1 Culture media

The following media were used:

Luria Bertani broth (LB-broth) contains: 10 g Bacto-Tryptone; 5 g yeast extract; 10 g NaCl, per litre. The pH was adjusted to 7.5 with NaOH. LB-agar contains 15 g per 1 of Bacto-Agar in addition to the components above.

2xTY broth contains: 16 g Bacto-Tryptone; 10 g yeast extract; 5 g NaCl, per litre.

Minimal agar (MGB1) was prepared by adding 100 ml of 5xMM salts (see below) to 400 ml of water containing 7.5 g of Bacto-Agar which had been freshly melted by autoclaving. After cooling to 55 °C, the following supplements were added: 0.5 ml of 20% (w/v) $MgSO_4$; 5 ml of 20% (w/v) glucose and 0.25 ml of (1% w/v) vitamin B1 (Thiamine). Stock solutions of the supplements were sterilised by filtration through 0.22 μm pore size filters. Plates were poured in 90 mm petri-dishes.

5xMM salts contains: 52.5 g K_2HPO_4; 22.5 g KH_2PO_4; 5 g $(NH_4)_2SO_4$ and 2.5 g sodium citrate, per litre. The solution was divided into 100 ml aliquots and sterilised by autoclaving.

H-Top agar (or agarose) contains: 10 g Bacto-Tryptone; 8 g NaCl; 6 g Bacto-Agar (or agarose) per litre.

Bacto-Tryptone, yeast extract and Bacto-Agar were purchased from Difco.

Table 2.1: Bacterial Strains.

Strain	Genotype	Source/Ref.
E. coli LM1035[1]	F$^-$,*hsd*S20 (r$_B^-$,m$_B^-$) *rec*A13,*ara*-14,*pro*A2, *lac*Y1,*gal*K2,*rps*L20(Sm^r), *xyl*-5,*mtl*-1,*sup*E44,λ^-	A.J.P.Docherty
E. coli DH1	F$^-$,*rec*A1,*end*A1,*gyr*A96, *thi*-1,*hsd*R17 (r$_K^-$,m$_K^-$) *sup*E44,*rel*A1,λ^-	D. Hanahan (1983)
E. coli JM101	*thi*,*sup*E,Δ(*lac-pro*AB), [F',*tra*D36,*pro*AB,*lac*I^q, *lac*ZΔM15]	J. Messing *et al* (1981)
E. coli BMH71-18 *mutL*	K12,Δ(*lac-pro*AB),*thi*, *mut*L::Tn^{10},[F',*pro*AB,*sup*E, *lac*I^q,*lac*ZΔM15]	Kramer *et al* (1984)a
. *E. coli* TG2	K12,Δ(*lac-pro*AB),*thi*,*sup*E, *hsd*Δ5,*rec*A,*srl*::Tn10^{TcR}, [F',*tra*D36,*pro*AB,*lac*I^q, *lac*ZΔM15]	M.Biggin (1984)
B. amyloliquefaciens RUB500[2]	Prototroph,source of restriction endonuclease *Bam*H1	E.Kenny
B. subtilis IG20[3]	*trp*C2 (r$^-$,m$^-$)	A.J.P.Docherty
B. subtilis DB104	*his*,*npr*R2,*npr*E18,Δ*apr*A3.	Kawamura & Doi (1984)

Notes to Table 2.1:
    (1)   A highly transformable derivative of *E. coli* HB101.
    (2)   *Bacillus* Genetic Stock Centre (BSGC) code 10A2.
    (3)   BSGC code 1A436.
The *Bacillus* Genetic Stock Centre, Department of Microbiology, Ohio State University, 484 W. 12th Ave. Columbus, Ohio 43210.

Media and components were sterilised by autoclaving at 121 °C for 20 min unless stated otherwise. Where required, antibiotics and supplements were added to the final concentrations shown.

2.2 Bacterial strains

The bacterial stains used, their genotypes and source or reference are summarised in Table 2.1. For details on the genetic markers in *E. coli* strains see Bachmann (1984) and for *B. subtilis* genetic markers see Piggott and Hoch (1985).

E. coli strains LM1035 and DH1, and all *Bacillus* strains were propagated on LB-agar plates. *E. coli* strains JM101, TG2 and BMH71-18 *mut*L were propagated on MGB1-agar plates to select for maintainenance of the F factor. Strains were stored in 15% glycerol at -70 °C as decribed by Maniatis *et al.*, (1982).

2.3 Buffers and Solutions

Buffer compositions are shown in Table 2.2

2.4 Isolation of DNA

2.4.1 Cloning vectors

The plasmid and bacteriophage cloning vectors used in this study are summarised in Table 2.3

Table 2.2: Composition of buffers and other solutions.

Buffer/Solution	Composition
1xTAE	40 mM Tris-acetate pH 8.0; 1 mM EDTA
1xTBE	90 mM Tris, 90 mM Borate, 2 mM EDTA pH 7.8
TES	30 mM Tris-HCl pH 8.0, 5 mM EDTA, 50 mM NaCl
TE	10 mM Tris-HCl pH 7.6, 1 mM EDTA
TNES	10 mM Tris-HCl pH 7.6, 0.14 M NaCl, 1 mM EDTA, 0.1% SDS
1xSSC (Standard saline citrate)	0.15 M NaCl, 0.015 M sodium citrate
PM	20% (w/v) sucrose, 20 mM Tris-HCl pH7.5, 20 mM $MgCl_2$
1xKB (Kinase buffer)	50 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, 5 mM DTT, 0.1 M spermidine-HCl, 0.1 mM EDTA
HB (6xNETS/Denhardts/SDS Hybridisation Buffer)	0.9 M NaCl, 0.09 M Tris-HCl pH 7.4, 6 mM EDTA, 0.5% (v/v)NP40, 0.2% (w/v) SDS, 2xDenhardts (Filtered through 0.45 μm pore size nitrocellulose prior to use)
Denhardts solution (100x)	2% (w/v) Bovine serum albumin, 2% (w/v) Ficoll 70 2% (w/v) Polyvinylpyrollidone
1xCIAP buffer	10 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$
Lysis Buffer	50 mM Glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA
Restriction Buffers (Maniatis *et al.*, 1982)	
High salt	100 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT
Medium salt	50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT
Low salt	10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT

Table 2.3:   Summary of plasmid and phage vectors used.

Vector	Size (kb.)	Selective markers	Host	Reference
1. Plasmid vectors:				
pBR322	4.36 kb	amp^R,tet^R	*E. coli*	Bolivar *et al.*, (1977).
pAT153	3.65 kb	amp^R,tet^R	*E. coli*	Twigg and Sherratt (1980).
pUB1662	8.5 kb	amp^R,kan^R kan^R	*E. coli/* *B. subtilis*	A. J. P. Docherty (1980).
pUB110	4.5 kb	kan^R	*B. subtilis*	Gryczan *et al.*, (1978).
2. Phage vectors:				
M13mp8/9	6.4 kb		*E. coli* [F']	Messing and Vieira (1982).

2.4.2 Large scale plasmid isolation from *E. coli*

Plasmid DNA was isolated from 250 ml cultures of *E. coli* using the alkaline/SDS procedure of Birnboim and Doly (1979) as modified by Ish-Horowicz and Burke (1981). Caesium chloride density gradients were centrifuged in a Beckman Ti 70.1 rotor at 48,000 rpm for 20 hr at 20 ˚C. Plasmid DNA was recovered through the side of the centrifuge tube using a syringe and hypodermic needle. Ethidium bromide was removed by extraction with CsCl-saturated butan-1-ol. Following ethanol precipitation, residual CsCl was removed by extensive dialysis against TE. DNA was stored frozen at -20 ˚C.

2.4.3 Small scale plasmid isolation from *E. coli*

Isolation of plasmid DNA from small scale cultures (2 ml) was as described (Maniatis *et al.*, 1982) except for the inclusion of a single phenol/chloroform extraction step (Section 2.6) prior to ethanol precipitation (Section 2.7).

2.4.4 Modified plasmid isolation method used to recover pPT1 from *E. coli*

E. coli harbouring the recombinant plasmid pPT1 grew very slowly at 37 ˚C to give colonies only 0.2-0.5 mm in diameter after incubation overnight. A single "micro-colony" was picked from a freshly grown LB-agar plate containing ampicillin (100 μg ml^{-1}) and dispersed in 0.5 ml of LB-broth. Aliquots (0.25 ml) were spread on LB-agarose plates containing ampicillin (100 μg ml^{-1}) and the plates incubated at 37 ˚C for 36 hr. The colonies were suspended in lysis buffer (Table 2.2, 1 ml per plate) using a glass spreader then transfered to microcentrifuge tubes. Each plate was rinsed with a further 0.5 ml of buffer and the cells pooled. The cells were washed by centrifuging briefly, discarding the supernatant and resuspending

in fresh lysis buffer. After a second wash, the cells were resuspended in 100 μl of lysis buffer. From this point, the procedure was exactly as described for small scale plasmid isolation .

2.4.5   Isolation of M13 RF DNA

M13mp9 replicative form (RF) DNA was made by infecting an early log-phase culture (200 ml) of *E. coli* TG2, grown in 2xTY-broth, with 1 ml of phage stock. The culture was grown at 37 °C for 18 hr in a shaking incubator. RF DNA was isolated as described for plasmid DNA (2.4.2). The phage stock was prepared by innoculating a single M13 plaque into 10 ml of 2xTY containing 0.1 ml of an overnight culture of E. coli TG2. After growth for 8 hr at 37 °C the infected cells were removed by centrifugation (5,000 rpm, 20 min, 4 °C). The phage supernatant was decanted, sterilised by filtration through a 0.22 μm pore size filter and stored at 4 °C.

2.4.6 Large scale plasmid isolation from *B. subtilis*

Plasmid DNA was purified from *B. subtilis* by a modification of the alkaline/SDS method used for *E. coli* (Hardy, 1985). Cultures (500 ml) were grown at 37 °C for 18 hr in LB-broth containing kanamycin (25 μg ml^{-1}) to select for plasmid-bearing cells.

2.4.7 Purification of chromosomal DNA from *B. amyloliquefaciens*

B. amyloliquefaciens RUB500 was streaked on an LB-agar plate and grown at 37 °C overnight to provide single colonies. One colony was dispersed in 2 ml of LB-broth and used to inoculate 200 ml of prewarmed LB-broth containing 0.2% glucose in a 2 l flask. The culture was grown for 18 hr

at 37 °C. The cells were harvested by centrifugation (4,000 rpm, 4 °C, 20 min, Sorval GSA rotor). The cell pellet was resuspended in 40 ml of TES, transferred to a 50 ml "Oak-Ridge" tube then recentrifuged (4,000 rpm, 4 °C, 20 min, Sorval SS34 rotor) and the supernatant discarded to remove secreted nucleases. The washed cells were resuspended in 5 ml of PM containing lysozyme (2mg ml^{-1}) and incubated at 45 °C for 15 min. The resulting protoplasts were lysed by osmotic shock on addition of 25 ml of TES to give a clear solution which was highly viscous due to the released chromosomal DNA.

1 ml of 10% SDS and 0.5 ml of Proteinase K (10 mg ml^{-1} in TES) was added and then the lysate incubated at 45 °C for 1 hr. The lysate was extracted with an equal volume of a 2:1 (v:v) mixture of buffered phenol (prepared as described by Maniatis et al., 1982) and chloroform to denature and extract proteins. After gentle mixing, avoiding shearing the DNA, and incubation on ice for 20 min the aqueous and organic phases were separated by centrifugation (8,000 rpm. 20 °C, 30 min, Sorval SS34 rotor). The upper aqueous phase containing total nucleic acids was aspirated off avoiding debris at the interphase. The aqueous phase was re-extracted with phenol/chloroform twice. Residual phenol was removed by extraction with chloroform alone.

The aqueous phase was removed into a sterile 200 ml beaker. Sodium acetate was added to a final concentration of 0.3 M and the DNA precipitated by adding 2.5 volumes of chilled (-20 °C) 96% ethanol. The phases were mixed by swirling gently and the DNA allowed to precipitate at -20 °C for 30min.

The precipitated DNA formed a clot which floated at the surface and was hooked out using a bent, sterile glass pasteur pippette. The DNA was

52

drained for a few seconds and then dissolved in 5 ml of sterile water. Next sodium acetate was added to 0.3 M followed by 2.5 volumes (12.5 ml) of chilled ethanol. The DNA was then left to precipitate for 30 min at -20 °C. The chromosomal DNA was reprecipitated in this way a total of four times. The repeated precipitation of the high molecular weight DNA efficiently separates it from RNA and small DNA fragments, which precipitate more slowly and do not form a clot. These contaminants are therefore diluted out by the repeated precipitations. The use of ribonuclease is avoided since it is difficult to obtain completely free of DNAase.

The final DNA precipitate was dissolved in sterile water and adjusted to 1xTE. The DNA solution (approximately 15 ml) was dialysed against TE (2x1 litre) at 4 °C for 18 hr. The DNA was quantified from its absorption at 260 nm. The purified chromosomal DNA was stored frozen at -70 °C.

2.4.8 Quantification of DNAs

DNA concentrations were determined spectrophotometrically using a Perkin-Elmer λ5 UV/VIS spectrophotometer. Wavelength scans (200-300 nm) were performed on DNA samples in TE, using TE as the blank. DNA concentrations were calculated using the assumption that an absorbance at 260 nm of 1 corresponds to 50 μg ml^{-1} of double stranded DNA. See Section 2.12.3 for the quantification of oligonucleotides.

2.5 Restriction endonuclease cleavage of DNA

Restriction enzymes were purchased from commercial sources and used in accordance with the suppliers directions. Digests were performed in the

Low, Medium or High ionic strength buffers described by Maniatis *et al.*, (1982). Reactions typically contained 1-10 μg of DNA at a concentration of approximately 100 μg ml^{-1} in a reaction volume of 10-100 μl. Digests were allowed to proceed for 1-2 hours at the appropriate temperature. The reactions were terminated by heat inactivation at 70 °C for 10 min or by phenol/chloroform extraction.

2.6 Phenol/chloroform extraction of DNA

Proteins were removed from solutions of DNA by phenol/chloroform extraction essentially as described by Maniatis *et al.*, (1982), except that a 2:1 volume ratio of buffered phenol to chloroform was used for each extraction. Residual phenol was removed by extraction with either water saturated di-ethylether or chloroform. DNA was recovered by ethanol precipitation (see below). The preparation of buffer saturated phenol was as decribed by Maniatis *et al.*, (1982).

2.7 Ethanol precipitation of DNA

The DNA solution was adjusted to 0.3 M sodium chloride (or sodium acetate) and 2.5 volumes of chilled (-20 °C) 96% ethanol were added. After mixing, the DNA was allowed to precipitate at -20 °C for at least 2 hours or at -70 °C for 30 min. The precipitated DNA was recovered by centrifugation for 10 min at 4 °C in a micro-centrifuge, or at 8,000 rpm at 4 °C for 30 min in a Sorval RC5 centrifuge (SS34 rotor). Precipitated chromosomal DNA was not centrifuged since this can result in the shearing of high molecular weight DNA (see Section 2.4.7). After centrifugation the supernatant was discarded and the pellet of DNA rinsed with a small volume of chilled 70% ethanol, to remove traces of salt. The DNA was briefly recentrifuged, the supernatant discarded and the DNA dried under

vacuum for 5 min. The DNA was dissolved in an appropriate volume of sterile water or TE and stored at -20 °C. Alterations to this procedure are described where applicable.

2.8 Agarose gel electrophoresis

2.8.1 Electrophoresis conditions

Chromosomal and plasmid DNA, and restriction fragments thereof were separated by electrophoresis through horizontal, submerged agarose gels. Gel tanks were supplied by BRL (model H4) and IBI (model QSH). Agarose (Agarose-NA) was purchased from Pharmacia. Gels, (0.3, 0.8 or 1.0%) were cast and run in either half strength Tris-borate buffer (0.5xTBE) at a constant 75 V or in Tris-acetate buffer (TAE) at a constant 25V. Analytical gels were run in 0.5xTBE, and preparative gels in TAE. Electrophoresis was performed at room temperature and the gels run until the tracking dye (bromophenol blue) had migrated 70% of the gel length or as required. Gels were stained in ethidium bromide (1 μg ml^{-1}) for 20 min then destained in water for 15-45 min. The DNA was visualised by fluorescence using a 305 nm UV trans-illuminator (UV Products) and photographed through a Kodak 23A Wratten filter on Polaroid Type 665 film. DNA samples were applied in 10% glycerol, 1 mM EDTA and 0.02% bromophenol blue. A *Hind*III digest of λ DNA (bactriophage lambda cI-857 Sam7, obtained from Boehringer) provided restriction fragment size markers of: 23,130; 9,416; 6,557; 4,361; 2,322; 2.027; 564 and 125 base pairs. Size estimation of unknown restriction fragments was performed by comparison to these standards.

2.8.2 Recovery of DNA fragments from gels

DNA fragments were recovered from agarose gels by electrophoretic adsorption onto strips of DEAE membrane (NA45 membrane, Schleicher & Schuell, Product code-417082). The membrane was pre-treated by soaking in TE containing 2.5 M NaCl for 1 hr at room temperature then rinsed in sterile water. Membrane strips were stored in 1 mM EDTA at 4 °C. Immediately prior to use, the strips were rinsed in water and then soaked for 5 min in 1xTAE electrophoresis buffer. DNA fragments were separated on gels cast and run in 1xTAE, visualised by UV fluorescence (305 nm) and then electrophoresed onto a strip of NA45 placed in a slit cut just below the required band. After removal from the gel the strip was cut into 2 mm^2 pieces and the DNA eluted into 200 μl of 1 M NaCl on incubation at 70 °C for 20 min. The membrane pieces were rinsed with a further 200 μl of salt solution and again incubated at 70 °C for 20 min. The eluates were pooled and extracted with a 2:1 (v:v) mixture of phenol and chloroform as described (Section 2.6). The DNA was recovered from the aqueous phase by ethanol precipitation (Section 2.7). DNA purified in this fashion was found to be suitable for subsequent re-cloning.

2.9 Dephosphorylation of vector DNAs

Removal of 5' phosphates from vector DNAs, to prevent religation, was achieved using calf intestinal phosphatase. 1-10 μg of the vector DNA was completely digested with the appropriate restriction enzyme in a 10-100 μl reaction volume. The restriction enzyme was inactivated by incubation at 70 °C for 10 min, 20 μl of 10xCIAP buffer was then added and the volume made up to 200 μl with sterile water. One unit of calf intestinal alkaline phosphatase (CIAP, Boehringer, Molecular Biology grade) was added and the reaction incubated at 37 °C for 30 min. TNES (200 μl) was

56

added and the phosphatase fully inactivated by incubation at 70 °C for at least 30 min. The dephosphorylated DNA was phenol/chloroform extracted twice and recovered by ethanol precipitation. The long heat inactivation was found to be necessary to prevent active phosphatase interfering with subsequent ligation reactions.

2.10 Ligations

Ligations were performed in 50 mM Tris-HCl pH 7.6, 10 mM MgCl$_2$, 1 mM ATP and 20 mM DDT at 15 °C for 4-18 hours. T4 DNA ligase was a gift from J. Knill-Jones. One unit of ligase was used per 20 μl of reaction volume. The concentration of DNA terminii was varied to promote the formation of monomeric or multimeric ligation products as required for transformation into *E. coli* or *B. subtilis*.

2.10.1 Ligation into *E. coli* vectors

Reactions typically contained 50-100 ng of dephosphorylated vector DNA and sufficient target DNA to give a 1:1 ratio of vector terminii to target terminii, thereby favouring the insertion of single target fragments into the vector. The total DNA concentration was 5-10 μg ml^{-1} in a 20 μl reaction volume.

2.10.2 Ligation of DNA into *B. subtilis* vectors

To favour the formation of multimeric ligation products which are required for the transformation of competent *B. subtilis* cells, ligations were performed at high DNA concentrations (50 μg ml^{-1}). Reactions typically contained 500 ng of vector DNA and sufficient target DNA to give an equimolar ratio of terminii. Reaction volumes were typically 20 μl.

2.11 Preparation and storage of transformation competent cells

2.11.1 Competent *E. coli*

E.coli strains were rendered transformation competent by the method of Viesturs Simanis (personal communication and described by D. Hanahan, 1985). Transformation efficiencies of better than 10^6 cfu (colony forming units) per μg of supercoiled pBR322 were routinely obtained.

2.11.2 Competent *B. subtilis*

Transformation competent *B. subtilis* cells were prepared using the method of Anagnostopoulos and Spizizen (1958) as modified by Dubnau and Davidoff-Abelson (1971). The procedure has been described by Hardy (1985). Transformed cells were selected on LB-agar plates containing kanamycin (25 μg ml^{-1}). Media was supplemented with histidine (50 mg ml^{-1}) for the growth of *B. subtilis* DB104 or tryptophan (50 μg ml^{-1}) for *B. subtilis* IG20.

2.12 Oligonucleotide synthesis and purification

2.12.1 Synthesis

Oligodeoxyribonucleotides (oligonucleotides) were produced on an automated DNA synthesiser (Applied Biosystems Inc. Model 380B), using optimised, solid phase, phosphite-triester chemistry (Caruthers *et al.*, 1982). Probes R and Y were synthesised manually by J. Brannigan.

2.12.2 Deprotection and purification

Oligonucleotides were fully deprotected by incubation in 35% (v/v) ammonia solution at 56 °C for 18hr, dialysed against water then quantified. The full length oligonucleotide was purified away from shorter products and protecting groups by electrophoresis through a 20% polyacrylamide gel (200 x 400 x 0.4mm) containing 1xTBE and 8 M urea. 2-5 OD. units of the crude oligonucleotide were lyophilised then dissolved in 20 μl of 98% deionised formamide containing 2 mM EDTA. The gel was pre-run for 30 min to remove the salt front which interferes with the migration of the oligonucleotide. The crude oligonucleotide was loaded at a maximum of 0.5 OD units per cm of gel, and electrophoresed at a constant 37W (approximately 1400 V, 25 mA). Marker dyes (0.1% BPB and 0.1% XC) were run in adjacent tracks, the gel was run for 1.5-2 hr until the BPB had migrated about 30 cm.

The required product was visualised by UV shadowing as follows. The gel plates were separated and the gel transferred to a sheet of "Saranwrap". The gel was placed over a fluorescent TLC plate then illuminated from above with 245 nm UV light. The required product was the slowest migrating material, which also cast the darkest shadow. The gel fragment containing the full length product was excised, cut into 2 mm squares and the oligonucleotide eluted into 1 ml of sterile water at 37 °C for 4-12hr. The eluate was centrifuged (12000 rpm, 10 min, room temp) to remove any gel fragments then extensively dialysed against water. The concentration of the purified oligonucleotide was determined from its absorption at 260 nm as described below. For short term storage, oligonucleotides were kept frozen at -20 °C. For long term storage oligonucleotides were lyophilised and kept at -20 °C.

2.12.3 Quantification of oligonucleotides.

The concentration of oligonucleotides was determined accurately using the relationship:

$$OD/\epsilon_{260} = x \ mM$$

where ϵ_{260} is the millimolar extinction coefficient. The value of ϵ_{260} was determined for each oligonucleotide by summing the contributions of each base, which are shown in Table 2.4. For convenience, concentrations were converted to pmol μl^{-1}.

2.13 5' ^{32}P labelling oligonucleotides

T4 polynucleotide kinase mediated end labelling of oligonucleotides was done using 100 pmol of oligonucleotide and 100 μCi of [γ ^{32}P] ATP (New England Nuclear, 7000 Ci mmol^{-1}). The reaction (50μl) was done in 1 x kinase buffer (Table 2.2) and contained 2 units of T4 polynucleotide kinase (Amersham International). After incubation at 37 °C for 30 min the reaction was terminated by adding 50 μl of 0.6 M ammonium acetate. The bulk of the unincorporated [γ ^{32}P] ATP was removed by chromatography on a Sephadex G25 spinning column as described by Maniatis *et al.*, (1982). The end labelled oligonucleotide was recovered in the void volume of about 100 μl, 900μl of hybridisation buffer (HB) was added giving a probe concentration of approximately 100 pmol ml^{-1}. Labelled probe was stored at -20 °C and used within 2 weeks for hybridisation screening.

Table 2.4: Contribution of bases to ϵ_{260}

BASE	CONTRIBUTION TO ϵ_{260} mM^{-1} cm^{-1}.
T	8.8
C	7.3
G	11.7
A	15.4

Data from Sproat and Gait (1984).

2.14 Hybridisation screening using oligonucleotide probes

2.14.1 Colony screening

Colonies were grown on selective media and transferred onto 82 mm nitrocellulose filters (Sleicher & Schuell, BA85 0.45 μm pore size), by placing a filter directly onto the plate. The filters were keyed to the master plates using a hypodermic needle to punch holes asymetrically at the margin of the filter. This allows the subsequent orientation of master plate, filter and autoradiogram. The filters were removed carefully from the plate and placed, colony side up, onto a fresh LB-agar plates containing ampicillin (100 μg ml^{-1}) and chloramphenicol (200 μg ml^{-1}). The filters were incubated at 37 °C for 16 hr to amplify plasmids harboured by the colonies. The master plates were also incubated a further 16 hr to allow the colonies to regrow, the plates were then sealed with "Parafilm" and stored at 4 °C. After amplification the colonies were lysed *in situ* and the released DNA immobilised on the nitrocellulose filter essentially by the procedure of Grunstein and Hogness (1975).

The filters were transferred, colony side up, to a plastic tray containing two layers of Whatman 3MM paper soaked in 0.5 M NaOH. The filters were exposed to alkali for 10 min, then transferred to fresh NaOH soaked 3MM paper for a further 10 min. The filters were neutralised by placing on 3MM paper soaked with 1 M Tris-HCl pH 8.0 for 2x10 min. The DNA was then bound to the nitrocellulose filters by placing them on 3MM paper soaked in 1 M Tris-HCl pH 8.0, 1.5 M NaCl for 15 min. The filters were air dried for 1 hr then sandwiched between sheets of 3MM paper and baked at 80 °C for 2 hr.

2.14.2 Hybridisations

Hybridisations were done in polythene bags made from double welded 500 gauge polythene sheet. Filters were placed "back to back" with up to 8 x 82 mm filters per bag. 6xNET/Denhardts/SDS hybridisation buffer (HB) containing denatured salmon sperm DNA (100 μg ml^{-1}) and *E. coli* tRNA (70 μg ml^{-1}) was added to the bag (2 ml per filter). Any remaining air was excluded and the bag heat sealed then placed in a 56 °C water bath to prehybridise for 2 hr. Prehybridisation prevents the radiolabelled probe from binding non-specifically to the nitrocellulose. 5' ^{32}P labelled oligonucleotide probe was added to a final concentration of 1 pmol ml^{-1} using a hypodermic syringe to inject the radiolabelled probe through the side of the bag. The puncture was sealed with waterproof adhesive tape (Scotch "Magic" Tape) and the probe distributed evenly throughout the bag. Hybridisation was allowed to proceed for 18 hr at a temperature at least 10 °C below the calculated Td for the probe used.

2.14.3 Washing filters

Filters were washed in 6xSSC to remove non-specifically bound probe. The stringency of the washes was controlled by altering temperature. The filters were removed from the bag and rinsed briefly in 6xSSC at 4 °C to remove excess labelled probe, followed by 3x45 min washes at room temperature (18 °C) to remove non-specifically bound probe. Washes at higher stringency were in 6xSSC maintained at the required temperature (eg. Td-2 °C). Filters were washed for a total of 10 min in two changes of buffer. The filters were blotted to remove excess buffer, but not allowed to dry, then sandwiched between layers of "Saranwrap" and autoradiographed for 6-24 hr or longer as required. Autoradiography was performed at -70 °C using pre-flashed X-ray film (Fuji RX) and an intensifying screen

(Cronex Lightning Plus, DuPont). Radioactive ink marks were used to key the autoradiogram to the filters. Hybridisation positives were identified on the master plates by reference to the keyed holes in the filter and the marks on the autoradiogram made by the radioactive ink.

2.14.4 Dot blots

1 μl of template DNA or 10 μl of resuspended phage were spotted onto gridded nitrocellulose. The filter was baked at 80 °C for 2hr. Hybridisation and washing was as described above except that prehybridisation was omitted and the hybridisation time reduced to 1hr. The exposure time for dot-blots was typically 4-6 hr.

2.14.5 Direct plaque hybridisation

Transfected cells were plated in H-top-agarose. This provides greater mechanical strength than H-top-agar and damage to the master plate during plaque transfer is avoided. The use of agarose also gives much lower background hybridisation. After incubation at 37 °C overnight, the plates were chilled at 4 °C for 2 hr. The plaques were then transferred to 82 mm nitrocellulose or nylon filters (Hybond-N, Amersham International). The filters were left in contact with the plaques for up to 15 min in order to maximise the number of phage transferred and also therefore the signal strength. The filters were keyed to the master plates then removed and baked for 1 hr at 80 °C, to disrupt the phage particles and immobilise the DNA. When nylon filters were used the DNA was covalently crosslinked to the membrane by exposure to 305 nm UV light for 10 min. Hybridisation and washing was as described earlier (Section 2.14.2).

2.15 M13 template preparation and DNA sequencing

Single stranded M13 template DNA for sequencing and mutagenesis was grown and purified as described by Bankier and Barrell (1983). Di-deoxy chain terminator sequencing (Sanger *et al.*, 1977) was also as described by Bankier and Barrell (1983). The products of sequencing reactions were resolved on thin (0.3 mm) 6% acrylamide gels containing 7 M urea and 1xTBE (Sanger and Coulson, 1978). DNA sequences were analysed using the *MICROGENIE sequence software* (Queen and Korn, 1983; distributed by Beckman).

CHAPTER 3: CLONING AND EXPRESSION OF THE SUBTILISIN GENE

3.1 Introduction: *Bacillus* as a host for recombinant DNA

3.1.1 Molecular cloning

Gene cloning technology developed from the study of plasmid biology and the phenomena of plasmid mediated transformation, the mechanism of host range restriction and the discovery of a range of restriction and DNA modifying enzymes in *E. coli*. These enzymes were used in concert to cut and rejoin segments of DNA *in vitro* thereby generating novel or recombinant DNA molecules which were subsequently reintroduced into *E. coli* by plasmid transformation. The process of isolation and amplification of a fragment of DNA and any genes thereon in a recombinant molecule being known as molecular cloning. The efficiency of this technology in *E. coli* and other species has been significantly improved through parallel developments in both host strains and cloning vectors derived from endogenous plasmids and bacteriophage. Cloning strategies and the major technical developments in this area have been documented in the texts by Glover (1984) and Old and Primrose (1985).

Whilst gene cloning is now technically feasable in several other organisms such as *Bacillus* and yeast, the process is sufficiently advanced in *E. coli* that it is unlikely to be surpassed as the host/vector system of choice for the construction of gene libraries.

Utilisation of these systems has resulted in a quantum leap in our understanding of genome organisation, function and regulatory mechanisms in both eukaryotic and prokaryotic organisms

3.1.2 Engineering the phenotypic expression of cloned genes

Many of the elements which are responsible for the regulated expression of genes have been characterised (see Lewin, 1985). Cloned genes may be expressed in the new host if fused to an appropriate promoter and ribosome binding site, such that the gene can be transcribed and translated. *E. coli* is frequently used as an expression host for heterologous genes from both other prokaryotes and eukaryotic organisms (reviewed by T. J. R. Harris, 1983). The gene product accumulates intracellularly or may be directed to the extracellular periplasmic compartment by fusion with the signal peptide of an outermembrane component such as OmpA (Ghrayheb *et al.*, 1984).

Species of the genus *Bacillus* are attractive hosts for the expression of secreted gene products since large amounts of secreted proteins are produced naturally during the stationary phase of growth prior to sporulation (Priest, 1977). The promoter elements and regulatory sequences for *Bacillus* RNA polymerases are well characterised (Doi, 1982) and several heterologous gene products have been expressed in this organism. Furthermore, the promoter and signal peptide of the α-amylase gene have been utilised to direct the expression and secretion of heterologous gene products such as α-interferon (Palva *et al.*, 1983) and *E. coli* β-lactamase (Palva *et al.*, 1982). The use of secretion based expression systems can ease subsequent purification of the target protein since it is already separated from the intracellular proteins. *B. subtilis* therefore seems an ideal host for the expression of the subtilisin gene from *B. amyloliquefaciens*. The subtilisin BPN' gene is expected to be transcribed from its own promoter and should be secreted normally. *B. subtilis* produces only low levels of secreted proteases, however a strain is available which is completely deficient in extracellular protease production (Kowamura and Doi, 1984). Use of this strain, *B. subtilis* DB104, as an expression host will allow wild-type subtilisin and the mutants constructed

to be isolated free from contaminating endogenous proteases.

3.1.3 Complications associated with *B. subtilis*

The use of *B. subtilis* as a host for recombinant DNA is complicated by its mechanism of transformation. Genetic transformation was first observed in this species by Spizizen (1958). Unlike *E. coli*, where competence can only be induced by treatments which modify cell wall components (Hanahan, 1983), competence for DNA uptake by *B. subtilis* cells is a normal physiological growth state induced towards the end of exponential growth as nutrients are depleted. The mechanism of DNA uptake is complex and has been reviewed in detail (Dubnau, 1982).

Competent *B. subtilis* cells can be transformed by plasmid DNA but there is an absolute requirement for multimeric plasmid molecules (Canosi *et al.*, 1978). This causes practical difficulties when transforming *in vitro* generated recombinant molecules since the ligation conditions must be designed to favour multimer rather than monomer formation. This can only be achieved by using high DNA concentrations and availability of material may therefore be a problem. To overcome this difficulty, "shuttle" or "bridge" vectors (Gryczan, 1982) have been constructed which contain two origins of replication allowing them to be propagated in both *E. coli* and *B. subtilis* and also selectable markers which function in one or both hosts. Thus, DNA can be simply manipulated in *E. coli*, then transferred to *B. subtilis* once the required plasmid has been constructed.

3.2 Molecular cloning of the structural gene encoding subtilisin BPN'

3.2.1 Strategy

Earlier attempts to identify subtilisin specific recombinants in λ libraries were unsuccessful (Kenny and Brannigan, personal communication). Several *Bacillus* genes have been successfully cloned in *E. coli* from libraries constructed in plasmids, examples include the *B. subtilis* xylanase gene (Bernier *et al.*, 1983) and the α-amylase genes from *B. coagulans* (Cornelis, *et al.*, 1982) and *B. stearothermophilus* (Tsukagoshi *et al.*, 1984).

Wells *et al.*, (1983) have sucessfully isolated a subtilisin clone from a gene bank constructed in an *E. coli* / *B. subtilis* shuttle vector using minimally redundant oligonucleotide probes. The DNA sequence of the gene was communicated to us prior to publication, but the cloned gene and appropriate strains were not made available. It was therefore necessary to clone the gene independently. Inspection of the published DNA sequence of the subtilisin gene revealed that the gene contained no *EcoR*1 sites within the coding region or the promoter elements. A plasmid library constructed from *EcoR*1 restricted *B. amyloliquefaciens* chromosomal DNA should therefore contain an uninterrupted subtilisin BPN' gene. The gene should be easily identified using specific oligonucleotide probes synthesised from the known sequence of the gene.

Two oligonucleotide probes were synthesised, probe R had the sequence 5'-CCGAGCGTCGCTTAC-3' and was complementary to nucleotides 372 to 386 in the coding stand (numbering according to Wells *et al.*, 1983).
Probe Y had the sequence 5'-GATCAGCCCTTTTCC-3' and was complementary to nucleotides 1206 to 1220 in the non-coding strand. Probes

R and Y were a gift from J. Brannigan. Any recombinant hybridising to both probes has a high probability of being a genuine subtilisin clone.

3.2.2 Construction of a *B. amyloliquefaciens* gene library in *E. coli*

A gene library consisting of *Eco*R1 digested *B. amyloliquefaciens* chromosomal DNA cloned into the *Eco*R1 site of the plasmid pBR322 (Bolivar *et al.*, 1977) was constructed. The integrity of the chromosomal DNA purified from *B. amyloliquefaciens* was checked by electrophoresis through a 0.3% agarose gel in 0.5xTBE. Unrestricted λ DNA (48 kb) provided a high molecular weight marker. The chromosomal DNA migrated appreciably slower than the marker DNA, indicating a mean size of considerably greater than 50 kb. No shorter DNA fragments were seen confirming that the material had not been sheared during isolation. No contaminating RNA was detected. Aliquots (300 ng) of the DNA were digested with *Eco*R1 (1 unit) or *Sau*3A (1 unit) for 30, 60, or 120 min and the products analysed by electrophoresis through a 0.8% agarose gel in 0.5xTBE. Digests with both enzymes went to completion within 30 min as there was no further change in the size distribution of the DNA fragments at later time points. The *B. amyloliquefaciens* DNA was therefore considered suitable for the construction of the gene bank.

Chromosomal DNA (10 μg) was digested to completion with 40 units of *Eco*R1 in a 100 μl reaction volume containing High salt buffer (Maniatis *et al.*, 1982). After incubation at 37 ˚C for 2 hours, a 2 μl aliquot was analysed by gel electrophoresis to confirm that the reaction had gone to completion. The remaining DNA was phenol/chloroform extracted then recovered by ethanol precipitation. The DNA was then dissolved in 200 μl of sterile water giving a concentration of approximately 50 ng μl^{-1}.

The vector DNA, pBR322 (10 μg), was similarly restricted with *Eco*R1 and complete digestion confirmed by gel electrophoresis. The linearised vector was then dephosphorylated using calf intestinal phosphatase as described (Section 2.9). The vector DNA was phenol/chloroform extracted twice to ensure complete removal of the phosphatase then recovered by ethanol precipitation. The linearised pBR322 was dissolved in water at 50 ng μl^{-1}.

*Eco*R1 digested vector and chromosomal DNAs (100 ng of each) were ligated at 15 °C for 18 hr under the conditions described (Section 2.10.1). Competent *E. coli* LM1035 cells were transformed to ampicillin resistance with the ligated DNA, giving >10^4 transformants after incubation at 37 °C for 18 hr. Control ligation and transformation experiments demonstrated that there was zero background due to unrestricted pBR322 and only 2% from religated pBR322 (i.e. non-phosphatased). Therefore approximately 98% of the ampicillin resistant transformants should contain inserts. Transformation efficiency was 8x10^5 cfu. μg^{-1} with supercoiled pBR322.

3.2.3 Identification of subtilisin recombinants by hybridisation screening

The colonies were transferred to nitrocellulose filters and the plasmids amplified by exposure to chloramphenicol as decribed (Section 2.14.1). Chloramphenicol amplification can increase the copy number of ColE1 related plasmids from 20-50 per chromosome to around 3000 (Clewell, 1972). This increases the signal from hybridisation positives during screening with radiolabelled probes. The master plates were regrown at 37 °C for 8 hr then sealed and stored at 4 °C. *In situ* colony lysis and hybridisation screening was as decribed (Section 2.14). The filters were hybridised with 5' ^{32}P-labelled probe R at 37 °C over night using a probe concentration of 1 pmol ml^{-1} in 6xNET/Denhardts/SDS hybridisation buffer. Following hybridisation, the filters were washed briefly in 6xSSC at 4 °C to remove

excess probe, then given a further 3 x 45 min washes in 6xSSC at room temperature (18 ˚C). Autoradiography (24 hr, -70 ˚C with an intensifying screen) showed that the probe had hybridised to all the colonies, whilst the background on the filters was very low. The filters were rewashed at higher stringency to identify hybridisation positives. The calculated Td of probe R using the Wallace rules (Suggs *et al.*, 1982, see Appendix 2) is 50 ˚C. The filters were given 2 x 10 min washes in 6xSSC at 48 ˚C (Td-2 ˚C), then autoradiographed for a further 24 hr. Seven candidate hybridisation positives were identified. These were located on the master plates and picked onto duplicate LB-agar selective plates. Colonies close to the presumed positives were also picked. The plates were grown at 37 ˚C for 18 hr. The colonies were transferred to nitrocellulose filters and these were processed as described (Section 2.14). Following prehybridisation one filter was hybridised with probe R and the other with probe Y. Hybridisations were done at 37 ˚C for 16 hr with 1 pmol ml^{-1} of the required probe. The filter hybridised with probe R was washed in 6xSSC at 48 ˚C for 3x15 min, and the filter hybridised with probe Y at 44 ˚C for 3x15 min (Td for probe Y = 46 ˚C). Both filters were autoradiographed for 24 hr at -70 ˚C using preflashed film and an intensifying screen. A single recombinant remained hybridised with both oligonucleotide probes (Figure 3.1) and this was designated RY+. Cells from this colony were immediately stored at -70 ˚C as described (Section 2.2).

3.2.4 Analysis of RY+

The recombinant RY+ exhibited a "micro" colony morphology (<0.5 mm diameter) when propagated on selective plates and failed to grow at all in liquid media. On repeated re-streaking on selective media "normal" morphology colonies (about 1 mm diameter) arose spontaneously and at high frequency. These grew well in liquid media but were negative when
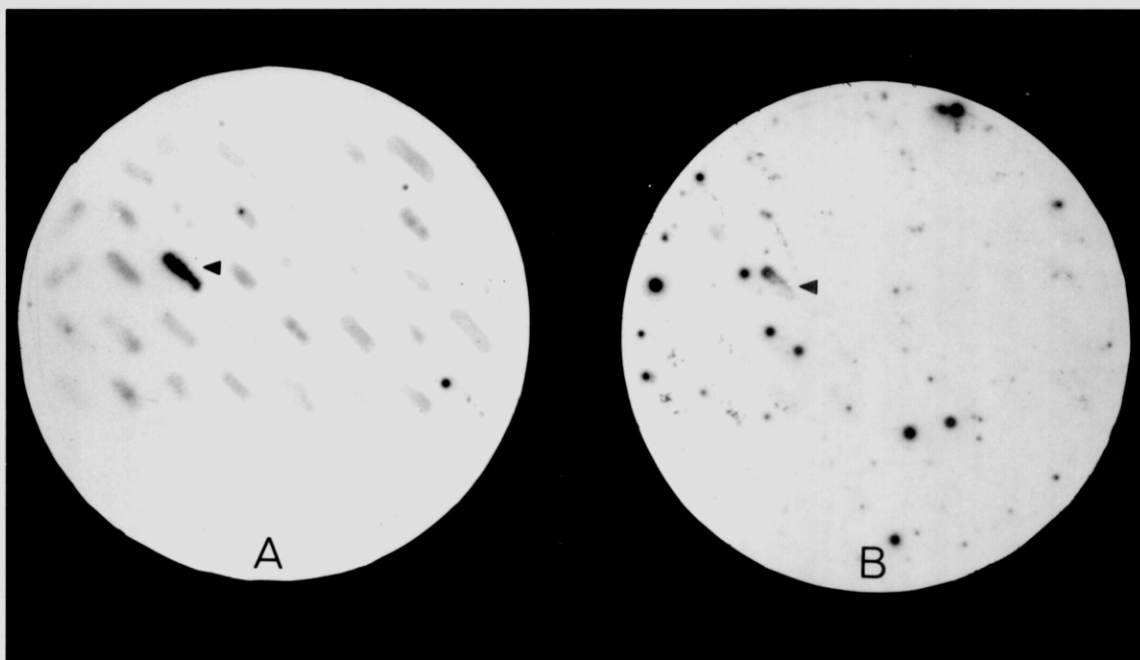
FIGURE 3.1:   Identification of RY+ by colony screening

Duplicate filters of putative hybridisation positives from the first round screen were reprobed with the two subtilisin specific oligonucleotides.

Filter A: Hybridised with probe R at 18 °C overnight, washed 2x5 min at 50 °C in 6xSSC.   Autoradiographed 48 hr at -70 °C with preflashed film and an intensifying screen.

Filter B: Hybridised with probe Y at 18 °C overnight, washed 2x5 min at 44 °C in 6xSSC.   Autoradiographed as above.

A single colony, indicated by arrow heads on the duplicate filters, remained hybridised to both probes when washed under stringent conditions.

screened with the specific probes. Thus, the recombinant isolated was unstable. This phenomenon has been observed by several other groups attempting to clone *Bacillus* DNA in *E. coli*, particularly for genes encoding hydrolytic enzymes.

Initially, plasmid DNA from RY+ could only be isolated by a modified procedure (Section 2.4.4), since cells harbouring the plasmid could only be grown on solid media. Colonies were scraped off selective LB-agarose plates, the cells washed and plasmid DNA purified using the alkaline/SDS procedure. Plasmid DNA recovered by this method was found to be a suitable substrate for restriction enzymes. Plasmid DNA purified from colonies grown on LB-agar plates failed to restrict, this is most likely due to copurificaton of sulphated polysaccharides from the agar which can inhibit restriction endonucleases.

Plasmid DNA (5μl) purified from RY+ was restricted with *Eco*R1 and the products analysed by agarose gel electrophoresis. Figure 3.2 (Panel B) shows the recombinant plasmid digested with *Eco*RI (lane d) and run on a 0.8% agarose gel in 0.5xTBE. Size markers were provided by λ DNA digested with *Hind*III (lane c). The lower band in lane d corresponds to linear pBR322 (4.4 kb) and the upper band is the inserted fragment of *B. amyloliquefaciens* DNA. The insert is approximately 10 kb in length as estimated from the adjacent size markers. This plasmid was designated pPT1.

Prolongedgrowth of *E. coli* LM1035 harbouring pPT1 on selective media eventually gave rise to "normal" sized colonies which remained positive when screened using the specific probes. These colonies grew well in liquid media and contained a plasmid indistinguishable from pPT1 (see below). Plasmid DNA was purified from a 200 ml culture to provide sufficient

FIGURE 3.2   Southern hybridisation analysis of the recombinant plasmid pPT1

Subtilisin specific probe R hybridises to a 10 kb *Eco*RI restriction fragment in both chromosomal DNA and the recombinant plasmid pPT1.   Southern transfers were as described by Maniatis *et al.* (1982).

Panel A: Southern transfer, filter hybridised with radiolabelled probe R and washed under stringent conditions (2x5 min at 50 °C in 6xSSC).

Track a   *Eco*RI restricted chromosomal DNA from *B. amyloliquefaciens.*

Track b   Fragment size markers; 5' ^{32}P labelled *Hind*III restricted λ DNA.

Panel B: Ethidium bromide stained agarose gel.

Track c   Restriction fragment size markers; *Hind*III restricted λ DNA.

Track d   *Eco*RI restricted pPT1 (unstable), fragments sizes are 3.65 kb (pBR322) and 10 kb (insert).

Panel C: Southern transfer, filter hybridised with radiolabelled probe R and washed under stringent conditions as above.

Track e   Fragment size markers; 5' ^{32}P labelled *Hind*III restricted λ DNA.

Track f   Fragment size markers; 5' ^{32}P labelled *Hinf*I restricted pAT153.

Track g   *Hind*III restricted pPT1 (stable) hybridising fragment is 3 kb.

Track h   *Eco*RI restricted pPT1 (stable), hybridising fragment is 10 kb.

76
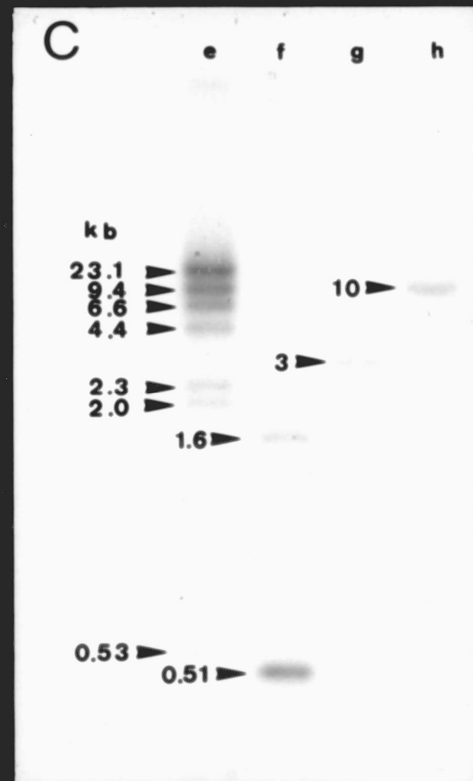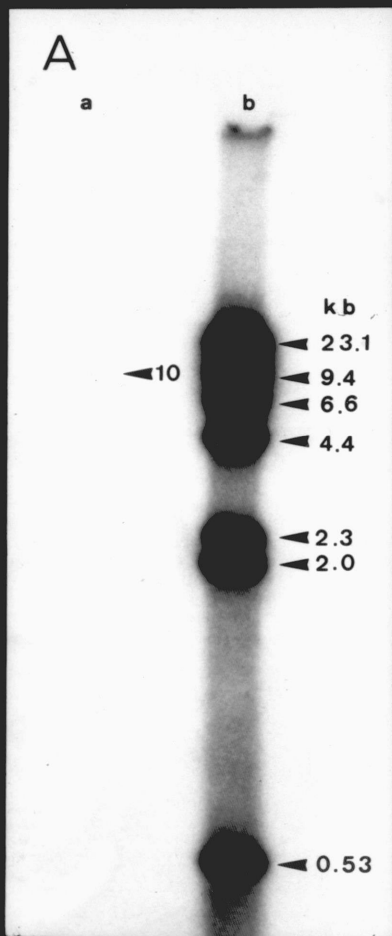
material for further analysis.

To confirm that the insert in the now stable pPT1 corresponded to a specific genomic DNA fragment, and that no gross alterations had occured as a result of DNA instability, Southern blot analysis of both genomic DNA and pPT1 was performed. Figure 3.2 shows that probe R specifically hybridises to restriction fragments of identical size in both pPT1 (Panel C, lane h) and genomic DNA (Panel A, lane a) on digestion with *Eco*R1. This result indicates that either no structural rearrangements or deletions had occurred, or that any such changes are too small to detect without extensive analysis. Figure 3.3 shows a physical map of the insert in pPT1 compared with that in pS4, an independent subtilisin clone (Wells *et al.*, 1983).

Fragments of *Eco*R1 digested pPT1 were separated on a 0.8% LGT agarose gel in 1xTAE and the 10 kb insert fragment purified as described by Maniatis *et al.*, (1982). The purified insert DNA was digested to completion with *Sau*3A then ligated with *Bam*H1 digested M13mp9. The plaques obtained on transfection of the ligated DNA into competent *E. coli* JM101 were screened with the specific probes. Hybridisation positives were identified with both probes. The probes R and Y were used as DNA sequencing primers on the hybridising M13 constructs. About 60 base pairs of subtilisin specific DNA sequence was obtained with each primer, confirming that the subtilisin BPN' gene was contained on the cloned fragment of *B. amyloliquefaciens* DNA (J. Brannigan, personal communication). Thus, the recombinant plasmid pPT1 was established to contain a 10 kb fragment of *B. amyloliquefaciens* DNA with both *Eco*R1 terminii correctly regenerated, and that the gene encoding subtilisin BPN' is contained within the fragment.

3.3 Expression of subtilisin BPN' in *Bacillus subtilis*

3.3.1 Vector selection

In order to obtain phenotypic expression of the subtilisin BPN' gene in a low protease producing stain of *B. subtilis*, the 10 kb insert was transferred to the shuttle vector pUB1662 (Docherty, 1980). This vector consists of the *Bam*H1 to *Eco*R1 fragments of the plasmids pUB110 (Lacey and Chopra, 1974; Gryczan *et al.*, 1978) and pBR325 (Bolivar *et al.*, 1978). pUB1662 was constructed by ligating *Bam*H1 digested pUB110 and pBR325 followed by deletion of the small *Eco*R1 fragment from the resulting plasmid. pUB1662 confers resistance to both ampicillin and kanamycin in *E. coli* and to kanamycin alone in *B. subtilis*. The structure of pUB1662 is shown in Figure 3.4. This plasmid is capable of replication in both host strains and should be suitable for the expression of subtilisin in *B. subtilis*, especially so since Wells *et al.*, (1983) used a shuttle vector, pBS42, to achieve expression of their subtilisin gene.

3.3.2 Construction of pPT10

pPT1 DNA (15 μl containing approximately 200 ng) purified from colonies propagated on LB-agarose plates (i.e. prior to the isolation of the stable form) was restricted with *Eco*R1. Following phenol/chloroform extraction and ethanol precipitation, the products were ligated with 100 ng of *Eco*R1 digested, phosphatased pUB1662 as described (Section 2.10.1). Transformation of competent *E. coli* LM1035 with the ligated DNAs gave rise to a high background of religated pBR322 after overnight growth at 37 °C, however after prolonged incubation (>24 hr) a number of smaller colonies became visible. These were similar in morphology to the original hybridisation positive RY+. Four of the transformants were positive when screened with

the specific probes, but only 3 of these grew in liquid culture. Digestion of plasmid DNA, isolated from small scale cultures, with *Eco*R1 gave fragments consistent with the 10kb fragment having been transferred to pUB1662 (8kb). This plasmid was designated pPT10. Restriction analysis of pPT10 purified from large scale (200 ml) cultures confirmed that it harboured a fragment of identical size to that in pPT1.

Competent *B. subtilis* IG20 cells were transformed with 500 ng of pPT10 DNA as decribed (Section 2.11.2). 840 primary kanamycin resistant transformants were obtained after overnight incubation on LB-agar plates containing kanamycin (25μg ml^{-1}) and glucose (0.2%) at 37 °C. The transformants were picked onto selective LB-agar plates containing 1% milk solids and incubated at 37 °C overnight. There was no significant increase in the amount of secreted protease activity compared with IG20 harbouring the parent vector. Furthermore, colonies presumed to harbour pPT10 failed to grow in liquid media. At this point further investigation of the unstable transformants was abandoned in favour of attempts to reduce the size of the fragment harbouring the subtilisin gene.

3.3.3 Construction of pPT2

Restriction mapping and Southern blot analysis of pPT1 indicated that the subtilisin gene was entirely contained within a 3 kb *Hind*III restriction fragment (Figure 3.2; Panel C, lane g) located towards one end of the 10 kb insert. In fact the *Hind*III site 3' of the subtilisin gene is located in the vector adjacent to the original *Eco*R1 cloning site. 2.5 μg of pPT1 was completely digested with *Hind*III and the products resolved by electrophoresis through a 1% agarose gel in TAE. The 3 kb band was well separated from any other fragments and was purified using NA45 membrane as described (Section 2.8.2). The vector DNA, pAT153 (2.5μg, Twigg and

Sherratt, 1980) was also digested with *Hind*III and dephosphorylated. Vector and target DNAs (50 ng of each) were ligated and then used to transform competent *E. coli* LM1035 cells. Transformants were selected by growth on LB-agar containing ampicillin (100 μg ml^{-1}). Plasmid DNA was isolated from 12 transformants. On restriction with *Hind*III, one of the transformants gave fragments of 3.65 kb and 3 kb corresponding to the vector and insert respectively. This plasmid was designated pPT2, it showed no signs of structural instability and was used in further restriction mapping experiments and as a potential source of insert DNA for later constructs. The restriction map generated for pPT2 was entirely consistent with the published DNA sequence of the subtilisin gene (Wells *et al.*, 1983) and mapping data of sites further 5' from the gene, with the exception of an additional *Pvu*II site located towards the 5' *Hind*III site. This site may have been overlooked in the map published by Wells *et al.*, (1983), or may have arisen as a result of DNA rearrangement. Alternatively it may reflect variation in DNA sequence between the donor strains from which chromosomal DNA was prepared. The structure and orientation of the insert in pPT2 is shown in Figure 3.3.

The 3 kb fragment isolated from pPT1 was also ligated into *Hind*III digested M13mp9 and transfected into competent *E. coli* JM101. Recombinants were obtained containing the insert in both orientations. The nucleotide sequence of the entire coding region of the enzyme was determined by using a series of oligonucleotides to prime DNA sequencing reactions at approximately 200 base intervals along the gene (Table 4.1). The sequence of the gene was confirmed to be identical to that published (J. Brannigan, personal communication; Wells *et al.*, 1983; Vasantha *et al.*, 1984).

Figure 3.3: Physical map of the cloned fragments in pPT1 and pPT2
Comparison with the insert in pS4 (Wells et al., 1983).

A: Insert in pPT1
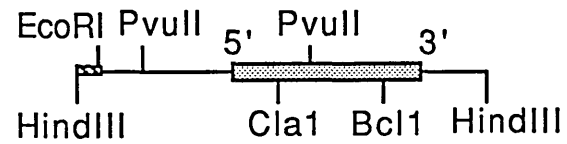
B: Insert in pPT2

C: Insert in pS4
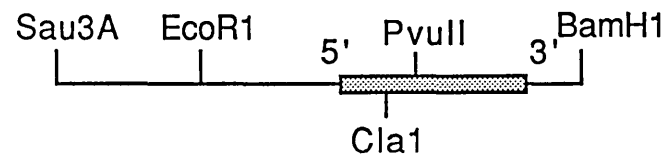
3.3.4 Construction of pPT20

Repeated attempts to purify the 3 kb insert fragment away from the 3.65 kb vector linear from pPT2, using either NA45 membrane or LGT agarose failed since the two fragments could not be separated sufficiently well by electrophoresis. Contamination of the insert DNA with vector DNA, which subsequently religated with high efficiency, gave anomalously high backgrounds of ampicillin resistant colonies. To overcome this problem one of the M13 recombinants harbouring the 3 kb *Hind*III insert was selected as a source of DNA, since vector religation in this case can not give rise to ampicillin resistant colonies.

Insert DNA was prepared using a modification of the procedure of Norris *et al.*, (1983). M13 universal primer (5 pmol) was annealed to 20 μl (containing approximately 2 pmol of DNA) of template DNA in a 40 μl reaction volume containing 10 mM Tris-HCl pH 8.0 and 10 mM MgCl$_2$. The mixture was boiled for 2 min, then slowly cooled to 20 $^\bullet$C, 5 μl of 10xdNTP mix (2 mM each of dCTP, dGTP, dATP and TTP in water) and 6 units of Klenow polymerase were added. The volume was adjusted to 50 μl with water and the reaction incubated at 25 $^\bullet$C for 30 min. Following heat inactivation at 65 $^\bullet$C for 5 min, the primer extended (partially double stranded) DNA was digested to completion with 40 units of *Hind*III in a 100 μl volume containing Medium salt buffer. An aliquot of the restricted DNA was analysed by gel electrophoresis through a 0.8% agarose gel in 0.5xTBE. Two fragments were present, the 3 kb insert and the 7 kb M13 vector. The remaining DNA was phenol/chloroform extracted and recovered by ethanol precipitation, then dissolved in 20 μl of water.

The DNA was then blunt-end repaired using 0.5 units of T4 DNA polymerase in the presence of 25 mM Tris-HCl pH 8.0; 6.6 mM MgCl$_2$;

5 mM β mercapto-ethanol; 0.5 mM EDTA and 200 μM of each dNTP. The reaction (30 μl) was incubated at 20 ˙C for 15 min then heat inactivated at 65 ˙C for 10 min. Phosphorylated EcoR1 linkers (d[GGAATTCC], New England Biolabs) were ligated to the blunt end repaired DNA. The reaction contained 10 μl of the T4 polymerase treated DNA and 3 μM linkers in 50 mM Tris-HCl pH 7.6; 10 mM $MgCl_2$; 20 mM DTT; 1 mM ATP and 2 units of T4 DNA ligase in a final volume of 20 μl. The reaction was incubated at 15 ˙C for 20 hr. The ligated DNAs were then digested to completion with 25 units of EcoR1 in HIGH salt buffer (50 μl) for 2 hr at 37 ˙C. The bulk of the excess linkers released by EcoR1 digestion were removed by the addition of ammonium acetate to 2 M and ethanol precipitation at -70 ˙C for 20 min. After centrifugation for 10 min at 4 ˙C, the precipitated DNA was washed with 70% ethanol, dried then dissolved in 20 μl of water.

The required EcoR1 linked fragment (3 kb) was purified using NA45 membrane after electrophoresis through a 1% agarose gel in TAE and dissolved in 20 μl of water. 10 μl of the linked DNA was ligated with 100 ng of EcoR1 digested, phosphatased pUB1662 in a 20 μl reaction volume. Competent E. coli DH1 cells were transformed to ampicillin resistance with 5 μl of the ligated DNA. Plasmid DNA was isolated from 12 of the 38 transformants obtained, digested with EcoR1 and analysed on a 1% agarose gel in 0.5xTBE, 11 of the transformants examined contained the 3kb fragment harbouring the subtilisin gene. The insert in all 11 was shown to be in in the same orientation by digestion with PvuII. One of these constructs designated pPT20, was selected for transfer into B. subtilis. Further restriction analysis of pPT20 DNA confirmed the presence of the additional PvuII site and showed the orientation of the subtilisin gene in the plasmid. The structure of pPT20 is shown in Figure 3.4.

A: pUB1662

SUBTILISIN GENE

pBR325 DERIVED

pUB110 DERIVED
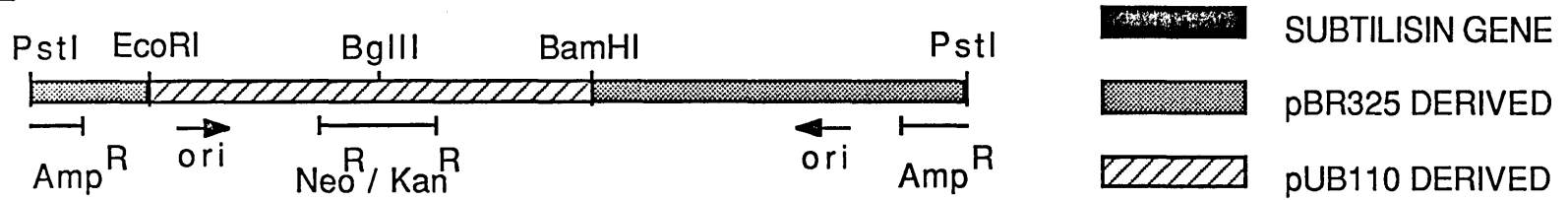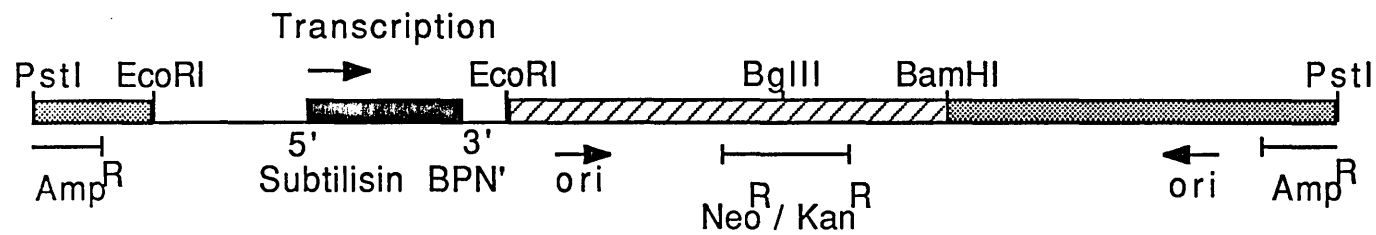
B: pPT20

Scale (kb)
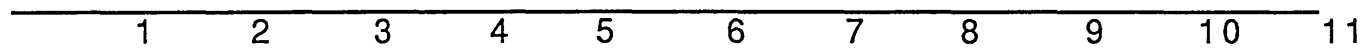
Figure 3.4: Physical maps of pUB1662 and pPT20

E. coli DH1 is *rec*A⁻ and therefore unable to perform homologous recombination, consequently it can not generate plasmid multimers. Since only multimeric plasmid DNA is able to transform competent *Bacillus* cells it was necessary to produce the plasmid multimers *in vitro*. Plasmid DNA was therefore digested to completion with a restriction enzyme which cuts the vector uniquely, then religated at high DNA concentrations to favour the formation of multimers.

3.3.5 Multimerization of pPT20.

pPT20 DNA (5 μg) was digested to completion with 20 units of *Bam*H1 in MEDIUM buffer in a reaction volume of 50μl. After incubation at 37 °C for 2 hr, the reaction was terminated by incubation at 65 °C for 15 min. The linear pPT20 was religated to itself to produce multimers in 50 mM Tris-HCl pH7.6; 10 mM MgCl$_2$; 20 mM DTT and 1 mM ATP with 2 units of T4 ligase in a final volume of 100 μl. The reaction was incubated at 20 °C for 2 hr. The formation of multimers, <5mers and probably some greater, was confirmed by electrophoresis of a 5 μl aliquot of the ligated DNA (containing 250ng DNA) on a 1.0% agarose gel in 0.5xTBE, using linear unligated pPT20 and uncut pPT20 as markers.

3.3.6 Transfer of pPT20 to *B. subtilis*

500ng of multimerized pPT20 was transformed into competent *B. subtilis* IG20 and transformants were selected on LB-agar plates containing kanamycin (25 μg ml⁻¹). Several kanamycin resistant transformants were picked and restreaked onto LB-agar plates containing kanamycin and milk solids (1% w/v Marvel). A clear halo of 1-2 cm diameter was visible around each colony harbouring pPT20 after incubation at 37 °C for 18 hr (Figure 3.5), whereas *B. subtilis* IG20 colonies harbouring the parent vector
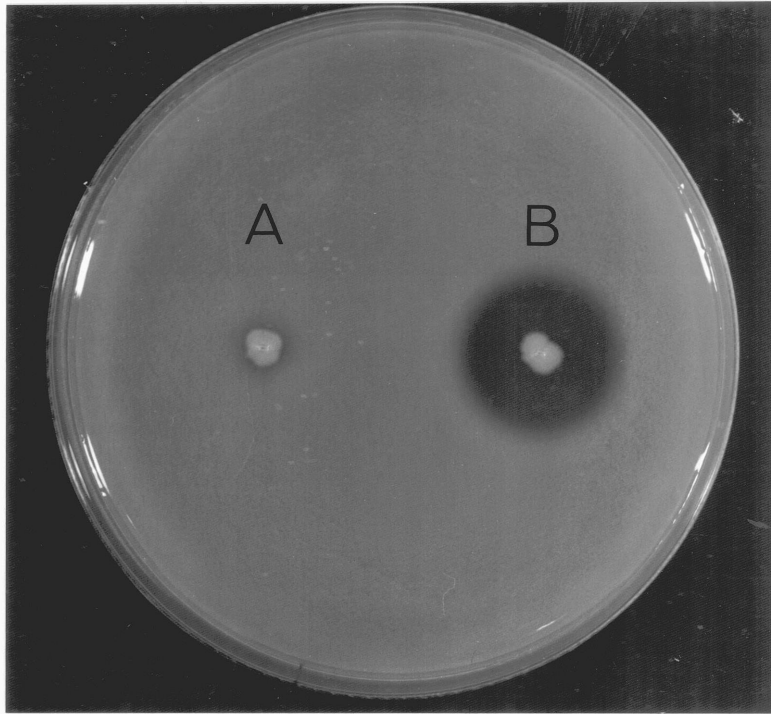
FIGURE 3.5:   Secretion of subtilisin by *B. subtilis* IG20 harbouring

pPT20.

Colonies were grown on LB-agar containing skim milk powder (1% w/v

Marvel) and kanamycin (25 μg ml^{-1}) at 37 $^\circ$C for 24 hr.

A: *B. subtilis* IG20 harbouring the shuttle vector pUB1662, note that after

prolonged growth (>20 hr) a faint zone of proteolysis is observed due to

endogenous protease production.

B: *B. subtilis* IG20 harbouring the subtilisin expression vector pPT20.

Secreted subtilisin results in a characteristic clear halo surrounding the

colony which is due to proteolysis of the milk proteins in the agar.

pUB1662 gave no clearing zones after 18 hr incubation and only very faint and small, 1-2 mm beyond the colony, clearing zones after prolonged incubation (>24hr, Figure 3.5). One protease producing colony was restreaked several times to ensure that it was clonally pure.

Subtilisin BPN' purified from *B. amyloliquefaciens* culture supernatants is inhibited by PMSF but not by EDTA (Wells *et al.*, 1983). 50 μl aliquots of *B.subtilis*/pPT20 culture supernatant, grown in LB-broth containing kanamycin (25 μg ml^{-1}) and glucose (1%) for 24 hr, were assayed for protease activity by monitoring the hydrolysis of the chromogenic substrate suAAPFpN (Section 5.4). The release of nitroanilide was inhibited by the addition of PMSF to 1 mM and was insensitive to EDTA at 10 mM (A.J. Russell, personal communication). There was no detectable protease activity in supernatants from cultures of *B. subtilis* IG20 harbouring the parent vector pUB1662.

These results are consistent with the secreted protease being subtilisin BPN'. Furthermore the kinetic constants, k_{cat} and K_m, for the hydrolysis of suAAPFpN by the enzyme secreted by *B. subtilis* harbouring pPT20 are 57 sec^{-1} and 0.15 mM respectively. These values are identical to those published for subtlisin BPN' for the same substrate (Polgar and Bender, 1983) and suggest that the secreted protease is authentic subtilisin BPN'.

3.3.7 Stability of pPT20 in *B. subtilis* IG20

It was observed that restreaking colonies of *B. subtilis* IG20 harbouring pPT20 which had been stored at 4 °C for periods from 1-4 weeks frequently gave rise to kanamycin resistant colonies which did not produce protease (<95% in one case). Similarly if colonies from plates more than a day old were used to innoculate liquid cultures, grown for 24hrs under

kanamycin selection and plated out onto LB-agar containing kanamycin and milk solids, then colonies also arose which did not produce protease. Plasmid DNA isolated from these colonies was smaller than pPT20, indicating that segments were being deleted and that the plasmid is unstable. It is generally assumed that structural instability is related to plasmid size and it is known that in this case the parent vector does not exhibit instability even on prolonged propagation in IG20 (Docherty, 1980).

A new expression plasmid, based on pUB110 was constructed to minimise further instability problems leading to a reduced yield of secreted subtilisin. This plasmid lacks an origin of replication which can function in *E. coli* but is considerably smaller than pPT20 and therefore should present a smaller target for recombination thereby reducing the risk of deletions. Unfortunately this prevents the re-construction of expression plasmids for each mutant in *E. coli*, requiring the DNA manipulations to be performed in *B. subtilis* directly.

The plasmid pUB110 is a naturally occuring *Staphylococcus aureus* plasmid which carries a kanamycin resistance gene (Lacey and Chopra, 1974). Gryczan *et al.*, (1978) demonstrated that this vector can also replicate in *B. subtilis* and confers kanamycin resistance. pUB110 is small (4.5 kb.) compared with pUB1662 (8 kb) and has a unique *Eco*R1 site suitable for the insertion of the 3 kb fragment harbouring the subtilisin gene. The expression plasmid could potentially be reduced from 11 kb to 7.5 kb and should be sufficiently stable for the production of large amounts of wild-type and mutant subtilisins. Figure 3.6 shows a physical map of pUB110.
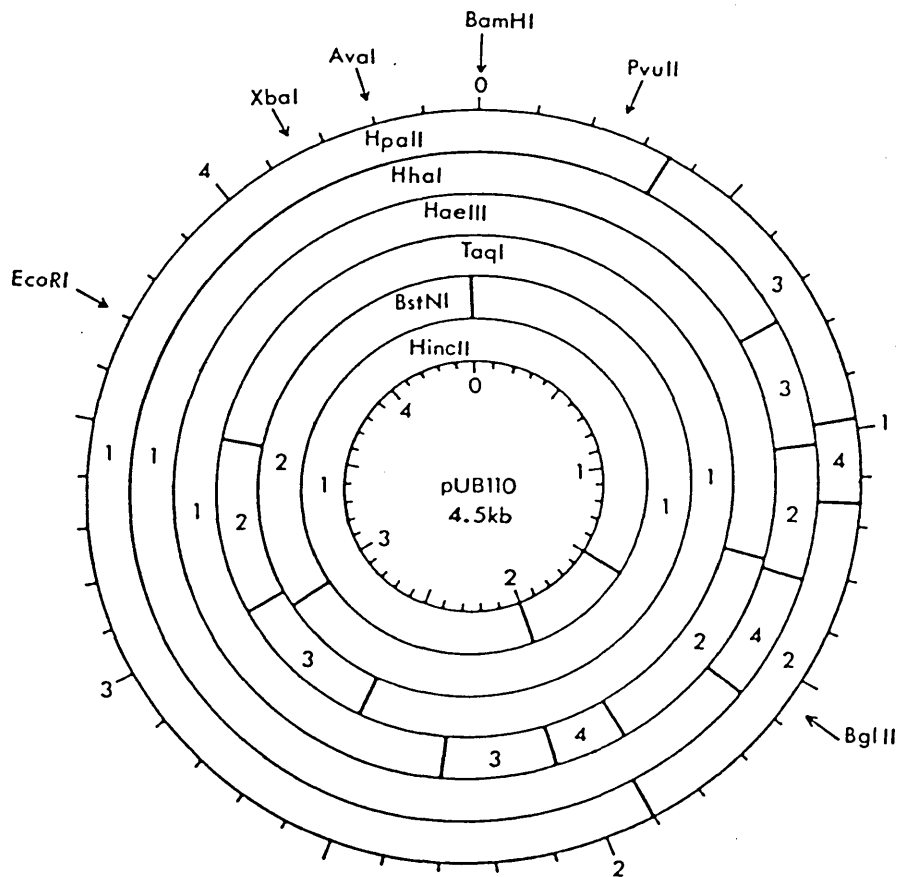
FIGURE 3.6:   Physical map of pUB110

Redrawn from Jalanko *et al.*, 1981.   The *Bgl*II site is within the neomycin (kanamycin) resistance determinant; the limits of the resistance gene are unknown.   The origin of replication has been mapped at 3.2 (Scheer-Abramowitz *et al*, 1981) with replication proceeding in an anticlockwise direction.

3.3.8 Construction of pPT30

20 μg of pPT20 DNA were digested to completion with 25 units of *Eco*R1. The products were separated on a 0.8% agarose gel in TAE and the required 3 kb insert purified using NA45 membrane. After recovery the DNA was dissolved in 25 μl of 0.1xTE. 2.5 μg of pUB110 was digested with *Eco*R1 and the resulting linearised DNA recovered by ethanol precipitation after phenol/chloroform extraction, then dissolved in 25 μl of 0.1xTE. Aliquots (1 μl) of each DNA were subjected to electrophoresis through a 0.8% gel in 0.5xTBE. Judging from the intensity of fluorescence of the bands after staining with ethidium bromide, a volume ratio of 1:2 (vector to target) was chosen to give a 1:1 ratio of cohesive ends during ligation. 500 ng of pUB110/*Eco*R1 (5 μl) was ligated with 10 μl of the target DNA under the conditions described (Section 2.10.2). The production of multimeric DNA was confirmed by gel electrophoresis of 2 μl aliquots of the ligated and unligated DNAs. The remaining ligation products were used to transform competent *B. subtilis* DB104 cells to kanamycin resistance. Many of the kanamycin resistant colonies obtained also exhibited a clearing zone characteristic of subtilisin production. Plasmid DNA isolated from 4 of the protease producing transformants, gave fragments of 3 kb and 4.5 kb on digestion with *Eco*R1 confirming transfer of the subtilisin gene into pUB110. This plasmid construct was designated pPT30. Plasmid DNA was purified from a 400 ml culture as described (Section 2.4.6). The orientation of the inserted fragment was determined by digestion with *Pvu*II. The structure of pPT30 is shown in Figure 3.7.

3.3.9 Stability of pPT30 in *B. subtilis* IG20

When *B. subtilis* IG20 cells harbouring pPT30 were restreaked either from freshly grown or stored plates there was no indication of plasmid
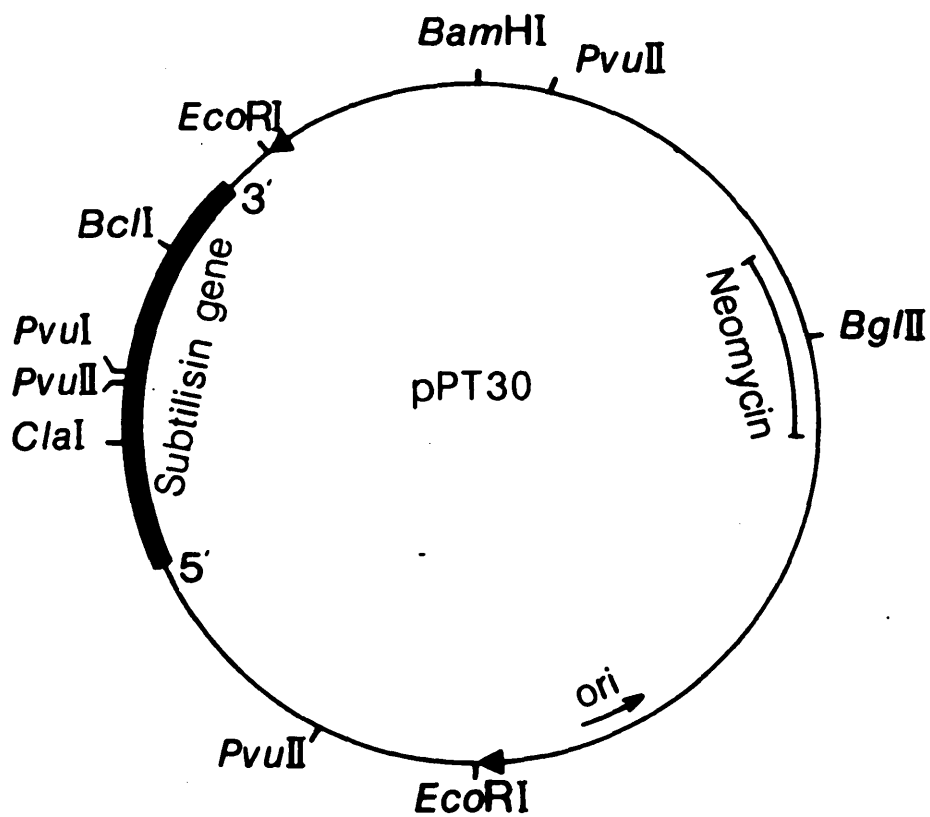
FIGURE 3.7:   Physical map of the subtilisin expression vector pPT30.
The subtilisin BPN' gene (solid block) is contained within a 3 kb EcoRI
fragment cloned into the unique EcoRI site of pUB110 (4.5 kb).   The
junctions of vector and insert sequences are indicated by the arrow heads.

instability. Similarly cultures grown for 24 hr contained no protease minus cells. Thus, pPT30 appears completely stable in *B. subtilis* over the period required for the production of enzymes from large scale culture.

3.4 Mutagenesis of the subtilisin BPN' gene

3.4.1 Construction of M13aprWT

Mutants were to be constructed by oligonucleotide directed mutagenesis of the subtilisin gene cloned in M13, then expressed in *B. subtilis* on transfer of the mutated gene to pUB110. Rather than have to add *Eco*R1 linkers to the *Hind*III fragment containing the subtilisin gene cloned in M13 each time expression of a mutant was required, it would be preferable to perform mutagenesis directly on the *Eco*RI linked fragment from pPT20 recloned in M13mp9.

To achieve this, 2 μg of pPT20 was digested to completion with 5 units of *Eco*R1 and the products separated by gel electrophoresis through 1% agarose in TAE, and the 3 kb fragment purified using NA45 membrane. M13mp9 DNA (1 μg) was digested with *Eco*R1 then dephosphorylated using calf intestinal phosphatase as described (Section 2.9). Purified insert DNA (approximately 25 ng) from pPT20 was ligated with 50 ng of the M13 vector. The ligation products were transfected into competent *E. coli* JM101 and plated out in H-top agarose containing X-gal (0.15 mg ml^{-1}) and IPTG (0.2 mg ml^{-1}), log phase JM101 cells provided the lawn. Template DNA was prepared from the 27 white plaques obtained after incubation at 37 $^{\circ}$C overnight. The presence and orientation of the insert was determined by dot-blotting using radiolabelled probes R and Y. Ten templates hybridised with probe Y and therefore contained the non-coding strand of the gene (i.e. identical to mRNA). Thirteen of the remaining templates hybridised

EcoRI

5'

SUBTILISIN BPN'

3'

M13aprWT
10.1 kb

Priming

EcoRI

▤▤▤   Polylinker

——   M13mp9

----   *B. amyloliquefaciens*
                  DNA
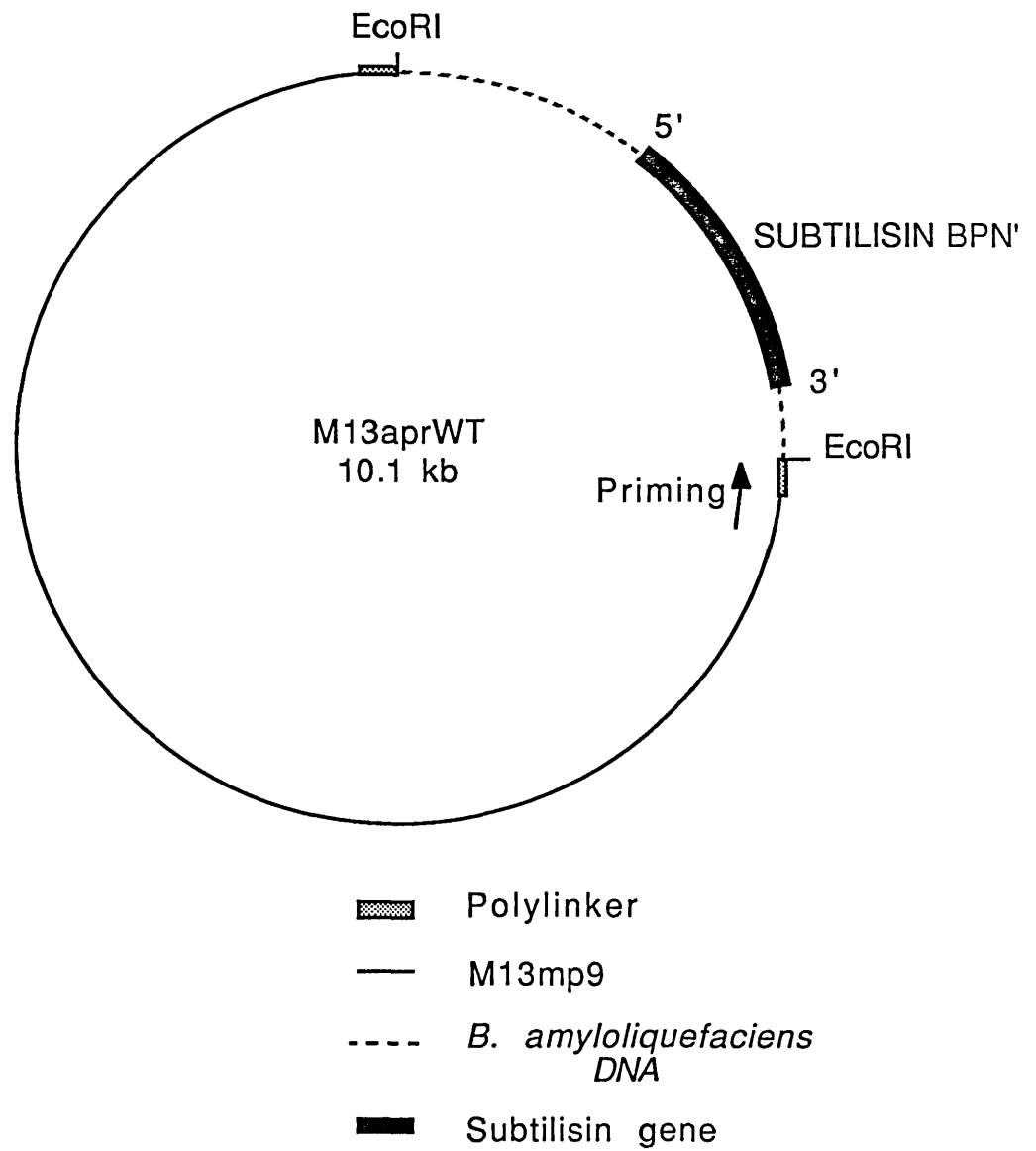
▬▬   Subtilisin gene

Figure 3.8:   Structure of M13aprWT

The non-coding strand of the subtilisin BPN' gene
is contiguous with the + strand of the viral DNA.

with probe R and these therefore contained the coding strand of the gene (i.e. complementary to mRNA). Four of the templates were negative and therefore presumed to have lost all or part of the subtilisin gene. Following plaque purification of the hybridisation positives, phage supernatants and template DNAs were stored for future use.

Restriction mapping data indicated that the EcoR1 site 3' of the subtilisin gene was about 400 bp down stream from the end of the coding region. To allow the universal primer to be used for sequencing into the 3' region of the gene, one of the M13 recombinants containing the non-coding strand of the gene was selected for further use. This construct was designated M13aprWT and provided the template DNA for sequencing and mutagenesis of the wild-type gene. The structure of M13aprWT is shown in Figure 3.8.

3.4.2 Stability of M13aprWT in JM101

Initially all manipulations involving M13 were performed in E. coli JM101. However, some batches of template DNA appeared to contain deletions as judged by their increased electrophoretic mobility on agarose gels. Dot-blot experiments confirmed the loss of the subtilisin gene. Other workers have observed that inserts in M13 of greater than about 1 kb can result in structural instability problems when JM101 is the host strain. The instability in this case is probably due to recA mediated recombination. RecA⁻ mutants have been used to succesfully overcome structural instability problems of both plasmid and phage recombinants (Yanisch-Perron et al., 1985).

E. coli TG2 is a recA⁻ derivative of JM101 (M. Biggin and T. Gibson, personal communication). This strain was constructed by P1 phage transduction and contains the strong recA⁻ allele from E. coli HH1210.

M13aprWT appeared to be completely stable when propagated in TG2. All subsequent manipulations involving M13 were performed in TG2 (except mutagenesis, see Section 4.4.3).

3.5 Discussion

3.5.1 Summary

The recombinant plasmid pPT1 was isolated from a gene bank of *Bacillus amyloliquefaciens* DNA cloned in the plasmid vector pBR322, using specific oligonucleotide probes. pPT1 has an insert of approximately 10 kb. and partial DNA sequence data confirmed that the subtilisin BPN' gene is contained within the inserted fragment. The subtilisin gene was sub-cloned on a 3 kb *Eco*R1 linked fragment into the shuttle vector pUB1662 to give pPT20, and into pUB110 to give pPT30. Active subtilisin is secreted at high levels by *B. subtilis* harbouring both pPT20 and pPT30. The same fragment was transferred to M13mp9 to give M13aprWT, which was used for sequence determination and site-specific mutagenesis of the subtilisin gene.

3.5.2 Structural DNA instability

The progress of this study was severely hindered by structural DNA instability problems which affected both pPT1 and M13aprWT in *E. coli* and pPT20 in *B. subtilis*. Initially *E. coli* LM1035 harbouring pPT1 exhibited a micro-colony morphology (<0.5 mm diameter after 18hrs growth at 37 °C) and failed to grow in liquid media. pPT1 DNA isolated from colonies grown on solid media was frequently contaminated with several smaller plasmid species. These were presumed to be deleted forms of pPT1, since segregants harbouring them were negative when screened with the specific oligonucleotide probes. Transfer of pPT1 to *E. coli* DH1, which is *recA⁻*,

did not alleviate the structural instability problems.

Yang *et al.*, (1983) experienced similar problems during the cloning of the *B. subtilis* α-amylase gene. A λ Charon 4A recombinant harbouring the α-amylase gene gave micro-plaques and was unstable during subsequent rounds of screening, giving rise to many negative plaques. An expressing sub-clone in pBR322 was also unstable, cells harbouring it frequently lost the ability to produce α-amylase and failed to grow in liquid media.

The α-amylase genes from *B. licheniformis* (Sibakov and Palva, 1984) and *B. coagulans* (Cornelis *et al.*, 1982) were similarly unstable during their isolation in *E. coli* and reconstruction into plasmid expression vehicles. However, the *B. stearothermophilus* α-amylase gene (Tsukagoshi *et al.*, 1984) and the xylanase gene from *B. subtilis* (Bernier *et al.*, 1983) were both stably maintained on plasmids in *E. coli*. Many other *Bacillus* genes are also stable when cloned in *E. coli*, for example: the *spo*0A locus (Ferrari *et al.*, 1984); tRNA genes (Vold and Green, 1984); glutamine synthetase (Fisher *et al.*, 1984). The instability observed with pPT1 was transient and a stable form was eventually isolated. *E. coli* cells harbouring the stable form grew satisfactorily in liquid media and sufficient DNA for subsequent manipulations could be isolated.

The reasons for the observed structural DNA instability remain unclear It is known that some *Bacillus* gene products are toxic when expressed in *E. coli*, for example the *spo*IIAC gene product (Yudkin, 1986). It is possible that instability arises from the selection for host cells with reduced or eliminated production of deleterious gene products from the cloned DNA. The deleterious proteins could be from the gene of interest or encoded elsewhere on the cloned segment of DNA. It is also possible that *Bacillus* DNA is intrinsically unstable in *E. coli* or susceptible to an as yet

uncharacterised host restriction/modification system. Over 100 restriction/modification systems have been identified for *E. coli* alone (Roberts, 1986). It is likely that the initial instability of pPT1 was not due to the subtilisin gene itself since pPT1 became stable and later plasmid constructs were also stable. Furthermore, the subtilisin gene is unlikely to express in *E. coli* due to its requirement for a minor RNA polymerase holoenzyme which is not known to occur in this species (Wong *et al.*, 1984).

M13*apr*WT was unstable when propagated in *E. coli* JM101, and gave rise to deleted forms which appeared smaller and failed to hybridise with the specific probes. It would appear that in this case instability was due to *rec*A mediated recombination, since the construct was perfectly stable in the *rec*A⁻ host TG2. It is widely known that inserts in M13 vectors of greater than 1 kb are frequently unstable (Messing, 1983). This may be due to the proximity of the inserted fragment to the origin of replication which may be sensitive to transcriptional interference. New *rec*A⁻ hosts for M13 have been constructed to alleviate these problems (Yanish-Perron *et al.*, 1985).

M13*apr*WT gave small plaques in both *E. coli* JM101 and TG2, this seems to be a common observation with M13 recombinants harbouring inserts of greater than 2 kb. Prior to the use of *E. coli* TG2 as the host strain, the differential growth of phage with full length inserts, compared to those containing deletions and giving larger plaques, allowed the deletants to be avoided.

pPT20 was constructed by inserting the subtilisin gene into the *E. coli* / *B. subtilis* shuttle vector (or bridge vector) pUB1662. Whilst pPT20 was stable in *E. coli*, it rapidly lost the ability to direct the expression of protease once transferred to *B. subtilis*, and deleted forms of the plasmid could be detected. The parent vector, in common with other

shuttle vectors, is stably maintained in both hosts and can be transferred between them without suffering deletions or rearrangements. There are several reports of recombinant shuttle vectors which become unstable on transfer from *E. coli* to *B. subtilis* (Goebel *et al.*, 1978; Kreft *et al.*, 1978; Kreft *et al.*, 1983; Ishiwa and Tsuchida, 1984).

The factors affecting the stability of shuttle vectors containing inserted DNA during transfer to and propagation in *B. subtilis* have been examined by Ostrof and Pene (1984). They concluded that an as yet uncharacterised host restriction-modification may exist which can damage heterologously propagated DNA. Whilst this mechanism appears responsible for low transformation efficiencies and deletions during actual DNA uptake by competent *Bacilli*, it fails to explain the instability of some plasmids during propagation. However they did observe that once the deletions had occured, the resulting plasmids became stable and did not undergo further rearrangement.

Uhlen *et al.*, (1981) demonstrated that deletions occured in segments of $\Phi105$ DNA, cloned into the shuttle vector pHV14 during propagation in rec^+ or *recE B. subtilis* hosts which were lysogenic or non-lysogenic for $\Phi105$. Extensive deletions occured during long term propagation irrespective of the *rec* character of the host or the existence of base pair homology between the plasmid and the chromosome. Thus, the use of a *recE* mutants of *B. subtilis* as hosts for subtilisin expression would probably not alleviate the instability problems with pPT20.

Whilst DNA instability problems diminish the value of *B. subtilis* as a host for cloning, the ability to engineer high level expression and secretion of a gene product into the culture supernatant makes this organism useful as an expression host.

The reason for the failure to obtain protease producing transformants of *B. subtilis* using pPT10 may have been due to the oversight that *E. coli* LM1035 is *recA⁻* and should not therefore produce plasmid multimers. The plasmid DNA used to transform *B. subtilis* was not multimerised *in vitro*. Whilst a number of primary kanamycin resistant transformants were obtained, none of these grew in liquid media. Given the difficulties experienced with a later construct, pPT20, also based on pUB1662, shuttle vectors were abandoned in favour of a small vector, pUB110, not known to suffer from instability problems on introduction of cloned segments of DNA. The plasmid pPT30 was therefore constructed, this contains the subtilisin gene on a 3 kb *EcoR1* fragment cloned in to the unique *EcoR1* site of pUB110.

The apparent stability of pPT30 even during long term propagation in *B. subtilis* suggests that the earlier problems could have been directly due to the segment of pBR325 used in the construction of pUB1662. Alternatively instability could be purely size related, the larger plasmid constructs placing an unacceptable metabolic load on the host cell, which responds by deleting non-essential regions of the plasmid. The properties of pPT30 seem to confirm that smaller plasmids are less likely to undergo structural alterations.

3.5.3 Frequency of subtilisin specific recombinants

Only one recombinant from over 10,000 screened was shown to contain the subtilisin gene. This begs the question as to whether the gene library large enough? Calculations of the probability that a particular gene library will represent the entire genome of the organism assume that the library is random, that the size distribution of the cloned fragments is quite narrow and that the average insert size is known. Clearly none of these

assumptions apply here. The library is not random, being constructed from a complete *EcoR*1 digest of *B. amyloliquefaciens* DNA, and the size of fragments cloned vary from <500 bp to >10 kb. Smaller fragments also tend to be ligated into the vector in preference to larger fragments resulting in a biased representation of sequences in the library. The larger fragments therefore being under-represented and the smaller fragments over-represented. It is possible that the recombinant identified was the only subtilisin specific clone generated. The insert in pPT1 was quite large (approximately 10 kb) and it might be expected that this recombinant would only occur at low frequency. Alternatively, the other 6 candidate positives identified by hybridisation could have contained the subtilisin gene initially but it was deleted rapidly during subsequent propagation and rescreening. The isolation of pPT1 may have been very fortunate.

3.5.4 Host strains for subtilisin expression

The alkaline and neutral protease deficient strain *B. subtilis* DB104 was generously provided by Dr R. Doi. This strain replaced *B. subtilis* IG20 as the host for the expression of wild-type and mutant subtilisins. Figure 3.5 shows *B. subtilis* IG20 harbouring pUB1662, it is clear that this strain produces low levels of endogenous subtilisin activity. DB104 produces no detectable secreted proteases and therefore constitutes the ideal host for expressing modified subtilisins since it minimises the potential problem of endogenous enzyme interfering with the analysis of mutants.

B. subtilis DB104 is *rec*E$^+$ and is transformable by multimeric plasmid DNA. The strain was constructed by gene conversion of *B. subtilis* DB100, a neutral protease deficient strain, using a 178 bp *in vitro* derived deletion in a promoter proximal segment of the *B. subtilis* alkaline protease gene (Kowamura and Doi, 1984). The strain carries two uncharacterised lesions

in the neutral protease gene. The derivative DB104 produces only 2-4% of the wild-type level of total protease. The residual activity most likely results from intacellular proteases released on lysis or possibly from a secreted esterase, bacillopeptidase F. DB104 should be an acceptable host for the expression of a variety of secreted gene products since they will not be subject to extensive proteolysis once released into the culture medium.

3.5.5 Other subtilisin clones

The *B. amyloliquefaciens* subtilisin gene has been cloned and expressed by at least two other groups. Wells *et al.*, (1983) isolated the subtilisin BPN' gene from a library constructed in *E. coli* using the shuttle vector pBS42 (4.4 kb). A pair of minimally redundant oligonucleotides, which covered all the possible nucleotide ambiguities between positions 117-121 in the amino acid sequence of the mature enzyme, were used to identify subtilisin specific recombinants. One of the recombinants, pS4, harboured a 4.4 kb insert and on transfer to a protease deficient *B.subtilis* host, directed the expression and secretion of subtilisin BPN'. Curiously, pS4 appeared to be stable in both *E. coli* MM294 and *B. subtilis* strains I-168 and BG86 and structural instability has not been observed (J. Wells, personal communication). The extent of the insert in pS4 is shown in Figure 3.3 in comparison with that contained in pPT1. The insert in pPT1 extends about 6 kb further 3' than the insert in pS4. It is possible that sequences responsible for the transient instability of pPT1 might be contained in this region. The total size of pS4 is 8.8 kb, compared with pPT20 (12 kb) and pPT30 (7.5 kb), this may account for the stability of pS4 during long term propagation in *B. subtilis.*

Wells *et al.*, (1983) have also determined the complete nucleotide sequence of

the subtilisin BPN' structural gene and identified the probable transcription initiation and termination signals. The time course of expression of the cloned subtilisin remains under sporulation control in spite of the increased gene dosage. The features of the subtilisin gene will be discussed in detail later.

Vasantha *et al.*, (1984) have cloned and expressed both the alkaline (subtilisin BPN')and neutral protease genes from *B. amyloliquefaciens*. The genes were cloned directly in *B. subtilis* using the marker rescue cloning system developed by Gryczan *et al.*, (1980). Alkaline and neutral protease specific recombinants were identified directly by the functional activity of the cloned genes, the secreted proteases producing zones of proteolysis around colonies grown on milk agar plates. This stategy avoids the use of *E. coli* as an intermediate host and reduces the associated risks of structural DNA instability. The nucleotide sequence of both genes was determined, that of the alkaline protease gene is identical to the subtilisin gene determined in this study and by Wells *et al.*, 1983).

3.5.6 Strategy for mutagenesis

The subtilisin BPN' gene has been cloned and is stably maintained in both pUB110, where it directs the expression and secretion of active subtilisin in *B. subtilis* DB104, and also in M13mp9. It should therefore be possible to mutate the gene harboured in M13 by oligonucleotide-directed mutagenesis. Mutant genes so produced should be similarly expressed in DB104 on transfer to pUB110.

Thus, the strategy for mutagenesis and expression of the mutant subtilisins is:

construct mutant genes in M13 using mismatched oligonucleotides;

confirm the mutation by sequencing the entire gene using a
series of oligonucleotide primers;

purify the fragment harbouring the mutant gene from M13 RF DNA;

ligate the fragment into pUB110 and transform competent
B. subtilis DB104.

Colonies secreting active mutant subtilisins should be identified from the zones of proteolysis. After several rounds of colony purification, the mutant enzymes will be purified from culture supernatants and subjected to kinetic analysis to determine the effects of the mutations on the properties of the enzyme.

CHAPTER 4: DNA SEQUENCING AND MUTAGENESIS

4.1 DNA sequence determination

4.1.1 Introduction

The development of techniques to determine the nucleotide sequence of any segment of cloned DNA rapidly has revolutionised molecular biology. Two DNA sequencing procedures have been developed, a chemical method (Maxam and Gilbert, 1977) and an enzymatic method (Sanger *et al.*, 1977). Whilst both procedures are still widely used, the latter has become accepted as the method of choice for determining long DNA sequences (>500 bp). This is due to the ease and speed of data acquisition and that a detailed restriction map need not be constructed before sequencing commences.

4.1.2 Dideoxy chain terminator sequencing

The enzymatic procedure involves the random incorporation of deoxynucleotide analogues into the growing DNA strand and results in specific termination of DNA synthesis at that position. The analogues are 2',3'-dideoxynucleotide triphosphates (ddNTPs), which lack an hydroxyl group at the 3' position of the ribose ring. They can be incorporated into the growing DNA strand by DNA polymerase, but then prevent further chain extension due to the lack of the 3' hydroxyl for formation of the next phosphodiester bond. DNA synthesis reactions primed from the same position each incorporating one of the ddNTPs and [α^{32}P] dATP, generates a set of radioactive polynucleotides of varying length in the 3'-direction terminating at a dideoxy insertion, and sharing the same 5'-end. The

reaction products are resolved in adjacent lanes on denaturing polyacrylamide gels (Sanger and Coulson, 1978). The gel is then dried and autoradiographed. The DNA sequence is deduced from the relative positions of the bands on the autoradiogram.

4.1.3 M13 vectors in DNA sequencing

The parallel development of bacteriophage M13 as a cloning vector permits the selection of recombinants and the simple production of single stranded template DNA required for sequencing (Gronenborn and Messing, 1978; Sanger et al., 1980). Furthermore, M13 can be used as a vector for shotgun cloning of short (200-400 bp), randomly generated fragments from a larger segment of DNA. The sequence data derived from the individual random clones is aligned using computer programs to generate the entire (contiguous) sequence of the the original fragment. The DNA sequences of many genes, operons and viruses have been determined to date.

4.1.4 Predicting the primary structure of proteins

Rapid DNA sequencing frequently makes it faster to determine the primary structure of a protein by cloning and sequencing its gene then deducing the the encoded order of amino acids, rather than by conventional protein sequencing methods. DNA sequencing is also less prone to the intrinsic errors to which direct protein sequencing is susceptible, for example the confusion between acid or amide groups (i.e. Glu/Gln and Asp/Asn).

4.2 Sequencing the subtilisin BPN' gene

The dideoxy chain terminator procedure of Sanger *et al.*, (1977) was used to determine the nucleotide sequence of the subtilisin BPN' gene. Since Wells *et al.*, (1983) and later Vasantha *et al.*, (1984) had already determined the sequence of both strands of the subtilisin gene, it was only necessary to demonstrate that the sequence of the gene cloned here was the same as those already published. Sequencing one strand of the gene is adequate for this purpose. A series of oligonucleotide primers was synthesised such that they primed DNA synthesis at intervals of 150-200 bases along the non-coding strand of the subtilisin gene cloned in M13mp9 (M13*apr*WT). This allowed the simple and rapid sequencing of the entire wild-type subtilisin gene and subsequent mutants without sub-cloning, as described for tyrosyl tRNA synthetase (Wilkinson *et al.*, 1984).

The sequence of each primer and its priming position is shown in Table 4.1. The amount of each oligonucleotide required for specific priming varied between 0.1-1.0 pmol. The optimal amount was determined experimentally by priming DNA sequencing reactions with a range of primer concentrations from 0.1-2 pmol using 0.5 pmol of template DNA. The products of sequencing reactions were resolved on 6% polyacrylamide gels containing 8 M urea and 1xTBE (Sanger and Coulson, 1978).

The DNA sequence of the coding region of the subtilisin BPN' gene cloned in this study is identical to that of the two previously isolated genes (Wells *et al.*, 1983; Vasantha *et al.*, 1984). Figure 4.1 shows the nucleotide sequence of the non-coding strand of the subtilisin BPN' gene harboured in M13*apr*WT. Only one difference between this and the previously published

Table 4.1:   Sequencing primers.

Oligonucleotide designation	Sequence 5' → 3'	Priming position[1]
SUBPRI.1	AAGATTAACGCTAAAGC	485
SUBPRI.2	TTTTCAGAAATGACATC	620
SUBPRI.3	CGTACGCATGTGCTACG	754
SUBPRI.4	AGCCGCAACTGTGCCGG	975
L2	TCGCCCACTCGATTCC	1097
SUBPRI.5	CCGGCTGCCGCAACGAC	1214
L1	CAGGTCCTACGCTTGA	1337
L (Probe Y)	GATCAGCCCTTTTCC	1559
SUBPRI.6	GACCGGAAACCGGGAAG	1769

Notes to Table 4.1:
Sequencing primers SUBPRI.1-6 were synthesised on the ABI 380 B.
Primers L, L1 and L2 were synthesised manually by J. Brannigan.
(1) Priming position refers to the position of the nucleotide in the non-coding strand of the subtilisin gene which is complementary to the 3'-end of the sequencing primer. The numbering is as shown in Figure 4.1.

```
                              30                                      60
TTT CCG CAA TTA TAT CAT TGA CAA TAT CAA CAT CAA TGA TAT TCA TTA TCA TTA TTT TTA
                              90                                     120
TAA AAT GGT TTC ACA GCT TTT CTC GGT CAA GAA AGC CAA AGA CTG ATT TCG CTT ACG TTT
                             150                                     180
CCA TCA GTC TTC TGT ATT CAA CAA AAG ATG ACA TTT ATC CTG TTT TTG GAA CAA CCC CCA
                             210                                     240
AAA ATG GAA ACA AAC CGT TCG ACC CAG GAA ACA AGC GAG TGA TTG CTC CTG TGT ACA TTT
                             270                                     300
ACT CAT GTC CAT CCA TCG GTT TTT TCC ATT AAA ATT TAA ATA TTT CGA GTT CCT ACG AAA
                             330            ·                 -35   360
CGA AAG AGA GAT GAT ATA CCT AAA TAG AAA TAA AAC AAT CTG AAA AAA ATT GGG TCT ACT
             -10            390                                     420
AAA ATA TTA TTC CAT ACT ATA CAA TTA ATA CAC AGA ATA ATC TGT CTA TTG GTT ATT CTG
                             450                                     480
CAA ATG AAA AAA AGG AGA GGA TAA AGA GTG AGA GGC AAA AAA GTA TGG ATC AGT TTG CTG
             Shine/Dalgarno      fMet Arg Gly Lys Lys Val Trp Ile Ser Leu Leu
                             510                                     540
TTT GCT TTA GCG TTA ATC TTT ACG ATG GCG TTC GGC AGC ACA TCC TCT GCC CAG GCG GCA
Phe Ala Leu Ala Leu Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gln Ala Ala
                             570                                 600 ►PRO
GGG AAA TCA AAC GGG GAA AAG AAA TAT ATT GTC GGG TTT AAA CAG ACA ATG AGC ACG ATG
Gly Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys Gln Thr Met Ser Thr Met
                             630                                     660
AGC GCC GCT AAG AAG AAA GAT GTC ATT TCT GAA AAA GGC GGG AAA GTG CAA AAG CAA TTC
Ser Ala Ala Lys Lys Lys Asp Val Ile Ser Glu Lys Gly Gly Lys Val Gln Lys Gln Phe
                             690                                     720
AAA TAT GTA GAC GCA GCT TCA GCT ACA TTA AAC GAA AAA GCT GTA AAA GAA TTG AAA AAA
Lys Tyr Val Asp Ala Ala Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys Lys
                             750                                     780
GAC CCG AGC GTC GCT TAC GTT GAA GAA GAT CAC GTA GCA CAT GCG TAC GCG CAG TCC GTG
Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val Ala His Ala Tyr Ala Gln Ser Val
                             810                               MAT   840
CCT TAC GGC GTA TCA CAA ATT AAA GCC CCT GCT CTG CAC TCT CAA GGC TAC ACT GGA TCA
Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu His Ser Gln Gly Tyr Thr Gly Ser
                             870                                     900
AAT GTT AAA GTA GCG GTT ATC GAC AGC GGT ATC GAT TCT TCT CAT CCT GAT TTA AAG GTA
Asn Val Lys Val Ala Val Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val
                         32      930   36                           960
                             Pro Asn              Asp
GCA GGC GGA GCC AGC ATG GTT CCT TCT GAA ACA AAT CCT TTC CAA GAC AAC AAC TCT CAC
Ala Gly Gly Ala Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn Asn Ser His
                                                              62      64
```

FIGURE 4.1: Nucleotide sequence of the subtilisin BPN' gene

Compiled from the published sequences of the subtilisin BPN' gene from Wells et al., (1983), Vasantha et al., (1984) and this study. The transcription initiation signals (-10 and -35 regions) and Shine/Dalgarno sequence are underlined. The transcription terminator sequence and 5 bp insertion are also indicated (see Figure 4.2). The cleavage sites between the mature enzyme and propeptide and the propeptide and signal peptide are indicated by arrow heads. The active-site residues Asp-32, His-64 and Ser-221 are indicated, also the target residues for mutagenesis; Asp-36, Asn-62, Asp-99 and Lys-213. Amino acids determined by direct protein sequencing which differ from those predicted from the gene sequence are shown.

```
GGA ACT CAC GTT GCC GGC ACA GTT GCG GCT CTT AAT AAC TCA ATC GGT GTA TTA GGC GTT
Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Asn Asn Ser Ile Gly Val Leu Gly Val
```
                              1050   Asp Ala   1080
```
GCG CCA AGC GCA TCA CTT TAC GCT GTA AAA GTT CTC GGT GCT GAC GGT TCC GGC CAA TAC
Ala Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly Ser Gly Gln Tyr
```
                              1110   99   1140
```
AGC TGG ATC ATT AAC GGA ATC GAG TGG GCG ATC GCA AAC AAT ATG GAC GTT ATT AAC ATG
Ser Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met
```
                              1170   1200
```
AGC CTC GGC GGA CCT TCT GGT TCT GCT GCT TTA AAA GCG GCA GTT GAT AAA GCC GTT GCA
Ser Leu Gly Gly Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala
```
                              1230   Ser Thr   1260
```
TCC GGC GTC GTA GTC GTT GCG GCA GCC GGT AAC GAA GGC ACT TCC GGC AGC TCA AGC ACA
Ser Gly Val Val Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser Thr
```
                              1290   1320
```
GTG GGC TAC CCT GGT AAA TAC CCT TCT GTC ATT GCA GTA GGC GCT GTT GAC AGC AGC AAC
Val Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly Ala Val Asp Ser Ser Asn
```
                              1350   1380
```
CAA AGA GCA TCT TTC TCA AGC GTA GGA CCT GAG CTT GAT GTC ATG GCA CCT GGC GTA TCT
Gln Arg Ala Ser Phe Ser Ser Val Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser
```
                              1410   1440
```
ATC CAA AGC ACG CTT CCT GGA AAC AAA TAC GGG GCG TAC AAC GGT ACG TCA ATG GCA TCT
Ile Gln Ser Thr Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser
```
                              213 1470   221   1500
```
CCG CAC GTT GCC GGA GCG GCT GCT TTG ATT CTT TCT AAG CAC CCG AAC TGG ACA AAC ACT
Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn Trp Thr Asn Thr
```
                              1530   1560
```
CAA GTC CGC AGC AGT TTA GAA AAC ACC ACT ACA AAA CTT GGT GAT TCT TTC TAC TAT GGA
Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Thr Lys Leu Gly Asp Ser Phe Tyr Tyr Gly
```
                              1590   INSERT   1620
```
AAA GGG CTG ATC AAC GTA GAG GCG GCA GCT CAG TAA AAC ATA AAA AAC GGC GTC GGC CTT
Lys Gly Leu Ile Asn Val Glu Ala Ala Ala Gln End        TERM
```
                              1650   1680
```
GGC CCC GCC GGT TTT TTA TTA TTT TTC TTC CTC CGC ATG TTC AAT CCG CTC CAT AAT CGA
```
                              1710   1740
```
CGG ATG GCT CCC TCT GAA AAT TTT AAC GAG AAA CGG CGG GTT GAC CCG GCT CAG TCC CGT
```
                              1770   1800
```
AAC GGC CAA GTC CTG AAA CGT CTC AAT CGC CGC TTC CCG GTT TCC GGT CAG CTC AAT GCC
```
                              1830   1860
```
GTA ACG GTC GGC GGC GTT TTC CTG ATA CCG GGA GAC GGC ATT CGT AAT CGG ATC AGA AGC
```
                              1890   1920
```
AAA ACT GAG CAC GGA TAT AAG CAG CAA TAA CAG CGG GAG AGC GGC CAG ATC TTC AGG CCC
```
                              1950   1980
```
CTT TAT ATG AAG CAT TCG GCC AAA GGA TTT CCA CAT CCG CCG CCT GTA AAG CTG GAA TTC
```

sequences was found.   A 5 bp insertion (indicated in Figure 4.2) was identified, outside the coding region of the gene, in the inverted repeat of the proposed transcription terminator (Wells *et al.*, 1983). In spite of the inserted bases being in the potential stem/loop structure of the proposed terminator, they seem unlikely to significantly affect the function of the terminator since alternative stem/loop structures could still be formed. Figure 4.2 shows the predicted secondary structures formed in the transcript prior to termination using the data from Wells *et* al (1983) and from this study.   The stability of the alternative secondary structures was calculated using the "Tinoco" rules (Tinoco *et al.*, 1973).   The inserted bases destabilise the secondary structure of the RNA by about 5 kcal mol^{-1} (at 25 ˚C). The alternative secondary structures have stabilities of -19.2 and -20 kcal mol^{-1} compared with -25 kcal mol^{-1} for the original terminator structure.

4.3 Features of the Subtilisin BPN' gene

Inspection of the nucleotide sequence   has revealed several important features of the subtilisin gene and given clues as to the nature of the primary translation product.   In addition to the sequence encoding the 275 amino acids of mature subtilisin, the open reading frame extends a further 108 codons (324 bp) upstream from the known amino terminus of the mature enzyme.   Gram positive organisms can use AUG, UUG or GUG to initiate translation (reviewed by Gold *et al.*, 1981).   There are seven potential in-frame translation initiation codons in the extended open reading frame.   AUG appears at positions —88, —60, and —57, GUG at —107 and —41, and UUG at —98 and —19.   Only the most distal of these, GUG —107, is immediately preceded by a plausible ribosome binding site

DNA sequence

Inserted sequence

GGCGT

AACATAAAAAACCGGCCTTGGCCCCGCCGGTTTTTTATATTTTT

Inverted repeats

1   Δ G = -25 kcal/mol

A U A A A A A A C C G G C
U A U U UU UU G G C C G

9 Base loop

2   Δ G = -19.2 kcal/mol

A U A A A A A A C G G C G
U A U U UU UU G C C G C
                    G

Bulge

13 Base loop

3

G   G
G   T   Bulge
G   T

A U A A A A A A C C G G C
U A U U UU UU G G C C G

9 Base loop

Δ G = -20 kcal/mol
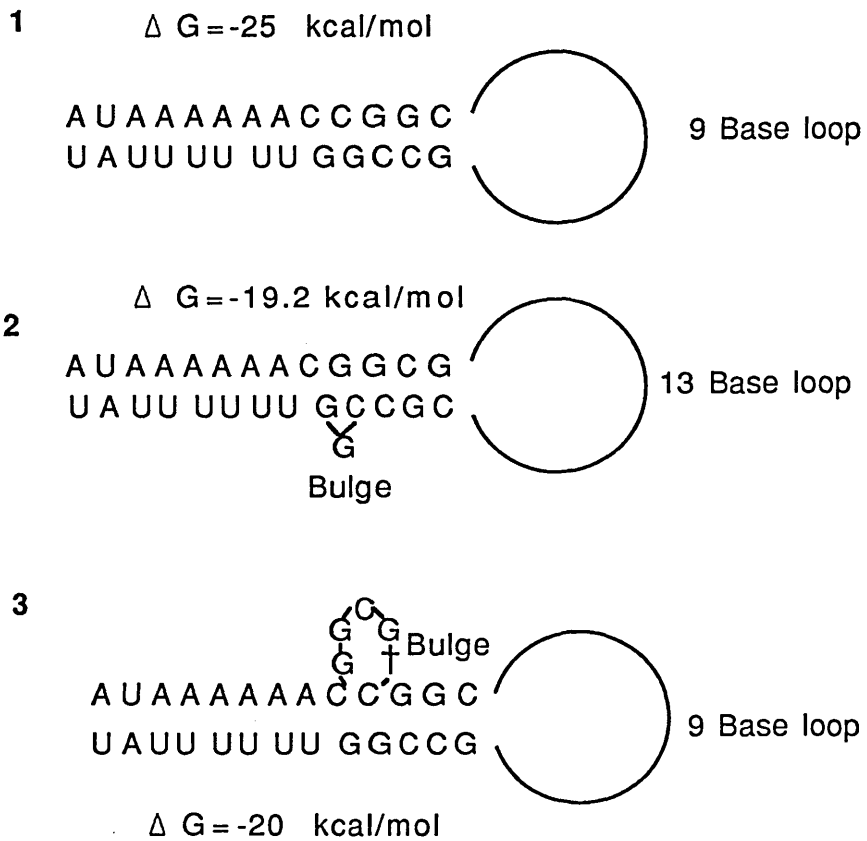
Figure 4.2: Stability of predicted terminator structures
1: From Wells et al., 1983
2,3: From this study

(underlined in Figure 4.1). Furthermore, this codon occurs at the start of a sequence which closely resembles a typical secretion signal peptide (Watson, 1984). There are 3 charged residues near the N terminus followed by 24 uncharged residues and then several possible signal sequence cleavage sites (see later). This left a region of some 75-77 amino acids of unknown function separating the signal peptide and the mature enzyme. Whilst the existence of a signal peptide was anticipated in view of the fact that subtilisin is a secreted enzyme, the presence of the leader peptide separating the signal from the mature enzyme was surprising. The peptide encoded by this region has a net positive charge, with the charged residues clustered, and therefore bears a resemblence to the pro-peptides of the pancreatic serine proteases. It is also similar in length to propeptide of streptococcal proteinase (80 amino acids). Given these similarities it has been suggested that this region functions as a propeptide, maintaining the enzyme in an inactive form until released from the cell (Wells *et al.*, 1983).

Power *et al.*, (1986) have used antibodies to subtilisin to perform immunoblots and radio-immunoassay studies which demonstrate that a 42 kd protein (the predicted size of the subtilisin precursor) accumulates in the cell membrane during vegetative growth and that it dissappears at the onset of sporulation with the concomitant appearance of a 27 kd protein free in the culture medium. The charged leader peptide may have an additional or alternative role in binding prosubtilisin to the negatively charged membrane until the end of log phase growth.

The presumed ribosome binding site or Shine-Dalgarno sequence (Shine and Dalgarno, 1975) identified by Wells *et al.*, (1983) has the sequence GAGAGG, which is similar to that utilised by the *B. licheniformis* α-amylase gene

(Stephens *et al.*, 1984) and that of the *B. subtilis* gluconate operon (Fujita and Fujita, 1986). However the homology to the 3' end of the 16S rRNA can be extended beyond the region indicated by both Wells *et al.*, (1983) and Vasantha *et al.*, (1984). The sequence AAAG<u>GAGAGG</u>AU is perfectly complementary to the 3'-end of the *B. subtilis* 16S rRNA in 10 out of 12 positions. This is consistent with the requirement for stringent homology between the Shine-Dalgarno sequence and the 16S rRNA for efficient initiation of translation in *Bacilli* (Band and Henner, 1984). In the case above, the calculated free energy for the extended interaction is -16 kcal mol^{-1} (using the rules of Tinoco *et al.*, 1973) compared with an average of -17 kcal mol^{-1} for other known *Bacillus* genes (Murray and Rabinowitz, 1982). This compares with an average value of -11 kcal mol^{-1} for *E. coli* ribosome-binding sites.

There are six discrepancies between the deduced amino acid sequence of subtilisin BPN' and that determined by direct protein sequencing. The order of adjacent amino acids has been inverted in four positions and there are two positions where single residues have been assigned as the incorrect acid or amide. The position of these differences is shown in Figure 4.1 It seems most likely that the differences at these positions result from protein sequencing errors. However, it is possible that these differences arise as a result of slight variations between the strains used as sources of DNA for cloning and protein for sequencing.

No promoter sequence was identified which could be transcribed by the major species of RNA polymerase in *Bacillus* species, which utilises the σ^{43} (formerly σ^{55}, Gitt *et al.*, 1985) transcription factor. However Wells *et al.*, (1983) have identified putative promoter elements which have homology with

114

the spoVC −10 and −35 regions (indicated in Figure 4.1). They suggest that like spoVC, the subtilisin gene is recognised and transcribed by the minor RNA polymerase which utilises the σ^{37} transcription factor. Furthermore, Wong *et al.*, (1984) have demonstrated that the highly homologous *B. subtilis* subtilisin gene has tandem promoters at least one of which is transcribed by the σ^{37} holoenzyme *in vivo*. It is probable that the subtilisin gene from *B. amyloliquefaciens* is similarly transcribed by the σ^{37} RNA polymerase holoenzyme, however it is not yet clear whether tandem promoters exist in this case.

Wong and Doi (1984) have demonstrated that the promoter from the *B. subtilis* subtilisin gene can function in *E. coli*, albeit at a very much reduced efficiency. This suggests either, that the *E. coli* RNA polymerase is able to recognise the σ^{37} promoter itself, or that there exist in *E. coli* hitherto unrecognised minor RNA polymerase holoenzymes which can recognise this promoter. This observation does not satisfactorily account for the structural instability of the subtilisin gene cloned in *E. coli* since stable constructs retain an intact promoter.

The expression of the cloned subtilisin gene in *B. subtilis* remains under sporulation control since protease activity only accumulates in the culture supernatant during stationary phase growth (A. J. Russell, personal communication and Wells *et al.*, 1983).

Studies have been undertaken to identify the precise cleavage site between the signal peptide and the leader (pro) peptide in subtilisin and to identify the proposed intermediates in the maturation process. Power *et al.*, (1986) showed that a mutant subtilisin, which had the catalytically critical residue

Asp32 changed to an Asn, was blocked for the secretion of the inactive enzyme. Similarly, deletion of the 3' portion of the gene encoding the carboxy terminal region of the enzyme, resulted in the failure of the truncated protein to be released from the cells. Radio-immuno assays of the sub-cellular fractions showed that in both cases the products of the mutant genes remained associated with the cell membrane in protease deficient hosts. However, in hosts which produced an endogenous subtilisin (which does not cross react with the subtilisin BPN' anti-serum), the inactive products were released from the membrane into the culture medium. These results were used to argue that the maturation and release of subtilisin from the membrane requires an autocatalytic process rather than proteolysis by a signal peptidase.

In vitro studies performed by Wong and Doi (1986) using coupled transcription-translation of the subtilisin gene under the control of the λ P_L promoter, demonstrated that the primary translation product is a 42 kd protein, this being the predicted size for pre-pro-subtilisin (Stahl and Ferrari, 1984). Addition of *B. subtilis* membrane vesicles and Triton X-100 (0.1%) resulted in the processing of the 42 kd product to 38 kd consistent with the formation of prosubtilisin. The 38 kd product was not processed further to the 28kd mature enzyme, a surprising result in view of the reported autocatalytic maturation (Power *et al.*, 1986). Perhaps catalytic amounts of active protease are required to initiate activation which subsequently proceeds autocatalytically.

The signal peptide is capable of directing the secretion of of *S. aureus* protein A (Vasantha and Thompson, 1986) and the TEM β-lactamase from *B. subtilis* (Wong and Doi, 1986) and is not dependent on subtilisin activity for

the release of the secreted product into the growth medium. Amino terminal sequencing of the secreted proteins and the 38 kd product described above, has shown that the signal peptidase cleavage site is at the sequence Ala-Gln-Ala-Ala where cleavage occurs between the two Ala residues at positions −78 and −77. This gives a signal peptide of 29 amino acid residues and therefore the pro peptide consists of 77 amino acids.

4.4 Mutagenesis as a tool in genetics.

4.4.1 Introduction.

"The *ignis fatuus* of genetics has been the specific

mutagen, the reagent that would penetrate to a given

gene, recognise it and modify it in a specific way."

Joshua Lederberg (1960).

Classical genetic analysis depends on identifying naturally occuring mutants or random mutagenesis (*e.g.* UV or chemical) of a population of organisms followed by screening or selection to identify mutants with altered phenotypes. The location of the mutation is subsequently identified by complementation and physical techniques. The major disadvantage of this approach is that "informative" mutations are produced at only very low frequencies and only those for which a suitable screen or selection can be developed are accessible for study. Those which are phenotypically silent are not detected.

The ability to isolate specific genes and segments of DNA by molecular

117

cloning techniques has made the goal identified by Lederberg a practical reality.

Genetic analysis has become "reversed" in that mutations in a specific gene can now be constructed *in vitro* with increased efficiency, the mutation confirmed by DNA sequencing and the mutated gene returned to an appropriate host, where the phenotypic effect of the mutation can be determined and correlated with the precise nature of the mutation at the molecular level.

Two strategies for genetic analysis are now possible, essentially random mutations can be targetted to the gene of interest (or a segment of it) allowing large numbers of potentially informative mutants to be isolated and characterised. This is especially useful when it is not possible to predict which mutation will result in an altered phenotype, and is termed targetted or directed mutagenesis. Alternatively, a predetermined mutation can be constructed at a specific location in the cloned gene. This permits, for example, the role of individual bases in regulatory DNA elements to be investigated, also specific codons can be altered such that specific amino acid substitutions can be made in a protein molecule. This is termed site-specific mutagenesis. Both strategies have the advantage over classical mutagenesis of producing mutants with relatively high efficiency and avoid biological selection (except where it remains useful) as a means of identifying mutants using as an alternative DNA sequencing or hybridisation screening.

Thus, the genetic analysis of an organism can now be performed at three levels:

the whole genome by classical methods; (random)

the whole gene by targetted or directed mutagenesis; (semi-random)

individual nucleotides by site-specific mutagenesis; (non-random)

The techniques of directed mutagenesis permit base substitutions, insertions and deletions. Directed mutagenesis can be achieved by a variety of methods including: bisulphite mutagenesis of single-stranded regions of DNA; nucleotide misincorporation and incorporation of nucleotide analogues. These procedures have been excellently reviewed by Shortle *et al.*, (1981) and are not considered in further detail here.

4.4.2 Site-specific mutagenesis

Site-specific mutagenesis is made possible by the availability of synthetic oligonucleotides. These provide the means to alter a defined site within a region of clone DNA in a predetermined fashion. Interestingly, Lederberg actually suggested that a polynucleotide might be able to perform this function. There is now an extensive body of literature describing developments in the technology of site-specific mutagenesis. Several excellent reviews have already been published including articles by Zoller and Smith (1983), Itakura *et al.*, (1984) and most recently Smith (1985) and

Carter (1986).

In addition to simple and rapid oligonucleotide synthesis, two further lines of investigation lead to the eventual development of oligonucleotide-directed mutagenesis. Firstly, it was demonstrated that point mutations in the single-stranded bacteriophage ΦX174 were efficiently corrected in marker rescue experiments (Weisbeek and van de Pol, 1970; Hutchison and Edgell, 1971). Restriction fragments from wild-type ΦX174 RF DNA were annealed to the region of the chromosome harbouring the mutation. Transfection of the resulting heteroduplex molecules into a host gave rise to both wild-type and mutant progeny. This demonstrated that both strands of the heteroduplex can give rise to progeny. The frequency of correction of the mutation was dependent on the length of the "rescuing" fragment, with those shorter than about 50 nucleotides the frequency decreased rapidly to undetectable levels.

Secondly, studies on the thermal stability of short complementary oligonucleotide duplexes demonstrated that they remained specifically hybridised at low temperatures even when they contained deliberate mis-matches (reviewed by Smith, 1983). Since oligonucleotides had already been shown to function as specific primers of DNA synthesis on both DNA and RNA templates (Sanger *et al.*, 1977; Smith, 1983), it was a logical step to combine the three systems and incorporate a mis-matched oligonucleotide into a closed circular genome by enzymatic means in order to generate mutants on transformation of the heteroduplex DNA into a suitable host. Thus, a mutagenic oligonucleotide is synthesised such that it is complementary to the target DNA apart from a specific mismatch which can involve one or more nucleotides. The mutagenic primer is then
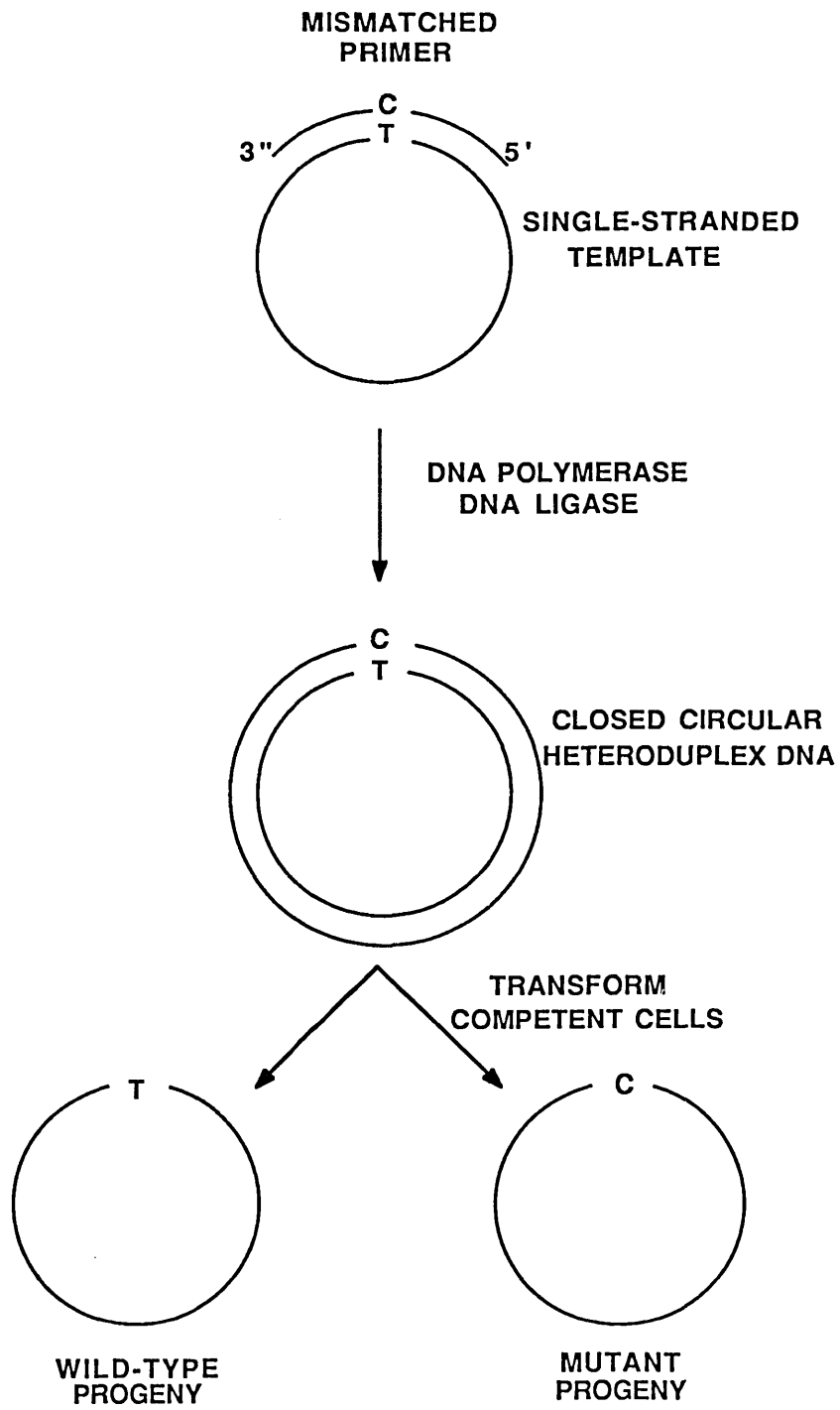
Figure 4.3: Strategy for oligonucleotide-directed mutagenesis

annealed to single stranded target DNA and used to prime DNA synthesis to generate a double stranded heteroduplex product incorporating the mis-matched bases. On introducion into the appropriate host, the parent and mutated strands segregate and mutant progeny are obtained with high efficiency. Figure 4.3 outlines the basic strategy of site-specific mutagenesis. The mutagenic primer itself is used as a radiolabelled probe to discriminate between wild-type and mutant progeny under stringent conditions.

Oligonucleotide primers can be designed to achieve simple transition or transversion mutations (Montell *et al.*, 1982) multiple point mutations (Thomas *et al.*, 1985) and insertions or deletions (Chan and Smith, 1984).

The so called "Wallace" rules (Suggs *et al.*, 1981; Appendix 2) are used to predict the temperature at which the oligonucleotide primer/probe will be able to discriminate between wild-type and mutant DNA sequences under defined conditions. The Td for an oligonucleotide perfectly hybridised to its complementary sequence is defined as the temperature at which 50% of the probe remains specifically bound.

The Td is calculated by summing the contribution of each base: 2 °C for an A or T, 4 °C for a G or C. Probe hybridised to the wild type sequence is mismatched and therefore destabilised to such an extent that it can be washed off at a temperature several degrees below the Td. Probe bound to the mutant sequence remains specifically hybridised at the same temperature and mutants can therefore be identified autoradiographically. The calculated value for the Td of a primer begins to break down for primers greater than about 22 bases long, and it becomes increasingly

122

difficult to discriminate between wild-type and mutant sequences especially in the case of single base mutations.

Vectors based on the single standed bacteriophage M13 are frequently used to provide the template DNA, however oligonucleotide-directed mutagenesis can also be performed directly on plasmid DNA using the "gapped duplex" approach (Morinaga *et al.*, 1984; Kramer *et al.*, 1984b; Hollenberg *et al.*, 1984). This method has the advantage (where *E. coli* is the expression host) of allowing mutations to be constructed directly in the expression vector avoiding the need to sub-clone in and out of M13.

4.4.3 Improving the efficiency of mutatagenesis

The procedures described above typically give efficiencies of mutagenesis of 1-15%. Strategies have been developed in several laboratories which further improve the efficiency of site-directed mutagenesis, in some cases to such a degree that hybridisation screening can be omitted and mutants identified directly by DNA sequencing. The increased efficiency depends on either reducing the frequency of mis-match repair of the heteroduplex DNA by the host or providing a biological selection mechanism for the mutant progeny.

Correction of the mutation back to wild-type can be reduced by the use of mutator strains, which are deficient in the enzyme systems for performing mis-match repair, as a hosts for the heteroduplex DNA (Kramer *et al.*, 1984a). Alternatively, the hosts mis-match-repair machinery can be confused by either growing the template DNA in a *dam⁻* host (Kramer *et al.*, 1982) or by *in vitro* methylation of the heteroduplex DNA using the *dam*

methylase (Horton and Lord, 1986). In both cases the host is unable to distinguish the parent (+) strand from the (-) strand synthesised *in vitro* (*in vivo* DNA replication results in hemi-methylated daughter stands) and therefore repairs the mutation at much lower frequency.

"Coupled priming" employs a second mutagenic oligonucleotide to remove a selectable marker elsewhere in the template strand. Carter *et al.*, (1985) constructed a series of M13 vectors which harboured an amber mutation in the viral gene 4, which encodes an essential phage function, and the virus is therefore unable to grow in non-suppressor hosts. Template DNA was grown in a su^+ host and during mutagenesis of the cloned tyrosyl-tRNA synthetase gene a second mutagenic primer was employed to correct the amber mutation back to the wild-type sequence. Transformation of the double-heteroduplex DNA into a su^- mis-match repair deficient host gave a significantly enhanced yield of mutants (<70%).

This strategy is very efficient but allows only a single round of mutagenesis since the selectable marker is necessarily lost. The construction of multiple mutations in the same gene would require that the mutant insert be recloned into the amber vector. This strategy has been refined to allow sequential rounds of mutagenesis. Synthetic oligonucleotides containing *Eco* K or *Eco* B restriction sites were inserted on the 3' side of the cloned gene. Use of two selection primers allow the *Eco* K site to be converted to an *Eco* B site or *vice versa*, the progeny being selected in the appropriate r_K^+ or r_B^+ mis-match repair deficient strain. This strategy was similarly efficient (<70%) and allowed consecutive rounds of mutagenesis to be performed in a cyclic fashion as shown in Figure 4.4.
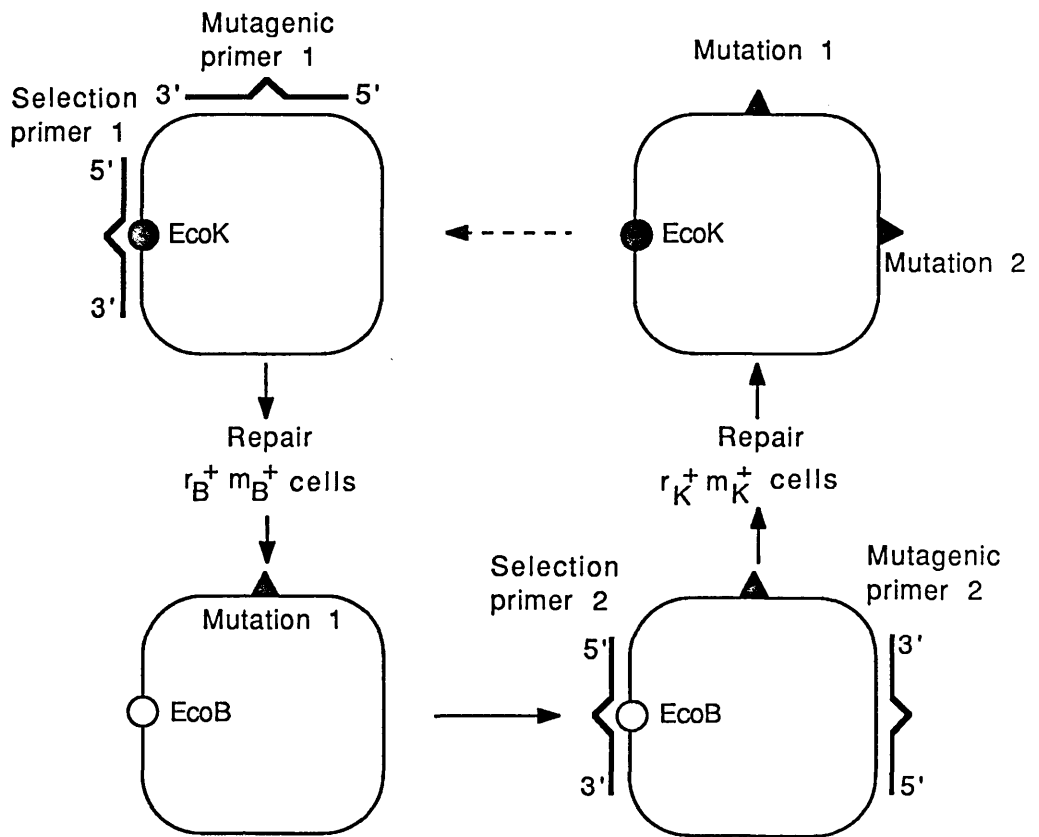
Figure 4.4:

Strategy for high efficiency mutagenesis using reciprocating EcoK/EcoB selectable markers (Carter et al., 1985)

Template DNA isolated from *E. coli dut⁻ ung⁻* strains contains uracil rather than thymidine and is unable to transfect wild-type strains. This can be exploited in site-specific mutagenesis experiments since the *in vitro* synthesised strand harbouring the mutation contains thymidine and therefore is able to transfect wild-type host strains. Selection can therefore be applied against progeny derived from the parent (+) strand (Kunkel, 1985).

Site-specific mutagenesis is clearly an extremely powerful technique which has application in the fine manipulation and functional analysis of DNA segments and also permits the alteration of coding sequences to produce novel proteins. The technique provides an invaluable tool for biochemists and enzymologists for the analysis of sequence/structure/function relationships in proteins. The emergence of "Protein Engineering" bears testimony to the importance of a routine technology which permits specific amino acid substitutions. In this study, site-specific mutagenesis is used to construct mutant forms of the serine protease subtilisin BPN' in an attempt to investigate the role of surface charge on the *p*H dependence of enzyme catalysis.

4.5 Mutagenesis of Asp99→Ser in subtilisin BPN'

4.5.1 Strategy for mutagenesis

The rationale for the construction of this mutant is described elsewhere (Chapter 6). Mutagenesis of the Asp codon at position 99 in the mature enzyme was achieved using an oligonucleotide containing a single mis-match essentially as described by Zoller and Smith (1983). The mutagenic primer was annealed to the template DNA and used to prime DNA synthesis all

Table 4.2:   Mutagenic primers.

Mutation	Sequence $5' \rightarrow 3'$	Length (bases)	T_d
DS99	CGGAACCGC̈T̈AGCACCG	17	58 ˙C
DQ36	TGAGAAGAČTǦGATACCGC	19	58 ˙C
ND62	TGAGAGTČGTTGTCT	15	44 ˙C
KT213	CCCCGTATǦTGTTTCCA	17	52 ˙C

Notes to Table 4.2:
* indicates bases mis-matched to the wild-type sequence.

the way around the template until ligation of the 3'-end of the nascent strand to the 5'-end of the primer generated a closed circular heteroduplex DNA molecule. Heteroduplex DNA was purified away from partial extension products and unprimed WT template DNA by centrifugation through an alkaline sucrose gradient, then transformed into *E. coli* BMH71-18 *mut*L, a mismatch-repair deficient host.

4.5.2 Design of the DS99 mutagenic primer

The codon for Asp99 is GAC and there are six possible Ser codons from which to select a replacement; TCT, TCC, TCA, TCG, AGT and AGC. Of these alternatives, only TCC and AGC could be constructed using a double mis-matched primer, the others would all require a triple mis-match. The codon AGC occurs 12 times elsewhere in the gene, whereas TCC occurs in only 4 positions. To minimise the risk of reduced translation efficiency through utilisation of a "bad" codon, AGC was selected as a suitable replacement. The mutagenic heptadecamer primer (17mer) synthesized to achieve the required mutation is shown in Table 4.2. The mis-matches were located centrally within the primer to ensure specific priming and to prevent strand displacement by the nascent strand during the extension/ligation reaction. DNA sequencing reactions (0.5 pmol of template DNA) primed with 1 pmol of the DS99 primer generated DNA sequence from the appropriate region of the subtilisin gene. This confirmed that priming was specific and that the oligonucleotide was suitable for mutagenesis of the M13*apr*WT template. Oligonucleotide primers for the construction of other mutants were designed with similar considerations.

4.5.3 Annealing primer and template

Previously 5' phosphorylated DS99 primer (2 pmol) was annealed with 2 μl of M13aprWT containing approximately 0.5 pmol of template DNA, in a 10 μl reaction volume. Annealing was done in 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM DTT. The DNAs were placed in a 90 °C water bath which was then allowed to cool to room temperature over 1 hr, during which annealing took place.

4.5.4 Extension/ligation

10 μl of a solution containing 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT , 1 mM each of dCTP, dGTP and TTP, 0.005 mM dATP, 1 mM ATP and 10 μCi of [α^{32}P]dATP, was added to the annealed primer/template mix. 2 units of DNA polymerase (Klenow fragment) was added and the reaction incubated at room temperature for 15 min to pulse label the newly synthesised DNA strand. 1 μl of 10 mM dATP was added as a chase and 2 units of T4 DNA ligase. The extension/ligation reaction was incubated at 15 °C over-night.

4.5.5 Purification of heteroduplex DNA

Unincorporated [α^{32}P]dATP, partial extension products and unprimed template DNA were separated from closed circular heteroduplex DNA by centrifugation through a 5-20% (w/v) alkaline sucrose gradient. The volume of the primer extended material was adjusted to 70 μl with water and 30 μl of 20% (w/v) PEG6000 in 2.5 M NaCl, added. The mixture was placed on ice for 15 min, then centrifuged for 5 min at 4 °C in a

microcentrifuge. The supernatant was discarded and the pellet washed with 100 µl of the PEG/salt diluted as above. After recentrifuging for 5 min, the supernatant was discarded and the DNA dissolved in 180 µl of 1xTE. 20 µl of 2 M NaOH was added and the mixture left at room temperature for 5 min prior to cooling on ice. The alkaline denatured DNA was layered onto the top of a 5-20% alkaline sucrose gradient (Zoller and Smith, 1983). Centrifugation was at 40,000 rpm for 2 hr at 4 °C in a Beckman SW60Ti rotor. The gradient was fractionated by puncturing the tube just above the end (to avoid solid debris) using a 21 g hypodermic needle and allowing the gradient to drain out under gravity. Two drop fractions were collected in 1.5 ml microfuge tubes and the location of the ccDNA determined by liquidscintillationcounting (Čerenkov). The closed circular heteroduplex DNA formed a peak in the lower part of the gradient whilst other DNAs remained near the top of the gradient. Figure 4.5 shows the profile of radioactivity down the fractionated gradient. Fractions 15-20 contained the ccDNA and were pooled and neutralised with 3 M Tris-Acetate pH5. Aliquots of the heteroduplex DNA were transfected into competent *E. coli* BMH71-18 *mutL* cells which were plated onto H-agar plates in H-top agar. Log phase *E. coli* TG2 cells provided the lawn to isolate the progeny phage from the mutator background.

4.5.6 Hybridisation screening

About 100 plaques were picked onto fresh H-plates and grown as infected colonies at 37 °C overnight. The colonies were transferred to nitrocellulose filters and probed with radiolabelled DS99 primer exactly as described (2.14.1). After hybridisation at room temperature overnight, the filters were washed at room temperature in 6xSSC for 3x45 min then exposed to

FIGURE 4.5: Alkaline sucrose gradient profile
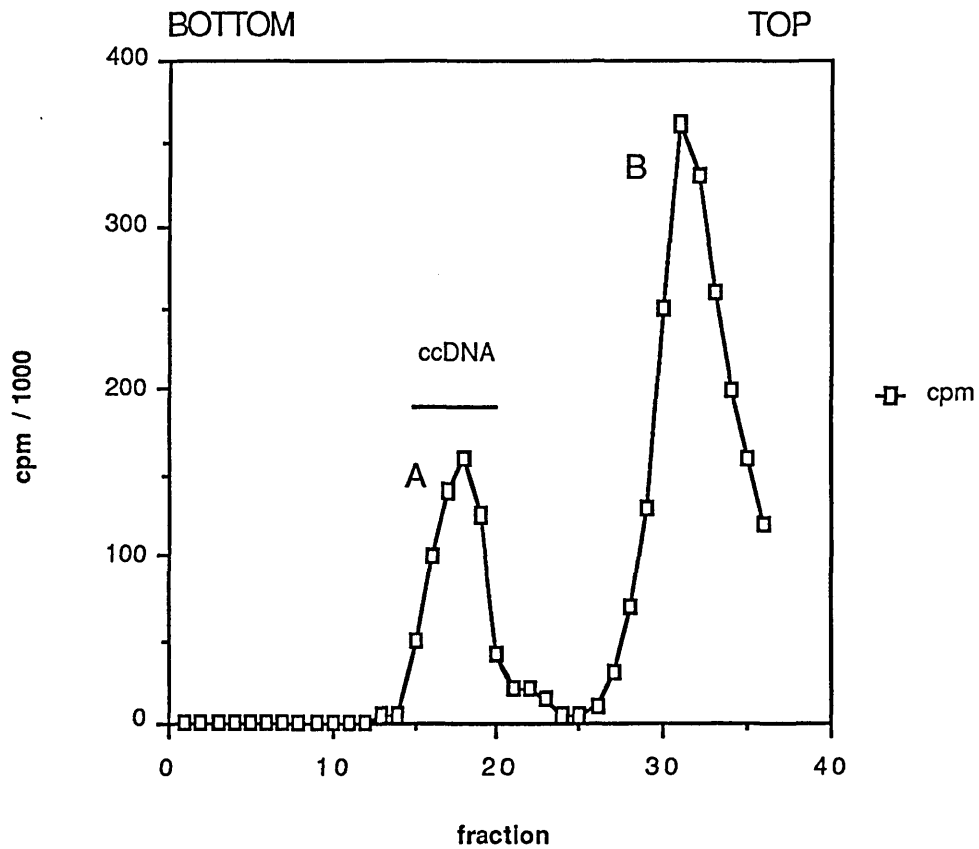
Profile of radioactivity in fractions down an alkaline sucrose gradient. Heteroduplex DNA harbouring the DS99 mutation is contained in peak A. Radiolabelled material from partially double stranded reaction products remains near the top of the gradient (peak B)

The appropriate amounts of vector and target DNA required for ligation were estimated by the intensity of fluorescence of aliquots run on agarose gels. Recovered target DNA (8 μl containing approximately 300-500 ng of fragment) and *Eco*R1 digested pUB110 (400 ng) were ligated at 20 °C for 4 hr under the conditions described (Section 2.10.2). Ligation to form multimeric products was confirmed by gel electrophoresis. Competent *B. subtilis* DB104 cells were transformed with 10 μl of the ligated DNA. Transformants were selected by growth on LB-agar containing kanamycin (25 μg ml^{-1}) and milk solids (1% w/v). Kanamycin resistant transformants were obtained, some 4% of which also exhibited the characteristic clearing zones indicative of subtilisin secretion. The mutation of Asp99→Ser (abbreviated DS99) does not appear to affect the activity of the enzyme against a natural substrate for this protease - casein.

Competent *Bacillus* cells are able to receive more than one plasmid during transformation and DNA taken up may be subjected to rearrangements. Therefore a single, well separated, protease producing colony was subjected to four rounds of sequential single colony purification to ensure that the plasmid species harboured by the protease producing cells was clonally pure. Plasmid DNA was isolated from one of the putative DS99 producing transformants grown on a small scale. Restriction analysis of the DNA confirmed that the correct fragments were present, 3 kb for the insert and 4.5 kb for the vector, demonstrating that no gross rearrangements had occured. The orientation of the fragment was the same as in pPT30 harbouring the wild-type gene. This construct was designated pPT30-DS99. Purification and part of the kinetic analysis of the subtilisin DS99 mutant was done by A.J. Russell.

preflashed X-ray film for 6 h at -70 °C with an intensifying screen. On development the autoradiogram showed that the probe had hybridised to every colony. The filter was rewashed at 55 °C (Td = 58 °C) for 2x5 min then re-exposed for 72 h as before. Figure 4.6 shows that 6 colonies remain strongly hybridised to the DS99 probe when washed under the more stringent conditions.

One of the hybridisation positives was selected for DNA sequence analysis to confirm the mutation. Plaque purified phage from the hybridisation positive were used as a source of template DNA. The DNA sequencing reactions were primed using the sequencing primer L2. The sequence obtained confirmed that the required -GA- to -AG- double mutation had occurred. Template DNA harbouring the confirmed mutation was designated M13*apr*DS99. The entire coding region of the gene was subsequently determined in full using the battery of sequencing primers to ensure that no adventitious mutations had occurred as a result of propagation through the *mut*L host or by unseen secondary priming events during mutagenesis. M13*apr*DS99 contained only the required mutation (A.J. Russell, personal communication).

4.5.7   Expression of subtilisin DS99

M13*apr*DS99 RF DNA was purified from 2 ml cultures of infected *E. coli* TG2 by the alkaline/SDS procedure. The recovered DNA was digested to completion with *Eco*R1 and the products separated by electrophoresis through a 0.8% agarose gel in 1xTAE. The 3 kb fragment harbouring the DS99 mutation was purified using NA45 membrane as described (Section 2.8.2). The recovered DNA was dissolved in 20 μl of water.
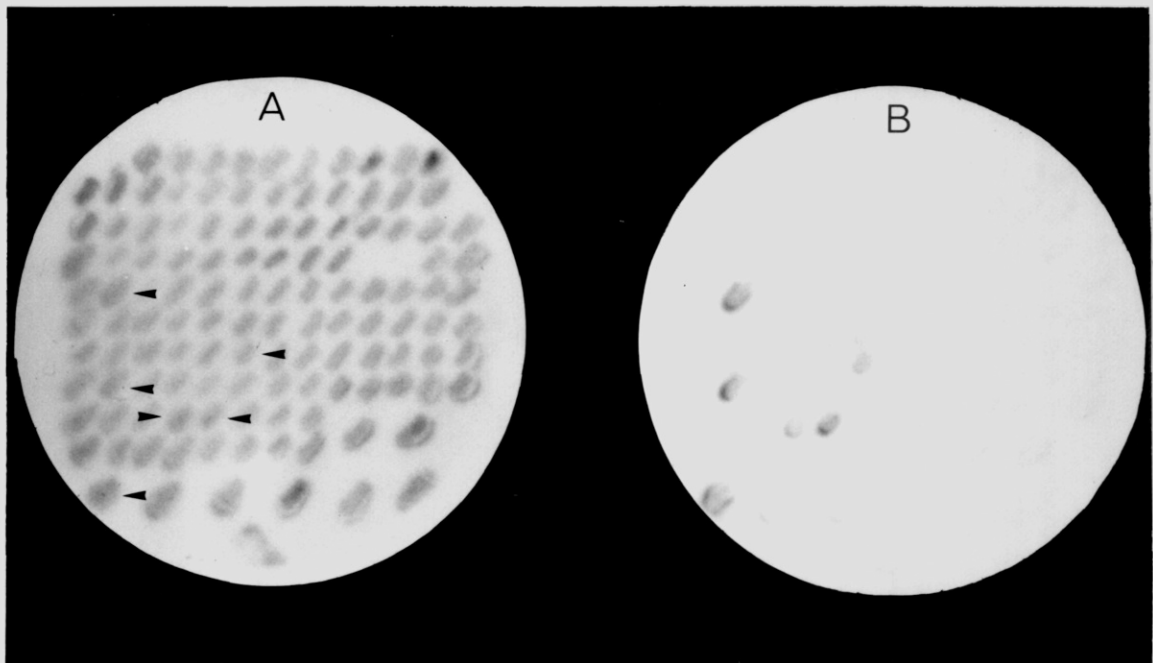
Figure 4.6:   Identification of Asp99→Ser mutants by hybridisation
                    screening.

Colony lysis and DNA immobilisation was as described in section 2.14.1.
The filter was probed with 5' ^{32}P labelled DS99 primer (Table 4.2) at a
concentration of approximately 1 pmol.ml^{-1}.   Hybridisation was at room
temperature overnight.

Panel A: Low stringency wash, 3x45 min at room temperature in 6xSSC.
Autoradiographed overnight at -70 ˙C with an intensifying screen.

Panel B: Stringent wash, 2x5 min at 55 ˙C in 6xSSC.   Autoradiographed
as above.

The six candidate DS99 mutants are indicated by arrow heads in panel A.

4.6 Construction of mutants by double priming

All other mutants in this study were constructed using a double primed procedure developed from several published methods (Norris *et al.*, 1983; Smith *et al.*, 1985; Carter *et al.*, 1985). The strategy for double primed mutagenesis is outlined in Figure 4.7. As an example, the construction of the mutant Asn62→Asp (ND62) is described in detail. Any alterations to this procedure for the construction of other mutants will be described where appropriate. The rationale behind the selection of this mutation and the design of the ND62 mutagenic primer are described elsewhere.

10 pmol each of the ND62 mutagenic primer and the M13 universal primer (both 5' phosphorylated) were annealed with 0.5 pmol of M13*apr*WT template DNA in a 10 μl reaction volume containing 10 mM Tris-HCl pH8.0 and 10 mM MgCl$_2$ (TM). The DNAs were heated at 80 °C for 2 min then allowed to cool slowly to room temperature. An equal volume (10 μl) of a solution containing 500 μM of each dNTP, 500 μM ATP and 10 mM DTT in TM was added to the annealed DNAs. One unit of DNA polymerase (Klenow fragment) and 2 units of T4 DNA ligase were added and the reaction incubated at 15 °C for 18 hr. The primer extended DNAs were diluted to 200 μl with TM and aliquots transfected into competent *E. coli* BMH71-18 *mut*L cells for mutant identification by direct plaque screening (Section 2.14.5).

The ND62 mutation was confirmed by DNA sequencing and the mutant gene expressed in *B. subtilis* DB104 on transfer to pUB110, exactly as described for the DS99 mutation.

Mutagenic primer

3'____/____5'

3'
Second
primer
5'

DNA polymerase (Klenow)
DNA ligase

Closed circular
heteroduplex DNA

Transform competent cells
without purifying ccDNA

Mutant

Wild-type
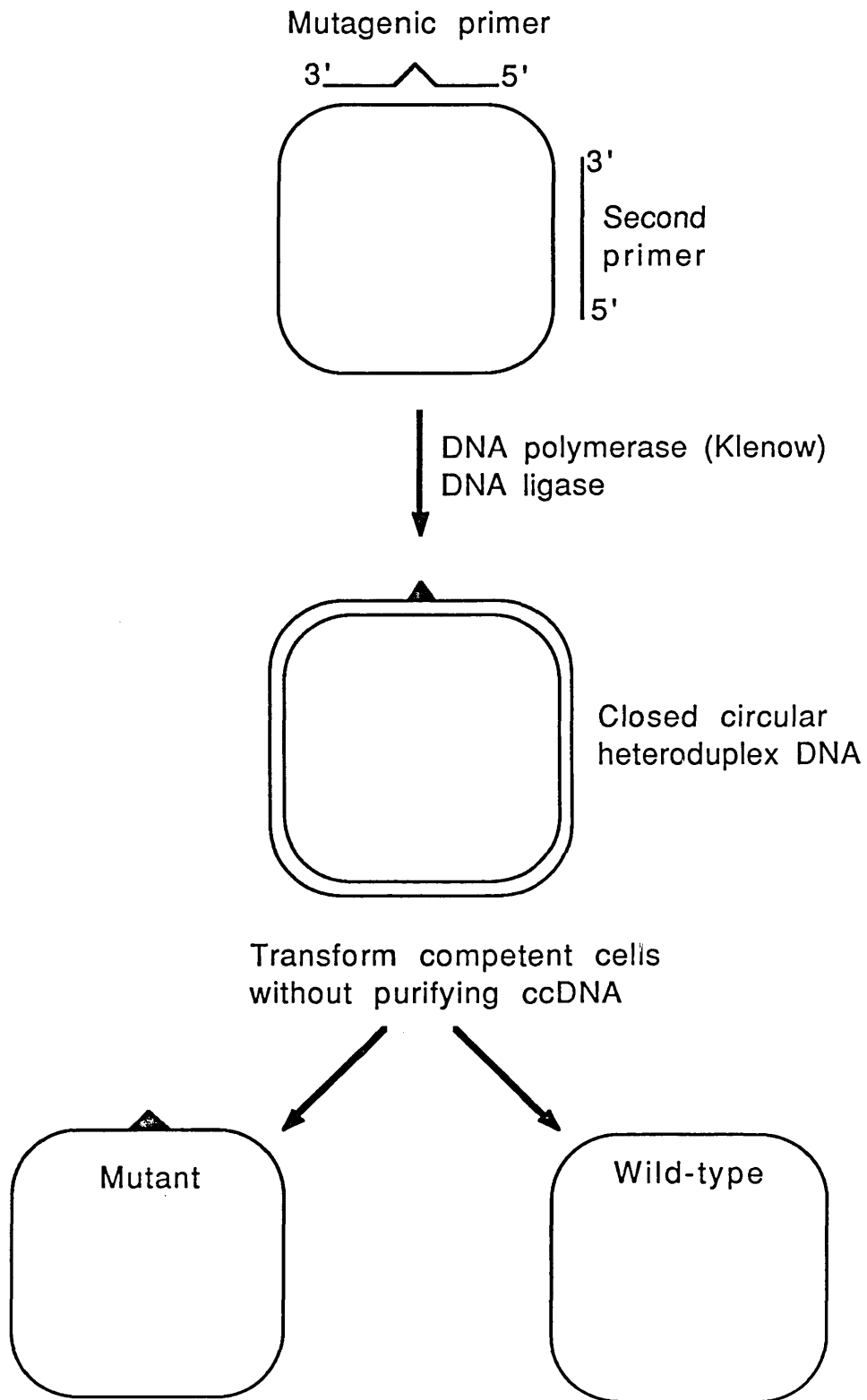
Figure 4.7:

Double primed oligonucleotide directed mutagenesis

4.7 Discussion

A system for the construction and expression of mutant subtilisin genes has been established. Mutations are produced by oligonucleotide directed mutagenesis of the subtilisin BPN' gene cloned in M13mp9. After confirmation of the required mutation by sequencing the entire coding region of the gene, using a battery of primers, the mutant genes are transferred into pUB110 and expressed in the protease deficient host *B. subtilis* DB104.

The single primed mutagenesis procedure was abandoned in favour of the double primed method to eliminate the time consuming step of purifying the heteroduplexes on the alkaline sucrose gradients. Furthermore, double primed mutagenesis frequently results in improved efficiencies of mutagenesis. Identification of the mutants by direct plaque hybridisation gave clean signals which were less susceptible to interference by background hybridisation to cell debris as is common in colony screening. Given the ease of identifying plaques containing the required mutant, even if efficiencies fell well below 1%, it was not necessary to incorporate any of the recent techniques which give improved yields of mutants.

The ratio of DNA polymerase to ligase in the extension/ligation reaction appears to be an important factor in the generation of biologically active DNA. The required event is the ligation of the 3'-end of the growing DNA strand to the 5'-end of the mutagenic primer, thereby sealing the mutation into the DNA. If the amount of ligase is reduced, or polymerase increased, then there is a significant loss of transforming activity and extremely high molecular weight material can be detected by gel

electrophoresis (T. Wells and M. Bycroft, observations). This is presumed to result from strand displacement by the nascent chain and leads to the formation of concatemeric single stranded DNA which is non-transforming. A unit ratio of 2:1 of T4 DNA ligase to DNA polymerase (Klenow) was found to give efficient mutagenesis and did not significantly affect the transforming activity of the DNA.

The design of the mutagenic oligonucleotide itself is an important factor in this situation. If the mismatch is located too near the 3'-end of the primer, then transient disassociation or "breathing" may occur between the primer and the template DNA, allowing strand displacement to occur. Attempts to mutate the chymosin B gene to chymosin A, involving a single nucleotide change, using a 15mer where the mismatch was only 5 bases from the 3'-end of the primer resulted in both a reduction in the transforming activity of the *in vitro* reaction products and the generation of mutants containing repetitions of the target DNA sequence (M. W. Bodmer, personal communication).

CHAPTER 5:   EXPERIMENTAL - ENZYME PURIFICATION AND
              ANALYSIS

5.1 Reagents

The synthetic peptide substrates

succinyl-L-alanyl-L-alanyl-L-prolyl-L-phenylalanyl *para*-nitroanilide,

(suAAPFpN)   and   benzoyl-L-valyl-L-glycyl-L-arginyl   *para*-nitroanilide,

(bzVGRpN) were purchased from Sigma Chemical Co.

N-*trans*-cinnamoyl imidazole (NCI) was purchased from Aldrich Chemical Co.

Ltd.

CM52-cellulose was obtained from Whatman Ltd.

CM-TrisAcryl was from IBF Biotechnics.

All other materials were "AnalaR" grade.

5.2 Enzyme purification

Protease activity was expressed and secreted when the plasmid pPT30, or it's

mutated derivatives, were harboured in the protease deficient host *B. subtilis*

DB104.   Wild-type and mutant subtilisin genes were purified from 36 hr

culture supernatants by a procedure modified from Estell *et al.*, (1985) and

J. A. Wells (personal communication).   Cultures were grown in LB-broth

containing kanamycin (25 μg ml^{-1}) and glucose (0.2%).   Small scale cultures

(1-2.5 l) were grown in flasks (500 ml broth per 2 l flask), incubated at

37 °C in a shaking incubator (275 rpm).   The cells were removed by

centrifugation (8,000 rpm, 4 °C, 30 min) and the pH of the culture

supernatant then adjusted to pH 6.2 at 4 °C using 1 M KH$_2$PO$_4$.   The

volume of the supernatant was reduced to approximately 200 ml either by

ultra-filtration (Amicon PM10 membrane) or ammonium sulphate precipitation

(100% saturation) of soluble proteins.   Ammonium sulphate precipitated

protein was recovered by centrifugation (8,000 rpm, 4 °C, 60 min) and

dissolved in 200 ml of 0.01 M phosphate buffer pH 6.2. The concentrated protein solution was dialysed extensively against 0.01 M phosphate buffer pH 6.2 until the conductivity was comparable to that of the buffer (approximately 1.2 mho). Any precipitated material in the dialysate was removed by centrifugation (8,000 rpm, 4 °C, 30 min) prior to chromatography on a CM-cellulose column (5x5 cm) equilibrated in 0.01 M phosphate buffer pH 6.2. Sample loading, washing and elution were automatically controlled using a Pharmacia FRAC 300 programmable fraction collector and peripherals. The flow rate was maintained at 2 ml min^{-1} throughout. The progress of adsorption, washing and elution was monitored by following the UV absorbance of the eluate at 280 nm.

The column was washed with 0.01 M phosphate buffer pH 6.2 until the eluate contained no UV absorbing material. Bound subtilisin was then eluted with 0.01 M phosphate buffer pH 6.2 containing 0.08 M NaCl. Fractions containing the eluted peak of protease activity were pooled, concentrated about ten fold by ultra-filtration (Amicon PM10 membrane), then frozen in liquid N$_2$ and stored at -70 °C.

The purification procedure was slightly modified for larger scale (10 l) cultures which were grown in a fermenter (MBR). Cultures were grown for 24-36 hr at 37 °C, media and supplements were as for flask cultures. During growth the air flow was maintained at 5 l min^{-1} and the culture stirred at 500 rpm to provide the aerobic growth conditions required by *B. subtilis*.

The cells and culture supernatant were separated by micro-filtration using a 0.45 μm pore size "Pellicon" membrane filter cassette (HVLP000C5, Millipore). The supernatant was concentrated to approximately 500 ml by ultra-filtration using a "Pellicon" membrane filter (PTGC00005, Millipore) with a nominal

10,000 M_r cut-off. The concentrated supernatant was dialysed by buffer exchange: the supernatant was repeatedly diluted with 0.01 M phosphate buffer pH 6.2 then reconcentrated, until the conductivity had been reduced to a minimum (approximately 1.2 mho). Typically ten cycles of dilution/concentration were required, this takes only 2 hr to complete and so is significantly faster than dialysis. Purification of the subtilisin on CM-cellulose was as described above.

The ion exchange resin and buffer system were eventually replaced by CM-TrisAcryl equilibrated and run in 0.01 M TES *p*H 6.2, bound subtilisin being eluted in 0.01 M TES *p*H 6.2 containing 0.1 M NaCl. Flow rates were maintained at 5 ml min^{-1}.

Yields of mutant subtilisins were in the range 10-13 mg l^{-1}, irrespective of the culture volume or the purification procedure used.

5.3 Quantification

The approximate concentration of subtilisin-containing fractions eluted from the ion exchange column was determined from the absorbance of the solution at 280 nm using an extinction coefficient $\epsilon = 1.17$ for a 1 mg ml^{-1} solution (Matsubara *et al.*, 1965).

Accurate concentrations of enzymically active subtilisin were determined by active site titration with N-*trans*-cinnamoyl imidazole (see below).

5.4 Active-site titration

5.4.1 Introduction

In any quantitative study the concentration of active enzyme must be accurately known. This can not usually be obtained from the the absorbance at 280 nm and the molar extinction coefficient, since the enzyme is frequently contaminated with inactive degradation products or other impurities. Concentrations derived from absorbance measurements are therefore frequently overestimated with respect to the concentration of active enzyme. This can result in significant errors where the enzyme concentration $[E_0]$ is used in subsequent calculations, for example the derivation of k_{cat} from V_{max}. The concentration of catalytically functional enzyme can be determined by active-site titration.

The principles of active site titration have been discussed in detail by Kezdy and Kaiser (1970). In general the requirements of an active site titrant are that on its reaction with the enzyme a measurable change should result which can be easily and accurately measured, for example a change in absorbance or fluorescence. The reaction must be specific to the enzyme of interest and must not occur with either contaminating proteins or inactive degradation products of the enzyme. The titrant and enzyme should react in a 1:1 stochoimetry (or other known ratio) and ideally result in the irreversible inactivation of the enzyme (eg by chemical modification). The latter point is frequently unobtainable and titrants which bind reversibly or give slowly hydrolysed products have to be used under conditions which minimise the unwanted reaction. Owing to the relatively low extinction coefficients of many of the active-site titrants available, the technique is inherently insensitive and can only be used if concentrated preparations of the enzyme are available.

5.4.2 Active-site titration of subtilisins

N-*trans*-cinnamoyl imidazole (NCI, Schonbaum *et al.*, 1961) has been used to determine the concentration of hydrolytic enzymes including subtilisin (Bender *et al.*, 1966). Subtilisin catalysed hydrolysis of NCI at pH 7 gives an initial rapid decrease in absorption at 335 nm (negative burst) on formation of the relatively stable cinnamoyl-subtilisin intermediate which is only slowly hydrolysed. Subtilisin and NCI react in a 1:1 stoichiometry. The number of moles of NC bound, and therefore the number of moles of subtilisin present, are calculated from the magnitude of the "burst" and the extinction coefficient of NCI (ϵ_{335} = 9.04x10^3 M^{-1} cm^{-1} in 0.025 M phosphate buffer pH 7.14).

The "observed" burst (π_{obs}) requires two corrections to derive the magnitude of the "actual" burst (π_{act}). There is a dilution factor (π_{dil}) on adding the enzyme solution (100-200 μl) to the buffered substrate and also the absorbance of the enzyme itself (π_{enz}) at 335 nm, both of which must be taken into account. Fortunately the difference in absorbance between subtilisin and cinnamoyl-subtilisin is negligible and need not be corrected for.

Thus,

$$\pi_{obs} = \pi_{act} + \pi_{dil} - \pi_{enz}$$

Figure 5.1 shows a schematic representation of the observed and actual bursts and the necessary corrections, which are probably best viewed as corrections to the base line.

FIGURE 5.1: Active-site titration using NCI

Addition of enzyme

Π_{dil}

Π_{enz}

BURST

Π_{obs}

Π_{act}

$$\Pi_{act} = \Pi_{obs} - \Pi_{dil} + \Pi_{enz}$$
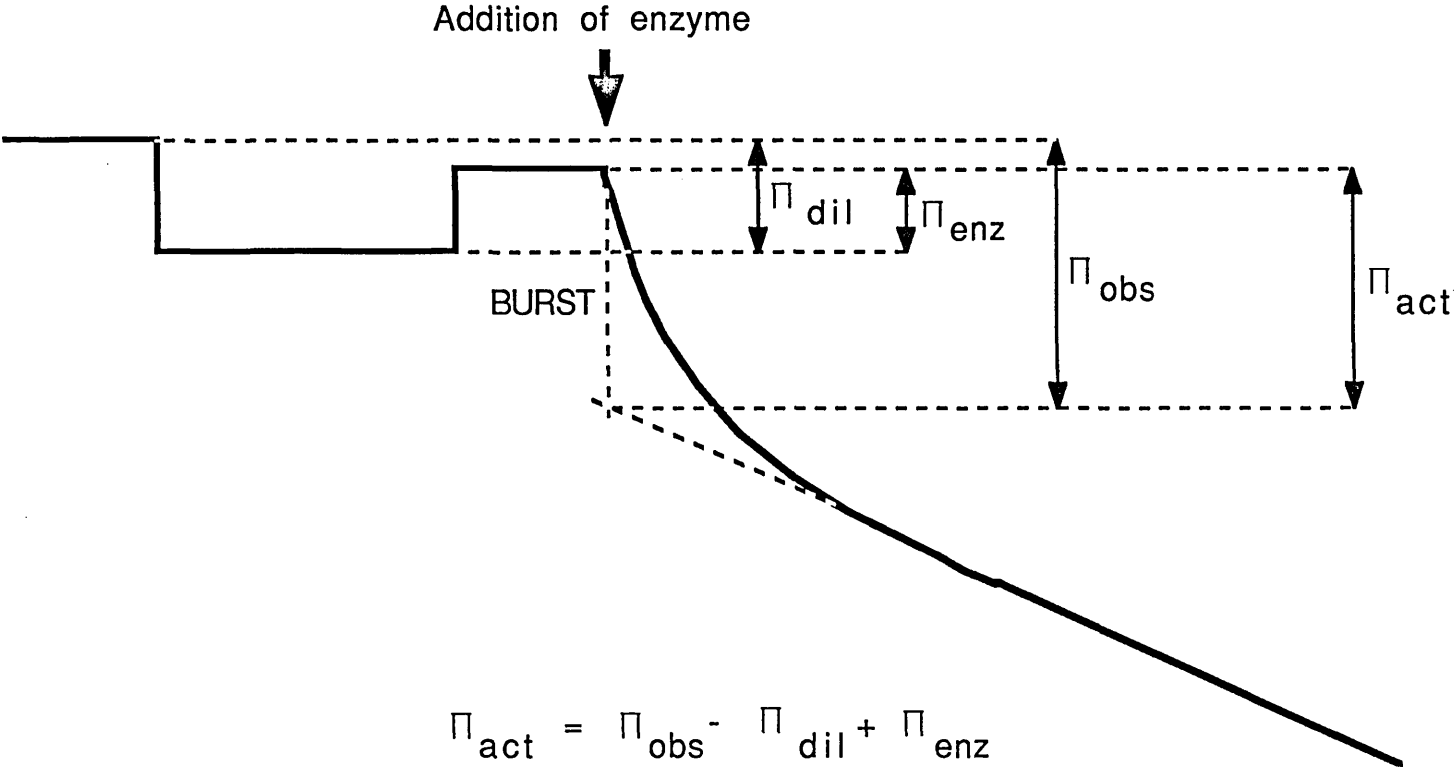
5.5.3 Experimental procedure

The reaction was carried out in 0.025 M sodium/potassium phosphate buffer pH 7.14 at 25 °C. 50 μl of NCI (0.66 mM in dry acetonitrile) was added to 2.35 ml of buffer in a 3 ml quartz cuvette maintained at 25 °C in a Gilford 2600 spectrophotometer, and the slow decrease in absorption at 335 nm, due to the spontaneous hydrolysis of the NCI, monitored for 60-100 sec. A 200 μl aliquot of the concentrated subtilisin stock (approximately 10 mg ml^{-1}) was quickly added and the solution rapidly mixed by aspiration and the burst recorded. The correction factor for the absorbance of the enzyme at 335 nm was determined from wavelength scans of a sample of enzyme. The correction factor for dilution was found experimentally by adding 200 μl of buffer alone to the buffered substrate.

5.5.4 Discussion

It was generally found that using smaller amounts of enzyme (100 μl or less of a 10 mg ml^{-1} solution) gave small and less reproducible burst sizes. Duplicate runs were inconsistent and appeared to give an underestimate of π_{obs}.

Assays performed using larger quantities of enzyme (200 μl or more) gave reproducible values for the observed burst, and the derived enzyme concentrations from duplicate assays were usually identical to ±5%. Enzyme concentrations obtained from active-site titration were typically 70-80% of the values calculated from UV absorbance measurements. This illustrates that enzyme concentrations determined by UV absorbance must be viewed with caution.

The disadvantage of active-site titration using NCI is the large amount of enzyme which must be sacrificed (up to 10 mg). The use of the fermenter and tangential flow filtration techniques make it simple to purify >100 mg of enzyme, and so is more satisfactory for the present studies.

Preliminary results using ^{35}S-PMSF as an active-site titrant for subtilisin, which it irreversibly inhibits, demonstrate that a radiometric active-site titration is possible for this enzyme (A. J. Russell, personal communication). Such a procedure offers significant advantages in that a very much smaller amount of enzyme is required, and is also simple and quick to perform. This would allow the enzyme concentration to be determined on each occasion of use, rather than only once for each batch of enzyme purified and assuming that the concentration remains unchanged on storage.

5.5 Catalytic constants for the hydrolysis of peptide substrates

5.5.1 Succinyl-L-ala-L-ala-L-pro-L-phe *para*-nitroanilide

The hydrolysis of succinyl-L-ala-L-ala-L-pro-L-phe *para*-nitroanilide (suAAPFpN) by subtilisin results in the release of *p*-nitroanilide. The resultant increase in absorbance at 412 nm was monitored spectrophotometrically over a range of substrate concentrations from 0.01-0.8 mM. Initial rates of hydrolysis were determined from the initial increase in absorption using the extinction coefficient of the nitroanilide ion ϵ_{412} = 8480 M^{-1} cm^{-1} (DelMar *et al.*, 1979).

Substrate stocks (0.8 mM) and dilutions were prepared in 0.1 M Tris-HCl buffer pH 8.6. Individual rate assays were initiated on addition of 10 μl of enzyme (typically 13-15 μM by active site titration in 0.01 M phosphate buffer pH 6.2 containing 0.08 M NaCl) to 3 ml of buffered substrate in a

cuvette maintained at 25 °C in a Gilford 2600 spectrophotometer. Assays were performed in duplicate for at least 10 different substrate concentrations covering the range 1/5 to 5 times the expected value of K_m.

Values of K_m and V_{max} were obtained from Eadie plots (v against v/[S]). k_{cat} was calculated from V_{max} using the equation

$$V_{max} = k_{cat} [E_0]$$

The enzyme concentration, [E_0], was known from active site titration.

The actual values of [S] were calculated from the final absorbance at 412 nm on complete hydrolysis of the substrate.

5.5.2 Benzoyl-L-val-L-gly-L-arg *para*-nitroanilide

Hydrolysis of benzoyl-L-val-L-gly-L-arg *para*-nitroanilide (bzVGRpN) by subtilisin also results in the release of *p*-nitroanilide. Rate assays were performed across the substrate concentration range 0.0125-0.4 mM in 0.1 M Tris-HCl pH 8.6. Reactions were initiated by adding 10 μl of enzyme (10-20 μM in 0.01 M phosphate buffer pH 6.2 plus 0.08 M NaCl) to 3 ml of buffered substrate maintained at 25 °C in a thermostatted Gilford 2600 spectrophotometer. The initial rates of hydrolysis were determined from the increase in absorbance at 412 nm as described above.

5.6 The pH dependence of k_{cat}/K_m

The variation of k_{cat}/K_m with *p*H reflects the ionisation of the free enzyme under conditions where the substrate itself does not ionise (Fersht, 1985). Thus, the pK_a of the catalytically functional histidine in subtilisin

was determined from the variation of k_{cat}/K_m with pH for the hydrolysis of suAAPFpN.

Assays were performed in phosphate or imidazole buffers at constant ionic strength, across the pH range 6-8 at approximately 0.1 pH unit intervals. Tables 5.1 and 5.2 show the composition of phosphate and imidazole buffer stocks at ionic strengths 1.0 and 0.1 respectively. Substrate stocks were made up at 8 mM in DMSO.

For assays performed in phosphate buffers at ionic strength 0.1 and below, the stock buffers were diluted appropriately. Ionic strength was adjusted by the addition of KCl for assays above ionic strength 0.1.

Assays in imidazole buffer, contained 0.001 mM imidazole-imidazolium hydrochloride and ionic strength was adjusted by addition of KCl.

Individual assays were initiated by ading 10 μl of enzyme (10-20 μM in phosphate or TES elution buffer) to 3 ml of buffered substrate (8 μM $<<K_m$), maintained at 25 °C in a disposable cuvette. Assays were performed at least in duplicate at each pH value.

The increase in absorbance at 412 nm was followed to completion. The digital output from the Gilford 2600 spectrophotometer was converted from parallel to serial using a device designed and supplied by D. G. Electronics Ltd, then stored directly on computer (BBC micro). The first order rate constants (k_{cat}/K_m) were derived using a non-linear regression algorithm to analyse the data (Leatherbarrow, 1987).

The pH of each sample was measured directly in the cuvette at 25 °C immediately after each assay using a PHM64 research pH meter and

Table 5.1: Composition of constant ionic strength phosphate buffers

*p*H (calculated at I=0.1)	g/500 ml Na$_2$HPO$_4$	KH$_2$PO$_4$	*p*H (measured at I=0.1)
6.00	7.16	47.46	6.01
6.20	9.63	40.32	6.22
6.40	12.55	33.32	6.41
6.60	14.98	24.97	6.63
6.70	16.21	21.45	6.73
6.80	17.33	18.22	6.84
6.90	18.33	15.31	6.94
7.00	19.22	12.75	7.03
7.10	20.00	10.53	7.14
7.20	20.66	8.64	7.24
7.30	21.21	7.05	7.34
7.40	21.67	5.72	7.44
7.60	22.37	3.73	7.63
7.80	22.83	2.40	7.84
8.00	23.13	1.53	8.04

Notes to table 5.1:
The weights given are for 10 x concentrated buffer stocks (ie I = 1.0) and the solutions prepared volumetrically.
Amounts required were calculated from:
$I = 1/2\Sigma mz^2$
where m is the molarity of the ion and z the charge on the ion;
and $pH = pK_a + \log$ [Base]/[Acid]
where the base is Na$_2$HPO$_4$ and the acid is KH$_2$PO$_4$.
The pK_a value used was 6.84

Table 5.2: Composition of constant ionic strength imidazole-imidazolium hydrochloride buffers

pH (calculated)	Vol. 2 M imidazole added to 50 ml of 0.2 M HCl. Final vol = 100 ml Ionic strength = 0.1 M	Final concentration of imidazole (mM)
6.10	5.50	110
6.21	5.65	113
6.30	5.80	116
6.40	6.00	120
6.49	6.25	125
6.59	6.55	131
6.70	7.00	140
6.80	7.50	150
6.91	8.25	165
7.03	9.25	185
7.10	10.00	200
7.21	11.25	230
7.30	13.00	260
7.40	15.00	300
7.50	17.50	350
7.57	20.00	400
7.70	25.00	500
7.81	30.50	610
7.91	37.50	750
8.00	45.00	900

Notes to table 5.2:
pK_a of imidazole taken as 7.1 at 25 °C.
Ionic strength constant at 0.1 M
Amounts calculated using $pH = pK_a + \log$ [Base]/[Acid]
where the base is imidazolium-HCl and the acid is HCL

GK2421C electrode (Radiometer, Copenhagen). The meter was calibrated at 25 °C using standard buffer solutions at pH 7.00 and 4.01 (Radiometer) and rechecked regularly to ensure that there was no drift in the electrode.

CHAPTER 6: DESIGN AND CONSTRUCTION OF MUTANTS

6.1 Introduction

The importance of electrostatic interactions in protein structure and activity has been discussed (Section 1.4). Chemical modification studies on chymotrypsin (Valenzuela and Bender, 1971), have shown that the pH dependence of enzyme catalysis by serine proteases alters with changes in overall surface charge. This suggests that a possible general method for modifying the pH dependence of enzymes is by altering the electrostatic environment of ionisable catalytic groups in the active site, so changing their $p\mathrm{K}_a$. The aim of this project is to test the generality of this approach to the deliberate alteration of the pH dependence of enzyme catalysis.

In the case of subtilisin, enzyme activity is critically dependent on the charge state of the active-site histidine residue (His-64). We can predict that mutations involving ionic groups in the vicinity of the active site will stabilise or destabilise the low pH protonated form of His-64 through simple electrostatics and raise or lower its $p\mathrm{K}_a$ accordingly. The magnitude of any $p\mathrm{K}_a$ changes produced by such mutations were unknown at the start of this study.

6.2 Rationale for mutant selection

Our philosophy for mutagenesis is to make substitutions which will result in small but measurable effects, and are unlikely to perturb the gross structure of the protein. Thus, where possible, mutations were limited to non-conserved residues in subtilisin BPN' and the substitutions were for amino acids known to occur at that position in the homologous enzymes.

To investigate electrostatic effects on the *p*H dependence of enzyme catalysis we are focussing on amino acids with ionic side chains which are located at the enzyme surface. Mutant design was therefore based on inspection of the crystal structure of subtilisin BPN' in light of the known primary structures of five other homologous enzymes.

Three non-conserved positions were initially identified as candidates for mutagenesis, where charged residues appear at the enzyme surface in one or more of the homologous enzymes. These residues are: Asp-99; Asp-36 and Lys-213, in subtilisin BPN'. It should be noted that due to a protein sequencing error the residue at position 99 was originally assigned as Ala and this was built into early crystallographic models (Kraut, 1971). A further non-conserved position even closer to the active site was also identified, Asn-62, however no charged residues appear at this position in the homologous enzymes. None of the residues selected for mutagenesis have been implicated in direct interactions with the substrate nor are they involved in the mechanism of catalysis.

6.3 Mutant selection and structural considerations

Interatomic distances and information regarding the environment of the target residues were obtained from the atomic coordinates of subtilisin BPN' (inhibited with CI-2, McPhalen *et al.*, 1985), using an Evans and Sutherland PS300 to run the graphics programme FRODO (Jones, 1978).

6.3.1 Mutation of Asp→Ser-99

The primary sequence of the known subtilisins around position 99 is shown in Table 6.1. The aspartate at this position in subtilisin BPN' can be replaced by either Ser or Thr in the subtilisins from other species of

Table 6.1	Primary sequences of homologous subtilisins around position 99

ENZYME	90								99							105
BPN'/Novo	L	Y	A	V	K	V	L	G	A	D̲	G	S	G	Q	Y	S	W
Carlsberg	L	Y	A	V	K	V	L	N	S	S	G	S	G	S	Y	S	G
DY	L	Y	A	I	K	V	L	N	S	S	G	S	G	T	Y	S	A
Amylosacchariticus	L	Y	A	V	K	V	L	D	S	T	G	S	G	Q	Y	S	W
168	L	Y	A	V	K	V	L	D	S	T	G	S	G	Q	Y	S	W
Mesentericus	L	Y	A	V	K	V	L	D	S	T	G	S	G	Q	Y	S	W

Table 6.2	Primary sequences of homologous subtilisins around position 36

ENZYME		30						36			40					
BPN'/Novo	S	G	V	I	D	S	G	I	D̲	S	S	H	P	D	L	K
Carlsberg	T	G	V	L	D	T	G	I	Q	A	S	H	P	D	L	N
DY	T	G	I	I	D	T	G	I	A	A	S	H	T	D	L	K
Amylosacchariticus	S	G	V	I	D	S	G	I	D	S	S	H	P	D	L	N
168	S	G	V	I	D	S	G	I	D	S	S	H	P	D	L	N
Mesentericus	V	A	V	I	D	S	G	I	D	S	S	H	P	L	D	N
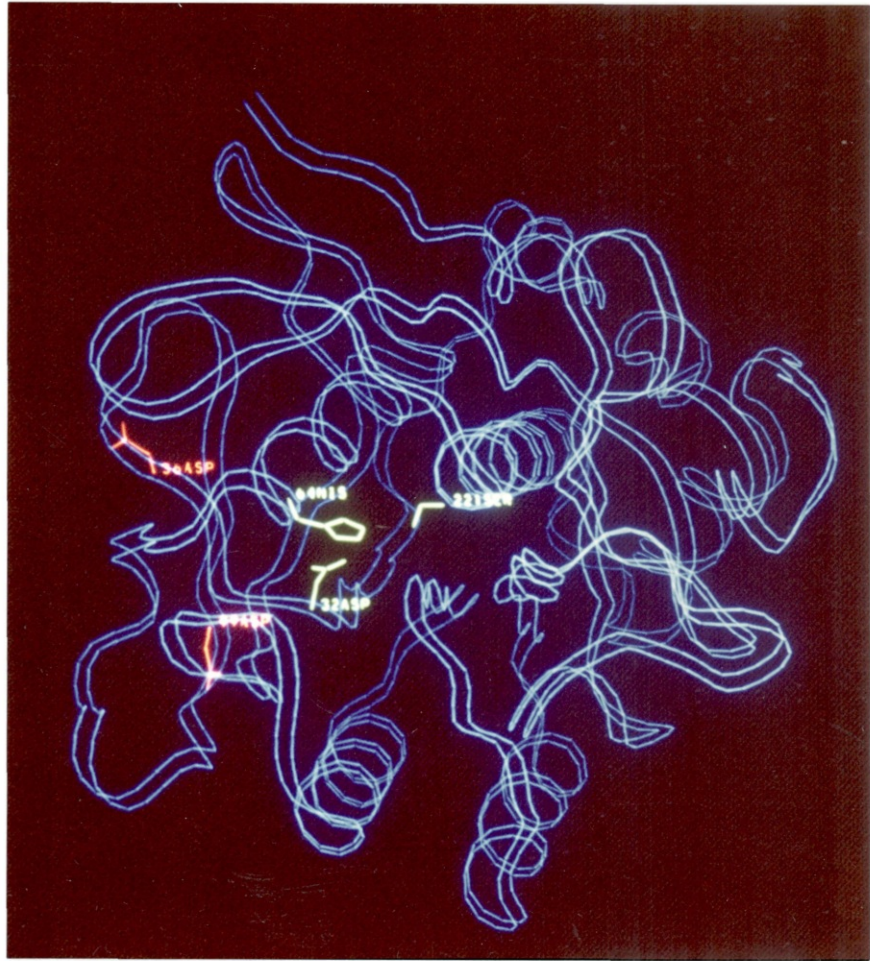
Figure 6.1   The location of Asp-99 and Asp-36 with respect to His-64.

Bacillus. The mutation Asp99→Ser was chosen to effect the replacement of the negatively charged aspartate with a neutral residue. Figure 6.1 shows that Asp-99 is located in a surface loop, just at the rim of the active-site cleft. The carboxylate group projects into the solvent and is some 12-13 Å distant from the imidazole of His-64 (R. Bott personal communication; McPhalen *et al.*, 1985). The environment separating Asp-99 and His-64 appears to consist mainly of solvent. The close similarity of the crystal structures of subtilisins BPN' and Carlsberg and the high degree of amino-acid sequence homology between the subtilisin family means that *a priori* the mutation Asp-99→Ser (abbreviated DS99) should avoid the possibility of structural perturbation and have no effect on the enzymes activity other than by altering an electrostatic interaction. The removal of a negative charge at this position is expected to destabilise the low pH protonated form of His-64 and therefore result in a lower pK_a in the mutant enzyme.

6.3.2 Mutation of Asp-36→Gln

Asp-36 is located in a surface loop completely outside the active-site cleft. It is separated from the imidazole of His-64 by a distance of approximately 15-16 Å. The environment separating Asp-36 and His-64 consists largely of protein side and main chain atoms of the α-helix containing His-64. The position of Asp-36 in relation to the active site is shown in Figure 6.1. The aspartate at this position can be replaced by either Gln or Ala in the homologous enzymes (Table 6.2). The mutation selected was Asp-36→Gln (DQ36), this will remove the negatively charged aspartate and replace it with a neutral residue. Since this is a naturally occuring substitution at this position, no artefacts due structural perturbations are expected. The mutation DQ36 is also expected to result in a reduced pK_a, since we expect that the removal of a negative charge will destabilise the protonated form
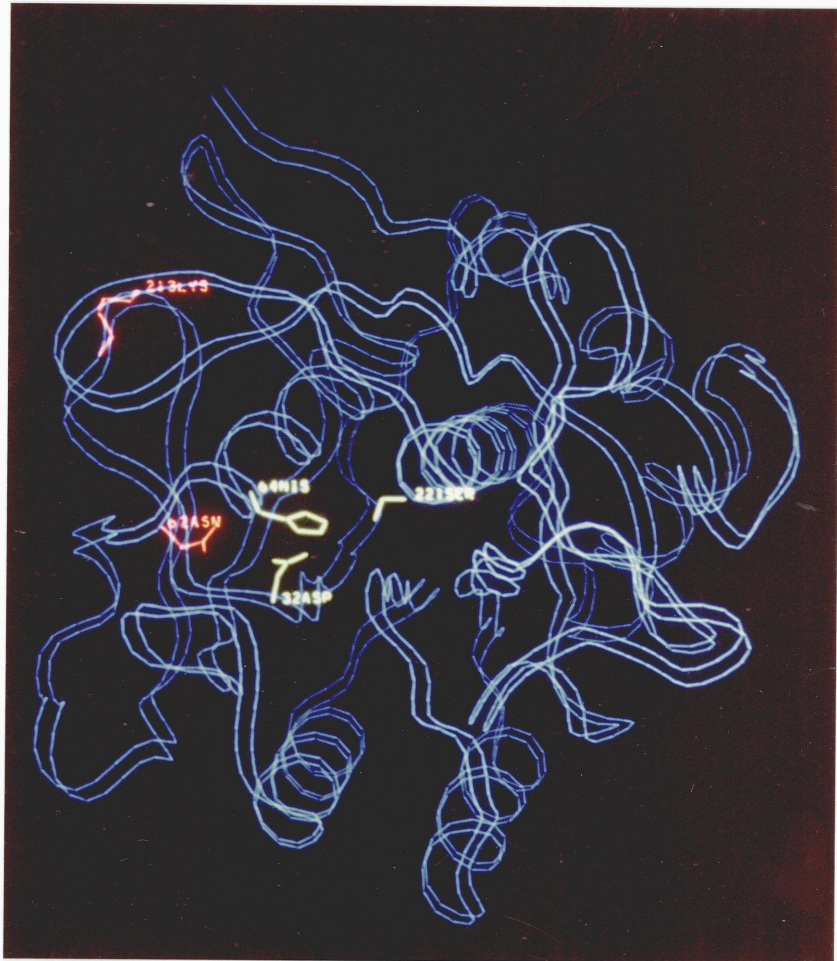
Figure 6.2   The location of Asn-62 and Lys-213 with respect to His-64.

Table 6.3	Primary sequences of homologous subtilisins around position 213

ENZYME	205							213						220
BPN'/Novo	I	Q	S	T	L	P	G	N	K̲	Y	G	A	Y	N G T
Carlsberg	V	Y	S	T	Y	P	T	N	T	Y	A	T	L	N G T
DY	V	Y	S	T	Y	P	S	N	T	Y	T	S	L	N G T
Amylosacchariticus	I	Q	S	T	L	P	G	G	T	Y	G	A	Y	N G T
168	I	Q	S	T	L	P	G	G	T	Y	G	A	Y	N G T
Mesentericus	I	Q	S	T	L	P	G	G	T	Y	G	A	Y	N G T

Table 6.4	Primary sequences of homologous subtilisins around position 62

ENZYME	55					62		64						70
BPN'/Novo	T	N	P	F	Q	D̲	N	N	S	H	G	T	H	V A G
Carlsberg	A	-	Y	N	T	D	G	N	G	H	G	T	H	V A G
DY	S	-	Y	N	T	D	G	N	G	H	G	T	H	V A G
Amylosacchariticus	T	P	N	Y	Q	D	G	S	S	H	G	T	H	V A G
168	T	N	P	Y	Q	D	G	S	S	H	G	T	H	V A G
Mesentericus	T	N	P	Y	Q	D	G	S	S	H	G	T	H	V A G

of His-64.

6.3.3 Mutation of Lys-213→Thr

Figure 6.2 shows the position of the Lys residue at position 213 in subtilisin BPN' in relation to the active site and the catalytic residues. Lys-213 is located in a surface loop, just outside the rim of the active site cleft, and the side chain projects out into the solvent. The electron density of this group is rather poorly defined and the temperature (B) factor is high. This side chain is considered to be highly mobile and the charged ϵ-amino group can not be accurately located. The distance between the charged amino group and the imidazole $N\epsilon2$ of His-64 can only be given approximately as 19 ± 4 Å (R. Bott, personal communication). The environment between positions 213 and 64 is predominantly aqueous.

Only in subtilisin BPN' is residue 213 a Lys, in all other cases this position is occupied by a Thr residue (Table 6.3). Mutagenesis of Lys-213→Thr (KT213) will reduce the net positive charge at the enzymes surface. This change in the surface electrostatic field should stabilise the positively charged, protonated form of His 64 and thereby increase its pK_a. The mutation KT213 should not affect enzyme activity other than by electrostatic effects since the substitution selected occurs naturally.

6.3.4 Mutation of Asn-62→Asp

To investigate whether the magnitude of any pK_a shift can be increased by engineering mutations closer to the active site, the region within a 10 Å radius of His-64 was examined for non-conserved positions. Unfortunately there were no non-conserved charged groups within this distance. The introduction of a charged residue at a non-conserved position which did not

160

normally harbour a charged group was therefore necessary, in-spite of the potential risk of stuctural perturbations resulting in an inactive enzyme or artefactual data.

Table 6.4 shows the amino acid sequence of the subtilisin family around the non-conserved position 62 which is normally an Asn or Ser residue. The introduction of an Asp at this position would give the required increase in net negative charge and should therefore result in an increased pK_a. The selection of an Asp for substitution at this position is based on the observation that at six other non-conserved positions throughout the enzyme Asp can replace Ser or Asn in the homologous enzymes. At two of these positions all three residues occur (positions 248 and 259). Furthermore, the side chains of Asn and Asp are closely isosteric so this substitution is likely to be satisfactory.

Figure 6.2 shows the position of Asn 62 in relation to the active site residues. The residue is located in a surface loop which joins the N-terminal end of the α-helix containing His-64. Asn-62 is situated just on the lip of the active-site cleft and the side-chain projects into this cleft. The distance from the amide group of Asn-62 to the imidazole N of His-64 is about 6.5 Å (McPhalen *et al.*, 1985), and this region consists entirely of solvent.

6.4 Oligonucleotide-directed mutagenesis

The construction of the mutations DS99, by single-primed oligonucleotide-directed mutagenesis (Section 4.5), and ND62 by a double-primed procedure (Section 4.6) has been described. The mutants KT213 and DQ36 were also constructed by the double-primed method mutagenesis.

6.4.1 Construction of KT213

A 17mer containing a central single mis-match was synthesised to mutate the Lys codon (AAA) at position 213 to an acceptable Thr codon (ACA) for expression in *B. subtilis*. The mutagenic primer had the sequence: 5'-CCCCGTATG̊TGTTCCA-3'. The mis-matched base is marked with an asterisk. The calculated T_d for this oligonucleotide is 52 °C. This mutagenic oligonucleotide specifically primed DNA synthesis from the correct position in the template strand when used in sequencing reactions. The primer was therefore considered suitable for mutagenesis. Site-directed mutagenesis was performed exactly as described (Section 4.6).

An aliquot of the primer extended/ligated material was transfected into competent *E. coli* BMH 71-18 *mut*L cells. The plaques obtained were transfered to nylon membranes and screened with 5' ^{32}P radiolabelled KT213 primer. Hybridisation was at 44 °C over night. The filters were given three 20 min washes in 6xSSC at room temperature then autoradiographed for 24 h. Virtually every plaque remained hybridised with the probe. The filters were rewashed at 50 °C (T_d-2 °C) in 6xSSC for 2x10 min, then autoradiographed. Approximately 23% of the plaques remained hyridised to the probe indicating that they contained the required mutant. The plaques were screened at a high density (approximately 1500 plaques per 9 cm plate) and it was not possible to select a single well separated mutant plaque. A patch of plaques was therefore picked and the phage diluted and replated at low density (<100 plaques per plate). These were rescreened by dot blotting to identify mutants.

Figure 6.3 shows the autoradiograms of the filter after low and high stringeny washes. Several of the putative mutants were picked and sequenced to confirm that the required mutation was present (Figure 6.4).
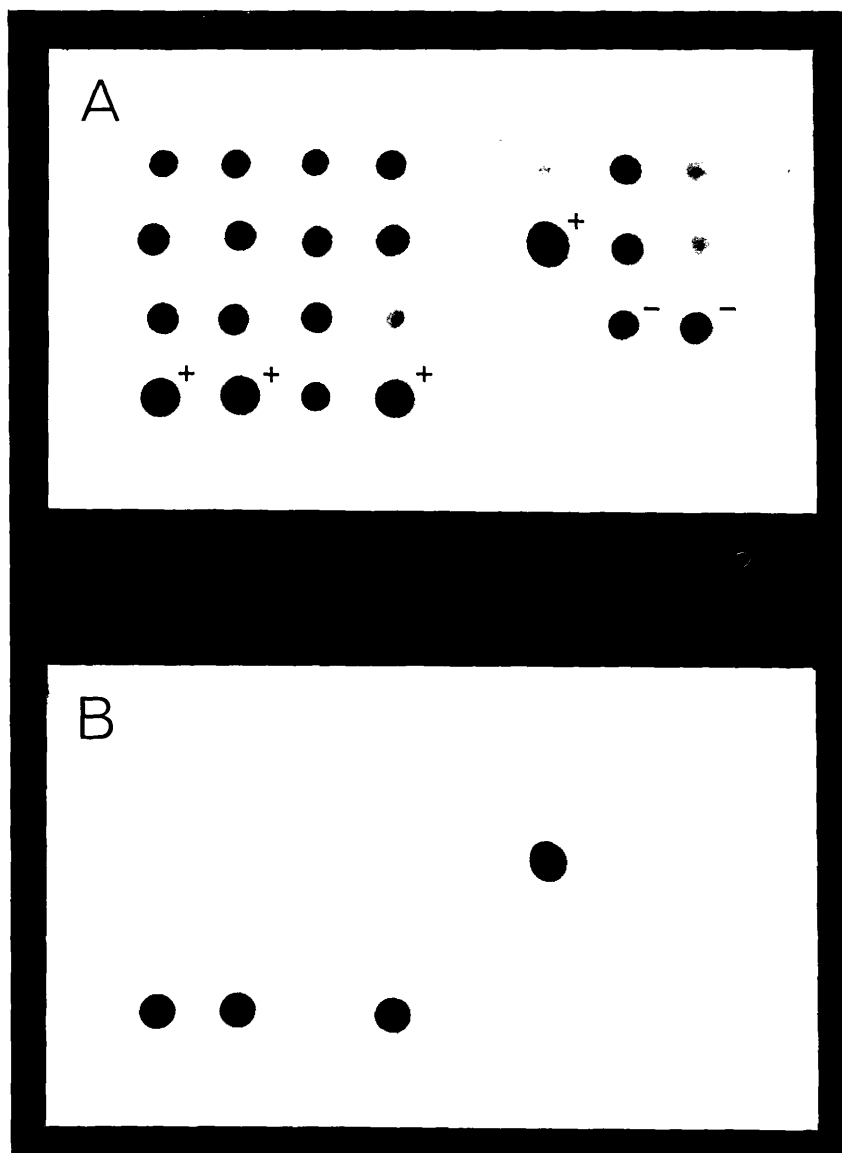
FIGURE 6.3:   Identification of KT213 mutants by dot-blot hybridisation
screening using ^{32}P radiolabelled KT213 probe.

A: Low stringeny wash; 3x10 min at 18 °C in 6xSSC. Exposed for 6 hr
at -70 °C with an intensifying screen.   Two control negative phage
supernatants are indicated.

B: Stringent wash; 1x5 min at 48 °C in 6xSSC.   Exposed for 6 hr at
-70 °C with an intensifying screen.   The four hybridisation positives
(presumed to be KT213 mutants) are also indicated in panel A.
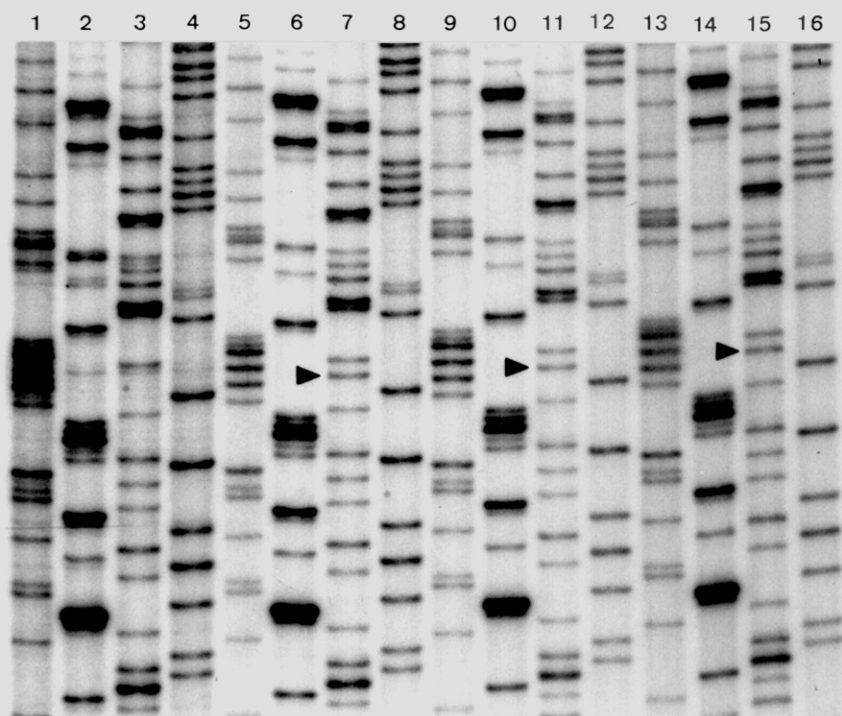
FIGURE 6.4:   Confirmation of the KT213 mutation by DNA sequencing

Tracks 1-4 are M13aprWT template DNA sequenced using the primer L, gel loading order is T C G A throughout.   Tracks 5-8, 9-12 and 13-16 are template DNAs prepared from three of the hybridisation positives identified in Figure 6.3, sequenced using the same primer.   Sequencing was done at 18 °C and the reaction products resolved on a 6% wedge gel using an aluminium plate to ensure even heat distribution.   Mutation of A at position 1408 to a C is confirmed by the replacement of a T by a G in the complementary strand synthesised during DNA sequencing.   All three templates contain the desired mutation (indicated by arrow heads).

A single confirmed mutant, designated M13aprKT213 was sequenced in full to confirm that no unwanted changes had occured.

The gene harbouring the KT213 mutation was purified from M13aprKT213 RF DNA and transferred to pUB110 for expression in *B. subtilis* DB104 as described (Section 4.5.7). *B. subtilis* DB104 transformants harbouring the reconstructed plasmid pPT30-KT213 exhibited zones of proteolysis characteristic of subtilisin expression. As anticipated, the mutant protein is active.

6.4.2 Construction and expression of subtilisin ND62

Conversion of the Asn codon (AAC) at position 62 to an acceptable Asp codon (GAC) for expression in *B. subtilis*, was achieved using a 15mer with a single, centrally located mis-match. The mutagenic primer had the sequence;

5'-TGAGAGT$\overset{*}{C}$GTTGTCT-3', and the calculated T_d is 44 °C. The construction and subsequent transfer of the mutated gene to the expression vector (to give pPT30-ND62) has already been described.

B. subtilis DB104 cells transformed with pPT30-ND62 exhibited zones of proteolysis when grown on selective media containing milk solids, indicating that this mutant is active in spite of the substitution of a residue which does not occur at this position in any known subtilisin.

6.4.3 Construction and expression of subtilisin DQ36

A double mis-matched primer was synthesised to effect the mutation of the Asp-36 codon (GAT) to a suitable Gln codon (CAG) for expression in *Bacillus*. The double mis-matched primer designed to effect this mutation

165

had the sequence: 5'-TGAGAAGACTGGATACCGC-3', the mis-matched bases are indicated by the asteriscs. The T_d for this primer is 58 °C. Sequencing reactions primed with 1 pmol of this oligonucleotide showed a very high degree of secondary priming from other positions within the template DNA, and resulted in unreadable sequencing autoradiograms. This problem was overcome by using reduced amounts of primer and performing the annealing step at 50 °C rather than at room temperature. The sequencing reactions were also done at elevated temperature in order to minimise non-specific priming. Sequencing reactions containing 0.1 p mol of the DQ36 primer and carried out at 37 °C gave specific priming from the required position. The same ratio of template to primer (5:1) was used for mutagenesis and the annealing also done at 50 °C. The extension/ligation reaction was initiated at 37 °C for 5 min to ensure specific priming then placed at 15 °C over-night. Even with an excess of unprimed template DNA resulting in high levels of wild-type progeny, direct plaque screening with the radiolabelled DQ36 primer showed that 1.8% of the plaques contained the required mutant. One of the hydridisation positives was sequenced in full to confirm the mutation and to ensure that no adventitious mutations had occured as a result of secondary priming.

The mutated gene was transferred to pUB110 to generate pPT30-DQ36 exactly as desribed (Section 4.5.7). Subtilisin DQ36 was expressed in *B. subtilis* DB104 harbouring pPT30-DQ36 and gave colonies exhibiting the characteristic halo of proteolysis on selective plates containing milk solids.

6.5 Summary

Oligonucleotide-directed mutagenesis of the subtilisin BPN' gene has been used to construct four variant enzymes differing only in the distribution of surface charges near the active site. All four mutants retain proteolyic

activity. The effects of these mutations on the pH dependence of the enzymes will be determined using enzyme kinetics.

CHAPTER 7: ANALYSIS 1 - ENZYME PURIFICATION AND KINETICS

7.1 Enzyme purification

Wild-type and mutant subtilisins were recovered from culture supernatants of *B. subtilis* DB104 harbouring pPT30, pPT30-DS99, pPT30-DQ36, pPT30-KT213 or pPT30-ND62 grown at 37 °C for 36 h in LB-broth containing glucose (0.2%) and kanamycin (25 μg ml^{-1}). Secreted protease was concentrated and then dialysed by buffer exchange using tangential-flow membrane ultra-filtration as described earlier (Section 5.2). Initially purification was by a single step of ion-exchange chromatography on CM-cellulose equilibrated and run in 0.1 M phosphate buffer *p*H 6.2 and bound subtilisin eluted with 0.1 M phosphate buffer containing 0.08 M NaCl. The procedure was later modified to utilise CM-TrisAcryl run in 0.01 M TES *p*H 6.2 and bound subtilisin eluted with 0.01 M TES *p*H 6.2 containing 0.1 M NaCl.

The switch to the use of CM-TrisAcryl equilibrated and run in 0.01 M TES *p*H 6.2 in preference to CM-cellulose in 0.1 M phosphate buffer *p*H 6.2 allowed higher flow rates to be achieved without compacting the resin and significantly reduced the purification time. Related NMR studies on mutant subtilisin required much larger amounts of enzyme (0.5-1 g) and cultures were grown in the presence of calcium chloride (10 mM) which stabilises the enzyme to autolysis (D. A. Estell, personal communication) and consequently gives much higher yields of product (approximately 100 mg per litre of recoverable enzyme compared with 15 mg per litre when grown in the absence of calcium). However, the use in this case of phosphate buffer would result in the precipitation of insoluble calcium phosphate. The buffer was therefore replaced by 0.01 M TES *p*H 6.2. Wild-type subtilisin BPN' purified by either method has identical kinetic properties (M. Bycroft, personal communication).

Table 7.1:   Summary of enzyme recoveries

Enzyme	Purification method	Total yield (mg)	% activity[1]
BPN'[2]	CM-cellulose phosphate buffer	90 mg	89%
DS99[2]	CM-cellulose phosphate buffer	36 mg	90%
DQ36	CM-Trisacryl TES buffer	29 mg	46%
KT213	CM-cellulose phosphate buffer	26 mg	77%
ND62	CM-cellulose phosphate buffer	125 mg	80%

Notes to table 7.1:
(1) Determined by active-site titration using NCI.
(2) A. J. Russell, personal communication.

Use of the term "purification" is perhaps rather inappropriate since the procedures used are only required to concentrate the secreted protease. The expression host produces no endogenous protease activity which might interfere with kinetic analysis (Kawamura and Doi, 1984; A. J. Russell, personal communication). Furthermore, active-site titration allows the actual concentration of catalytically functional enzyme to be determined even if contaminated with autolytic degradation products or other impurities. Purification to homogeneity is not therefore a requirement and significantly simplifies analysis of mutants.

Wild-type subtilisin BPN' and subtilisin-DS99 were purified by A. J. Russell. Table 7.1 summarises the purification details for the mutant subtilisins involved in this study.

7.1.1 Subtilisin DQ36

29 mg of subtilisin-DQ36 was recovered from a 10 l culture of *B. subtilis* DB104 harbouring pPT30-DQ36. Ion-exchange chromatography was on CM-TrisAcryl in TES buffer pH 6.2. The low yield in this case was due to the failure of the ultra-filtration membrane during processing, greater than 100 mg of enzyme would normally be obtained from this scale of culture. However, this amount of material was sufficient for analysis. Active-site titration of this mutant indicated that it was only 46% active compared with concentration determined from UV absorption at 280 nm. SDS-PAGE and iso-electric focussing confirmed that the protein was contaminated, presumably with autolytic degradation products.

7.1.2 Subtilisin KT213

Subtilisin KT213 was purified from a 2.5 l culture of *B. subtilis* DB104

harbouring pPT30-KT213. Proteins in the culture supernatant were precipitated using ammonium sulphate, and subtilisin KT213 purified by chromatography on CM-cellulose as described (Section 5.2).

A total of 26 mg of purified subtilisin KT213 was recovered. Active site titration with NCI indicated that the enzyme was 77% active compared with the concentration measured by UV absorption.

7.1.3 Subtilisin ND62

Subtilisin ND62 was purified from a 10 l culture of *B. subtilis* DB104 harbouring pPT30-ND62. 125 mg of enzyme was obtained after purification on CM-cellulose (Section 5.2). Active-site titration of the ND62 enzyme showed that it was 80% active compared with the concentration measured by A_{280}.

7.2 Kinetics of the subtilisins

The kinetic properties of the subtilisins have been comprehensively reviewed by Philipp and Bender (1983). Like the mammalian serine proteases, subtilisins catalyse the hydrolysis of both ester and amide substrates. Subtilisin is fairly non-specific towards ester substrates but shows specificity for amide substrates with bulky, hydrophobic side-chains.

The hydrolysis of ester and amide substrates is described by the following scheme:

$$E + S \underset{\rightarrow}{\leftarrow} E.S \xrightarrow{k_2} EA \xrightarrow{k_3} E + P_2$$

with K_S, k_2, k_3 above, and

$$+$$
$$P_1$$
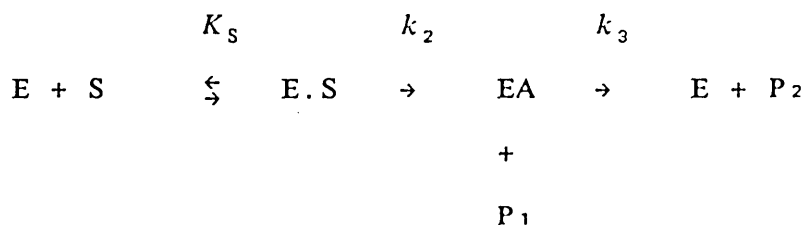
Where E and S represent the free enzyme and substrate. E.S represents the Michaelis complex, EA the acyl-enzyme intermediate and P_1 and P_2 the hydrolytic reaction products. The kinetic constants, k_{cat} and K_m, for this scheme are defined as:

$$k_{cat} = k_2 . k_3 / (k_2 + k_3)$$

and

$$K_m = K_S . k_3 / (k_2 + k_3)$$

For ester substrates, deacylation is the rate limiting step (ie. $k_2 \gg k_3$), whereas for amide substrates acylation is rate limiting (ie. $k_3 \gg k_2$) (Philipp and Bender, 1983). For nitroanilide substrates (amide) the ratio of k_3 to k_2 is 33 (Wells et al., 1986). Thus, steady state measurements of k_{cat} and K_m for suAAPFpN directly give k_2 and K_S respectively.

7.3 Catalytic constants for the hydrolysis of peptide substrates

The catalytic rate constant, k_{cat}, and the Michaelis constant, K_m, for the hydrolysis of synthetic peptide substrate suAAPFpN, by wild-type and mutant subtilisins, were determined from steady state kinetics at 25 °C in 0.1 M Tris-HCl pH 8.6. The assays were as described in section 5.5. The derived values of k_{cat} and K_m are summarised in Table 7.2.

7.3.1 Wild-Type subtilisin BPN'

The wild-type enzyme purified from cultures of B. subtilis DB104 harbouring pPT30 is characterised by values of k_{cat} and K_m of 57 sec^{-1} and 0.15 mM respectively for the hydrolysis of suAAPFpN (Thomas et al., 1985). These

Table 7.2:   Kinetic constants for the hydrolysis of peptide substrates by mutant and wild-type subtilisins.

Substrate	Enzyme	k_{cat} (sec^{-1})	K_m (mM)	k_{cat}/K_m [1] (sec^{-1} M^{-1})
suAAPFpN				
	BPN'	57	0.15	3.8×10^5
	DS99	45	0.13	3.46×10^5
	DQ36	59	0.14	4.21×10^5
	KT213	62	0.13	4.77×10^5
	ND62	52	0.57	9.12×10^4
bzVGRpN				
	BPN'	1.35	0.32	4.22×10^3
	ND62	0.08	0.16	5.00×10^2

Notes to Table 7.2:

Wild-type and DS99 data from Thomas *et al* (1985).

Assays performed in 0.1 M Tris-HCl *p*H 8.6 at 25 °C.

(1) Limiting value of k_{cat}/K_m at high *p*H.

values are identical to those of an authentic sample of subtilisin BPN' as used in the original crystallographic studies (Wright *et al.*, 1969; A. J. Russell, personal communication). Furthermore, the enzyme produced in this study has identical activity to that obtained from an independently cloned subtilisin BPN' gene (Wells *et al.*, 1983; D. A. Estell and J. A. Wells, personal communication). Other samples of supposedly subtilisin BPN' obtained from commercial suppliers are clearly not the authentic enzyme since values of k_{cat} and K_m for the same substrate are 642 sec^{-1} and 0.225 mM (Russell and Fersht, 1986). Thus, both kinetic data and DNA sequence data confirm that the enzyme under investigation is authentic subtilisin BPN'.

7.3.2 Subtilisin DS99

The activity of subtilisin DS99, at pH 8.6 in 0.1 M Tris-HCl, on the synthetic substrate suAAPFpN is similar to that of the wild-type enzyme. The catalytic rate constant, k_{cat}, is 45 sec^{-1} which is slightly lower than that of wild-type enzyme. K_m for this mutant is 0.13 mM (Thomas *et al.*, 1985). Control experiments have shown that the substrate used is only 88% pure, substrate concentrations were corrected to eliminate this error. The mutation Asp-99→Ser appears to have little effect on substrate binding and causes only a slight reduction in the activity of the enzyme against this substrate.

7.3.3 Subtilisin DQ36

The mutant subtilisin DQ36 also has similar activity to the wild-type enzyme against suAAPFpN. Figure 7.1 shows the initial rate (v) of hydrolysis plotted against substate concentration [S] and also the derivative "Eadie" plot of v against v/[S]. The derived values of k_{cat} and K_m are

Figure 7.1:   Subtilisin DQ36: dependence of rate of hydrolysis on

                   substrate concentration

Initial rate (AU min^{-1}) is plotted against the concentration of suAAPFpN

(mM).   Assays were done at 25 °C in 0.1 M Tris-HCl buffer pH 8.6.

Inset: Eadie-Hofstee plot.

Figure 7.2:   Subtilisin KT213: dependence of rate of hydrolysis on substrate concentration

Initial rate (AU min⁻¹) is plotted against the concentration of suAAPFpN (mM).   Assays were done at 25 ˚C in 0.1 M Tris-HCl buffer pH 8.6.

Inset: Eadie-Hofstee plot of initial rate against rate/substrate concentration.

59 sec^{-1} and 0.14 mM repectively. The mutation Asp-36→Gln has, therefore, no effect on the catalytic activity of the enzyme and results in a very small decrease in K_m, possibly due to the removal of an electrostatic interaction which would normally tend to repel the negatively charged substrate.

7.3.4 Subtilisin KT213

Figure 7.2 shows the initial rate of hydrolysis at varying concentrations of the synthetic substrate catalysed by subtilisin KT213. The data are shown graphically. The derived value of K_m is 0.13 mM (corrected for errors in substrate concentration). For this mutant k_{cat} is 62 sec^{-1}. These values are very similar to those of the wild-type enzyme. The mutation therefore appears to have a negligible effect on either substrate binding or catalytic activity.

7.3.5 Subtilisin ND62

The mutation ND62 gives values of 52 sec^{-1} for k_{cat} and 0.6 mM for K_m. The data are plotted graphically in Figure 7.3. The catalytic activity of this mutant is not detectably altered from wild-type for this substrate. However, the mutation clearly does affect substrate binding, K_m being increased about four fold. The mutation introduces a negatively charged group at a position where it projects into the active-site cleft. The reduced substrate binding can be interpreted as being due to electrostatic repulsion between the substrate succinyl group and the carboxylate group introduced by the mutation. The catalytic constants for the hydrolysis of the positively charged substrate bzVGRpN by subtilisin ND62 at 25 °C in 0.1 M Tris-HCl pH 8.6 are 0.16 mM for K_m and 0.08 sec^{-1} for k_{cat} (Figure 7.4). The corresponding wild-type values for this substrate are 1.35 sec^{-1}
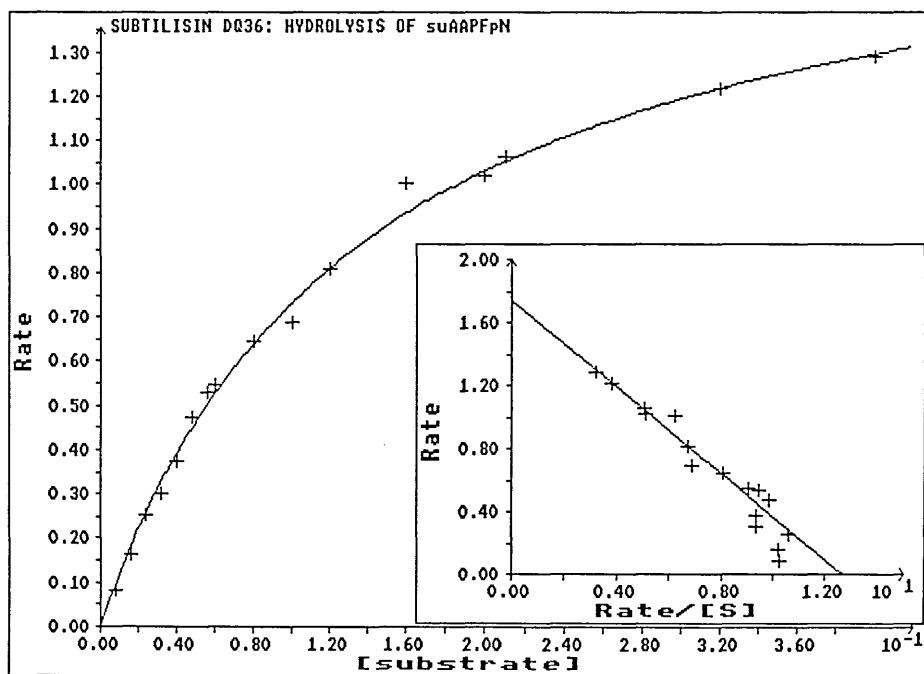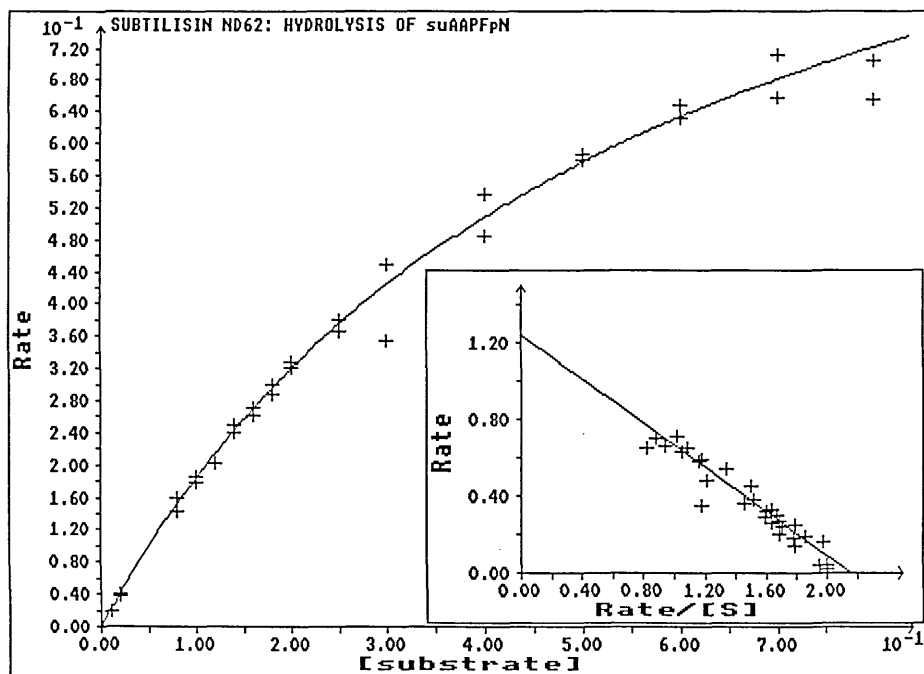
Figure 7.3: Subtilisin ND62: dependence of rate of hydrolysis on substrate concentration

Initial rate (AU min^{-1}) is plotted against the concentration of suAAPFpN (mM). Assays were done at 25 °C in 0.1 M Tris-HCl buffer pH 8.6. Inset: Eadie-Hofstee plot.
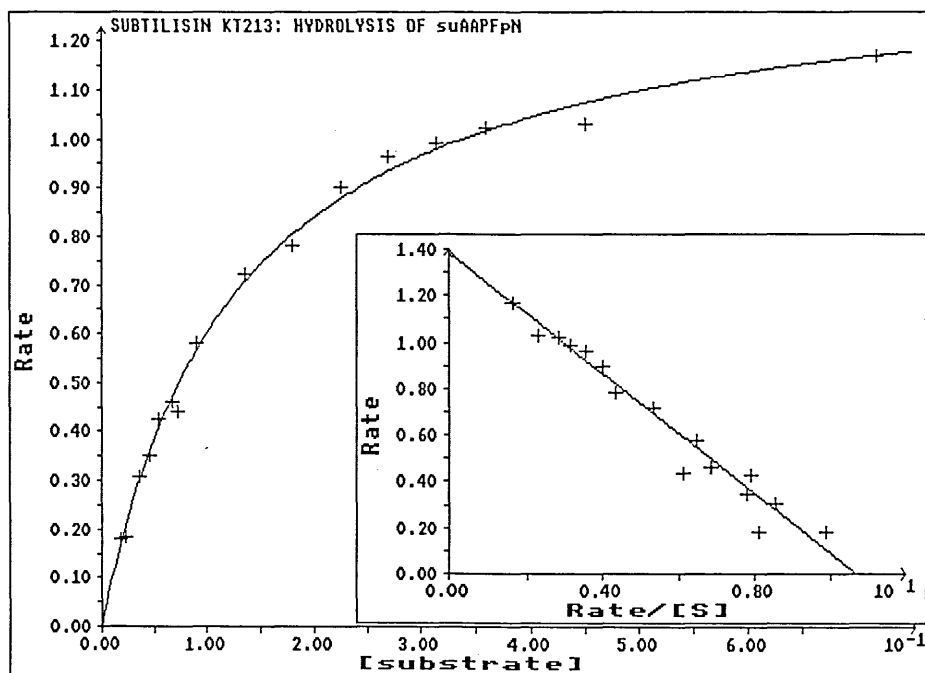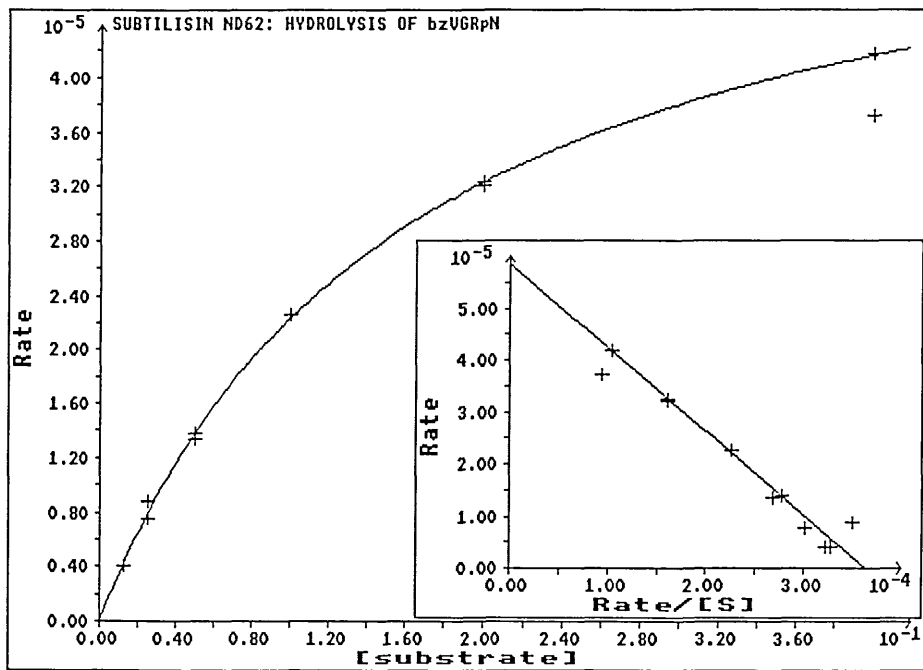
Figure 7.4:   Subtilisin ND62: dependence of rate of hydrolysis on concentration of the positively charged substrate bzVGRpN

Initial rate (AU min^{-1}) is plotted against the concentration of bzVGRpN (mM). Assays were done at 25 °C in 0.1 M Tris-HCl buffer pH 8.6. Inset: Eadie-Hofstee plot.

179

and 0.32 mM (A. J. Russell, personal communication). The introduction of the negative charge at position 62 in subtilisin therfore leads to a reduction in K_m to about half of the wild-type value for the positively charged substrate. This suggests stronger binding of the positively charged substrate to the ND62 enzyme, as might be expected from the changed electrostatic interaction. However, the catalytic rate constant for this mutant is reduced about 17-fold compared with the wild-type enzyme. This may reflect increased occupancy of a non-productive mode of binding with this substrate, a phenomenon known to occur with subtilisin and some other synthetic peptide substates (Poulos *et al.*, 1976).

7.4 Conclusions

Steady state kinetic analysis of the four mutant subtilisins constructed in this study, indicates that there have been no gross structural changes to the enzyme as a result of the mutations introduced by site-specific mutagenesis. The catalytic activities of the mutant subtilisins, as measured by k_{cat} for the hydrolysis of suAAPFpN, are not significantly different from that of wild-type. In three of the mutants substrate binding is unchanged from wild-type, only for the mutant ND62 is there a significant change in K_m which is increased to about four times the wild-type value. This increase is entirely consistent with the idea that the introduction of a negatively charged residue close to the substrate binding site will reduce the affinity of that mutant enzyme for a negatively charged substrate.

CHAPTER 8: ANALYSIS 2 - $p\text{K}_a$ DETERMINATION

8.1 Introduction

The ability of subtilisin to function catalytically is dependent on the charge state of the active site histidine. When His-64 is protonated it can no longer function as a general base and the enzyme is inactive. Thus the activity of subtilisin varies with pH following the ionisation of this residue. The effect of surface charge on the $p\text{K}_a$ of serine proteases has been discussed (Section 1.5).

Monitoring the pH dependence of k_{cat}/K_m gives the ionisation constants of the free enzyme and the free substrate (Fersht, 1985). $p\text{K}_a$ values determined from such plots should be independent of substrate in regions where the substrate does not ionise. The pH range investigated was restricted to the region around neutrality, thereby excluding the ionisation of the succinyl group on the substrate and also perturbations due to the ionisation of other residues on the enzyme, for example Asp and Glu at low pH and Lys, Tyr, Cys and eventually Arg at high pH.

8.2 $p\text{K}_a$ determination

The active-site $p\text{K}_a$s of wild-type and mutant subtilisins were accurately determined from the variation of k_{cat}/K_m with for the hydrolysis of suAAPFpN with pH as described (Section 5.4). The value of k_{cat}/K_m at each pH was determined from measurements of the release of products at substrate concentrations much lower than K_m, as described by Fersht and Renard (1974). When $[S] \ll K_m$ integration of the Michaelis-Menten equation gives:

$$\ln[S_t] = \ln[S_0] - (k_{cat}/K_m)[E_0]t$$

Plotting $\ln[S]$ aginst time (t) therefore gives a straight line of gradient $= -k_{cat}/K_m[E_0]$. Since the enzyme concentration is known and remains constant throughout the experiment, k_{cat}/K_m can be determined at each pH. Under the experimental conditions used first order kinetics should be obeyed which are independent of initial substrate concentration, thereby avoiding one possible source of error.

The increase in absorption at 412 nm, on hydrolysis of the substrate and release of the nitroanilide ion was followed to completion (over 10 half-lives) and found to obey simple first-order kinetics with rate constant k_{cat}/K_m x [subtilisin]. The value of k_{cat}/K_m at each pH could therefore be obtained directly from the first order plots. This was facilitated by recording the output of the spectrophotometer directly on a micro computer, allowing k_{cat}/K_m to be derived by fitting the data to a theoretical first order rate equation. k_{cat}/K_m could be determined manually from the data collected directly on the chart recorder, plots of $\ln([S_0]-[S_t])$ against t give a straight line of gradient $(-k_{cat}/K_m)[E_0]$.

In addition to satisfying conditions for first order kinetics, the use of very low substrate concentrations (8 μM $\ll K_m$) also avoids errors due to product inhibition.

Since the value of k_{cat}/K_m at any pH reflects the ratio of protonated to non-protonated forms of the enzyme, the data can be fitted directly to the Henderson-Hasselbalch equation (1) to derive the pK$_a$ of the free enzyme. The measured values of k_{cat}/K_m fit, with high precision, the ionisation curve for a single base and the pK$_a$ of the enzyme can be determined accurately.

Table 8.1   Summary of experimental conditions used for pK_a determinations.

Enzyme Buffer	pH range	No. pH's	No. assays	Ionic strength
Subtilisin BPN'				
phosphate	5.58-7.59	18	45	1.00
	6.20-7.75	18	47	0.0075
	6.26-7.77	18	54	0.005
imidazole	6.14-8.23	20	40	1.00
	6.30-7.99	20	39	0.001
Subtilisin DS99				
imidazole	6.20-8.16	20	38	1.00
Subtilisin DQ36				
imidazole	6.30-8.15	20	39	1.00
	5.89-7.94	20	47	0.001
Subtilisin KT213				
phosphate	5.58-7.61	18	48	1.0
	6.25-7.90	18	51	0.0075
	6.25-7.64	18	54	0.005
imidazole	6.00-7.88	20	54	0.001
Subtilisin ND62				
phosphate	5.58-7.60	18	52	1.00
	6.21-7.79	18	50	0.0075
imidazole	6.20-8.11	18	52	1.0
	6.01-7.92	18	46	0.001

$$pH = pK_a + \log ([base]/[acid]) \qquad \text{Equation (1)}$$

pK_as were obtained either manually from plots of k_{cat}/K_m against $k_{cat}/K_m.[H^+]$ which give a straight line of gradient $-1/K_a$ (Fersht and Renard, 1974), or by direct computer fitting to the theoretical ionisation equation (1) using the "ENZFITTER" data analysis programmes (Leatherbarrow, 1987).

At low ionic strengths electrostatic effects should be highly pronounced, since screening of the interacting charges by solvent is minimal. Therefore it is to be expected that any shifts in pK_a, resulting from the insertion or removal of ionic residues which may perturb the ionisation of His-64, will be maximal at low ionic strength. Conversely, the presence of high concentrations of dissolved ions should mask electrostatic interactions due to increased screening of the charged groups. Therefore at high ionic strength the pK_as of wild-type and mutant subtilisins are expected to converge on their intrinsic value in the absence of influence by other charges.

pK_a values were measured over a range of ionic strengths from $\mu = 0.005$ to $\mu = 1.000$ in phosphate buffer and also at $\mu = 0.001$ and $\mu = 1.000$ in imidazole buffer. The experimental conditions used in pK_a determinations for wild-type and the various mutant subtilisins constructed in this study are summarised in Table 8.1. The derived pK_a values of wild-type and mutant enzymes are shown in Table 8.2.

8.3 pK_a of wild-type subtilisin BPN'

Figure 8.1 shows the variation of k_{cat}/K_m for the hydrolysis of suAAPFpN with pH for wild-type subtilisin BPN' at low ionic strength ($\mu = 0.005$) in phosphate buffer at 25 °C. The derived pK_a is 7.06 ± 0.01. pK_a values

SUBTILISIN BPN' : pKa I=0.005 PHOSPHATE

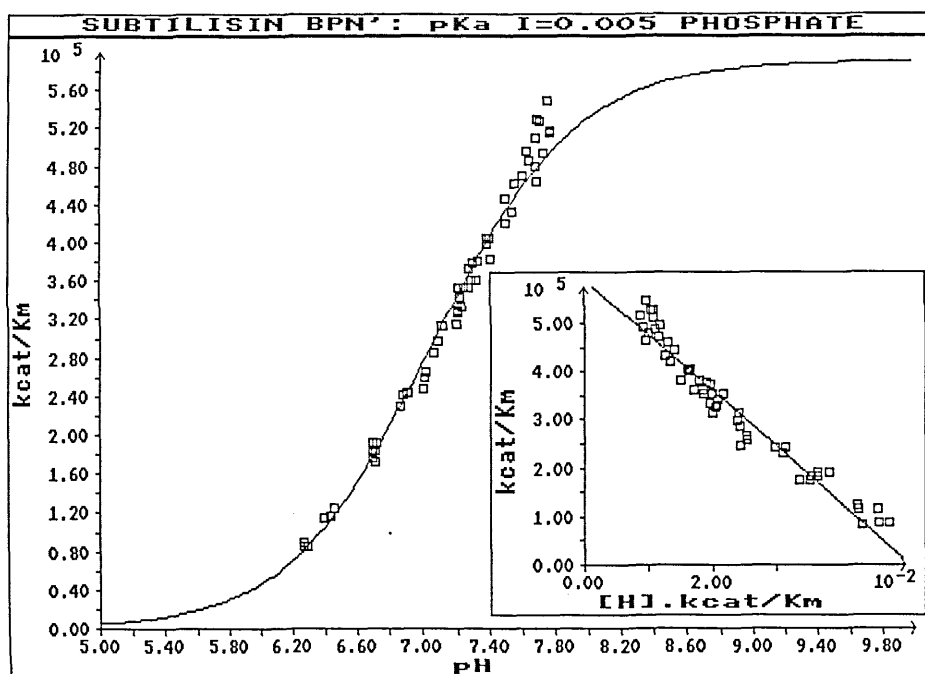**Figure 8.1:** *p*H dependence of subtilisin BPN'

Variation of k_{cat}/K_m (sec^{-1} mM^{-1}) with *p*H, for the hydrolysis of suAAPFpN by wild-type subtilisin at ionic strength 0.005 M in phosphate buffer at 25 °C.

Inset: linear transformation of data, k_{cat}/K_m plotted against k_{cat}/K_m x proton concentration ([H$^+$])

Table 8.2 pH dependence of hydrolysis of suAAPFpN by wild-type and mutant subtilisins.

Enzyme	Buffer	Ionic strength (M)	pK$_a$	ΔpK$_a$
Wild-type				
	Phosphate	1.00	7.29±0.01	-
		0.0075	7.07±0.01	-
		0.005	7.06±0.01	-
	Imidazole	1.00	7.32±0.01[1]	-
		0.001	6.95±0.01	-
DS99	Phosphate	1.00	7.20±0.01[2]	-0.09
		0.005	6.61±0.01[2]	-0.45
	Imidazole	1.00	7.25±0.01	-0.07
		0.001	6.51±0.01[2]	-0.44
DQ36	Imidazole	1.00	7.33±0.01	+0.01
		0.001	6.78±0.01	-0.17
		0.001	6.76±0.01	-0.19
KT213	Phosphate	1.00	7.28±0.01	-0.01
		0.0075	7.14±0.01	+0.07
		0.005	7.14±0.01	+0.07
	Imidazole	1.00	7.28±0.01	-0.04
		0.001	7.05±0.01	+0.10
ND62	Phosphate	1.00	7.23±0.01	-0.06
		0.0075	7.50±0.02	+0.43
	Imidazole	1.00	7.43±0.02	+0.1
		0.001	7.21±0.02	+0.26

Notes to table 8.2:
(1) T. Reinikainen, personal communication.
(2) A. J. Russell, personal communication.
All pK$_a$s determined at 25 ˙C.

derived at other ionic strengths are summarised in Table 8.2.

8.4 Effect of the mutation DS99 on pK_a

The mutation DS99 was predicted to lower the pK_a of the enzyme through simple electrostatics. At low ionic strength in phosphate buffer ($\mu = 0.001$) the derived pK_a for this mutant is 6.61 ± 0.01, which is 0.45 units lower than the wild-type value (7.06) derived under the same conditions. Increasing ionic strength should mask electrostatic interactions and the pK_as of wild-type and mutant enzymes are expected to converge. At ionic strength 1.00 in phosphate buffer containing KCl the pK_a of the mutant is 7.20 ± 0.01 and that of the wild-type enzyme is 7.29 ± 0.01, a difference of only 0.09 units. The convergence of these values indicates that the pK_a shift is due to electrostatics rather than to other effects (such as stuctural changes).

Further experiments have demonstrated that even at the low ionic strengths used here, phosphate ions are still able to screen electrostatic interactions quite effectively (A. J. Russell, personal communication). To avoid this potential problem pK_as were also determined in imidazole-imidazolium hydrochloride buffers at both low (0.001) and high (1.00) ionic strengths. Figure 8.2 shows the variation of k_{cat}/K_m with pH for subtilisin DS99 at low ionic strength in imidazole buffer. The derived pK_a is 6.51 ± 0.01, this represents a drop of 0.44 units relative to the wild-type enzyme whose pK_a under the same conditions is 6.96 ± 0.01. The pK_a shifts observed under both buffer systems are almost identical so specific buffer effects do not apply for this mutant. The experimentally determined pK_a confirms our prediction that the pK_a of the DS99 mutant would be lowered through destabilisation of the low pH protonated form of His-64.

Figure 8.2:   pH dependence of subtilisin DS99

Variation of k_{cat}/K_m (sec^{-1} M^{-1}) with pH for the hydrolysis of suAAPFpN by subtilisin DS99 at ionic strength 0.001 M in imidazole-imidazolium hydrochloride buffer at 25 ˚C.

Inset: linear transformation of data, k_{cat}/K_m plotted against

k_{cat}/K_m x proton concentration ([H$^+$])

8.5 Effect of the mutation KT213 on pK_a

At low ionic strengths in phosphate buffer (μ = 0.005 to 0.01) the pK_a of the mutant subtilisin KT213 is 7.14 ± 0.01. Figure 8.3 shows the variation of k_{cat}/K_m with pH at ionic strength μ = 0.005 graphically. The pK_a of the wild-type enzyme also at low ionic strength in the same buffer is 7.06 ± 0.01 (μ = 0.005) and 7.07 ± 0.01 (μ =0.0075). The mutation Lys → Thr at position 213 results in a small but clearly detectable increase in pK_a of about 0.08 units.

The increased pK_a is due to the removal of an electrostatic interaction (between residues 213 and 64) since the pK_a values obtained from measurements at high ionic strength converge and are identical within error. The derived values being 7.28 ± 0.01 for subtilisin KT213 and 7.29 ± 0.01 for wild-type subtilisin BPN'.

When the pK_as were measured in imidazole buffer at low ionic strength (μ = 0.001) values of 7.05 ± 0.01 and 6.95 ± 0.01 were obtained for the KT213 mutant and the wild-type enzyme respectively. For this mutant phosphate ions have only a small screening effect. When the screening effect of phosphate ions is removed the magnitude of the pK_a shift is mardinally increased to 0.1 units.

The prediction that the removal of the positive charge at position 213, by mutation of Lys-213→Thr, would stabilise the protonated, inactive form of the active site histidine and therefore result in an increased pK_a appears to have been correct.

The magnitude of the pK_a shift is smaller than that exhibited by the DS99 mutant (a decrease of 0.3 units at μ = 0.1), this probably reflects the

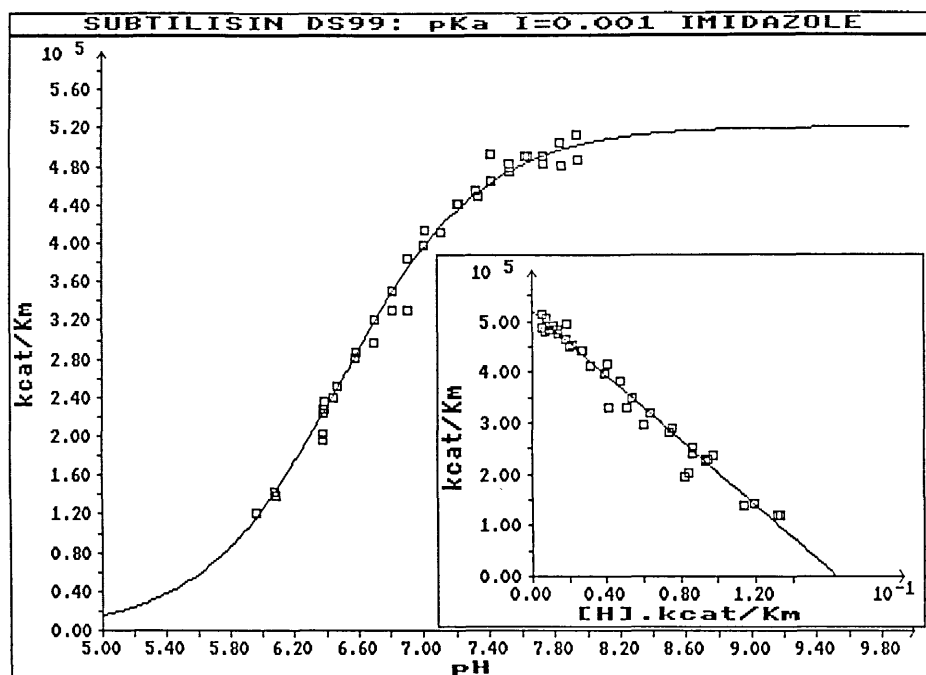**Figure 8.3:**   *p*H dependence of subtilisin KT213

Variation of k_{cat}/K_m (sec^{-1} mM^{-1}) with *p*H for the hydrolysis of suAAPFpN by subtilisin KT213 at ionic strength 0.005 M in phosphate buffer at 25 °C.

Inset: linear transformation of data, k_{cat}/K_m plotted against k_{cat}/K_m x proton concentration ([H$^+$])

greater distance separating the interacting charges in the KT213 mutant.

8.6 Effect of the mutation DQ36 on pK_a

The active-site pK_as for subtilisin DQ36 were determined in imidazole-imidazolium hydrochloride buffer at high and low ionic strengths. The derived pK_a values are summarised in Table 8.2. Figure 8.4 shows the variation of k_{cat}/K_m with pH at low ionic strength ($\mu = 0.001$) for this mutant. The high pH data (>6.5) fit the theoretical ionisation curve with high precision, however at lower pH values k_{cat}/K_m progressively drops such that the data points lie well below the expected curve. The deviation is even more apparent in the derivative plot (Figure 8.4, inset). This result is reproducible and therefore unlikely to be an experimental artefact, especially since assays performed in parallel on wild-type enzyme give a normal ionisation curve and the expected pK_a. Fitting the high pH data points from two separate assays at low ionic strength give pK_as of 6.78 ± 0.01 and 6.76 ± 0.01. Thus, the mutation has the expected effect of decreasing the pK_a of the enzyme by 0.17-0.19 units relative to the wild-type value (6.95 ± 0.01).

At ionic strength 1.00 there is no deviation of the data points from a theoretical ionisation curve. The derived pK_a for subtilisin DQ36 is 7.33 ± 0.01, which is virtually identical to the corresponding wild-type value of 7.32±0.01. The convergence of these values again confirms that the effect of the mutation on the pH dependence of the enzyme is screened by high concentrations of dissolved ions and must therefore be due to electrostatics. The reason for the deviation of the low pH data points at low ionic strength from the expected curve remains unknown.
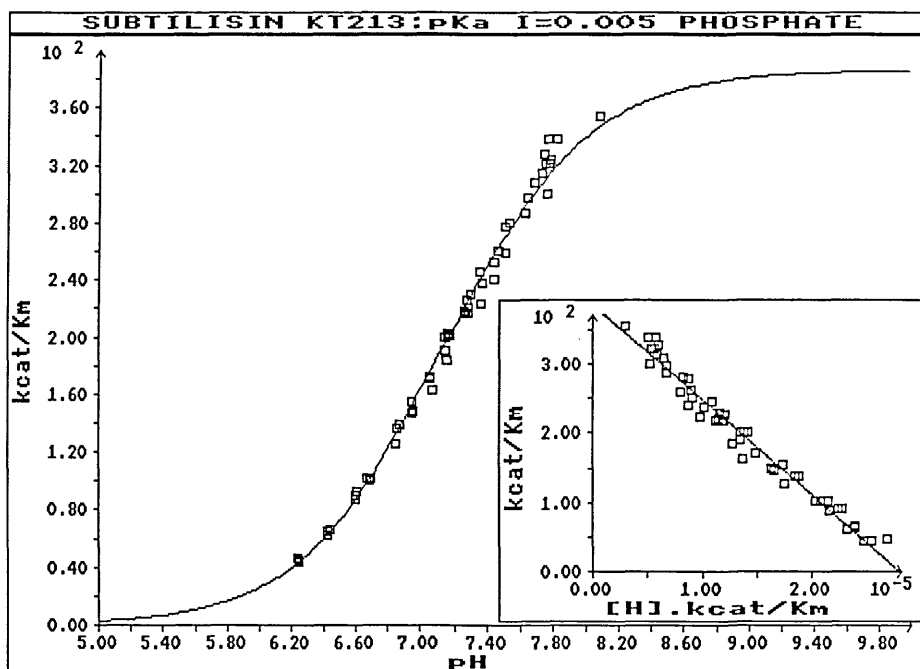
Figure 8.4: *p*H dependence of subtilisin DQ36

Variation of k_{cat}/K_m (sec^{-1}) with *p*H for the hydrolysis of suAAPFpN by subtilisin DQ36 at ionic strength 0.001 M in imidazole-imidazolium hydrochloride buffer at 25 ° C.

Inset: linear transformation of data, k_{cat}/K_m plotted against k_{cat}/K_m x proton concentration ([H$^+$])

8.7 Effect of the mutation ND62 on pK_a

The mutation Asn → Asp at position 62 also has an unexpected effect on the ionisation of His 64. Figure 8.5 shows the variation of k_{cat}/K_m with pH at low ionic strength (μ = 0.0075) in phosphate buffer for the hydrolysis of suAAPFpN by subtilisin ND62. Whilst the pK_a certainly appears to have been significantly increased, the values of k_{cat}/K_m at low pH are too high and the curve deviates from that of a single ionising base.

The data no longer fits a single ionisation curve and the derivative plot is clearly biphasic (Figure 8.5, inset). If the k_{cat}/K_m data points below pH 7 are omitted from the analysis, then the high pH data do fit a single ionisation curve with a pK_a of about 7.5, an increase of about 0.4-0.5 from the wild-type value at this ionic strength. Reducing the pH range over which the data is analysed obviously means that the derived pK_a values are somewhat less precise.

At high ionic stength in phosphate buffer, the ionisation curve for subtilisin ND62 remains anomalous, although the deviation is less pronounced. The pK_a derived using just the k_{cat}/K_m values from assays above pH 7.0, is 7.23 ±0.02. The pK_a of subtilisin-ND62 does converge with the wild-type value at high ionic strength (0.05 lower than WT), again indicating that the pK_a shift is due to electrostatics.

Similar anomalous ionisation curves were obtained when the assays were carried out in imidazole buffer. pK_a values derived from analysis of data above pH 7 are 7.21 ± 0.02 at low ionic strength (μ = 0.001) and 7.43 ± 0.02 at high ionic strength (μ = 1.0). Under these conditions the pK_a of the ND62 mutant appearsto be about 0.3 units higher than that of wild-type subtilisin. Thus, the mutation ND62 causes a significant increase
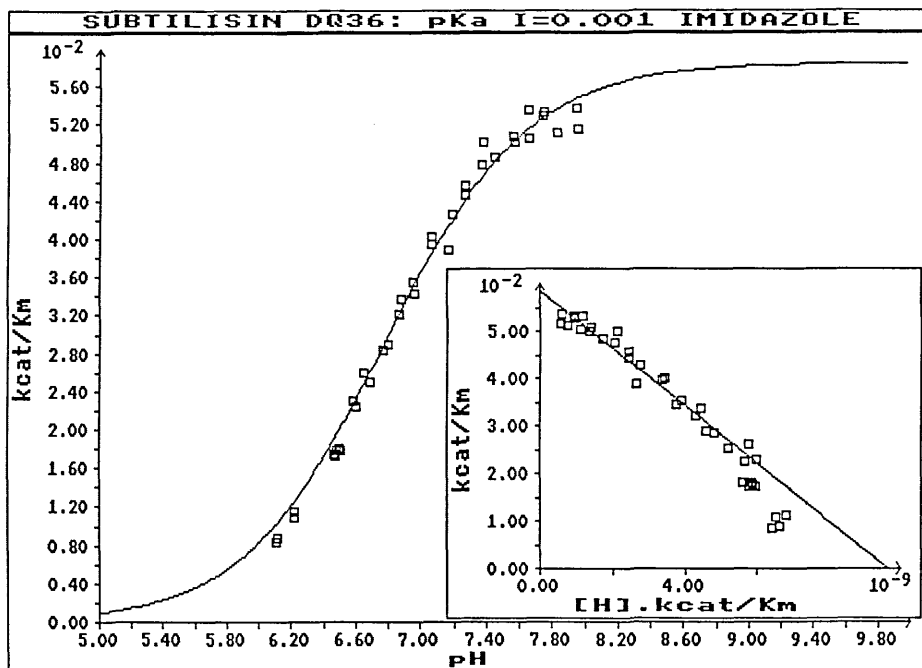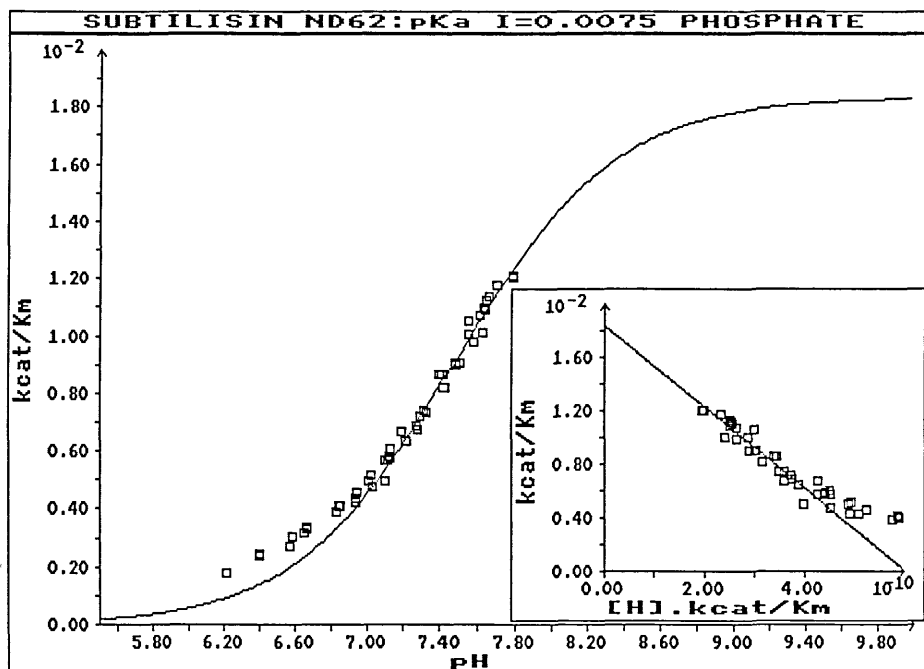
Figure 8.5: *p*H dependence of subtilisin ND62

Variation of k_{cat}/K_m (sec^{-1}) with *p*H for the hydrolysis of suAAPFpN by subtilisin ND62 at ionic strength 0.0075 M in phosphate buffer at 25 ˚C. Inset: linear transformation of data, k_{cat}/K_m plotted against k_{cat}/K_m x proton concentration ([H$^+$])

pK$_a$ of about 0.3-0.5 units at low ionic strength through the electrostatic stabilisation of the protonated inactive form of the enzyme as predicted.

8.8 Conclusions

Four variants of subtilisin BPN' have been constructed and analysed to investigate the effects of mutations which alter the charge environment of the active-site histidine. In all four cases the shifts in pK$_a$ were qualitatively as predicted from a simple model where electrostatic interactions are viewed as either stabilising the protonated form of His-64 whereupon the pK$_a$ increases, or destabilising the protonated form of His-64 resulting in a decreased pK$_a$. The mutations DS99 and DQ36 decreased the pK$_a$, whilst the mutations ND62 and KT213 had the effect of increasing the pK$_a$ of the active-site histidine.

In all cases the pK$_a$s of wild-type and mutant enzymes converge at high ionic strength, confirming that the pK$_a$ shifts are due to electrostatic effects on the ionisation of His-64. These data establish that a general principle for altering the pK$_a$s of ionising catalytic groups is by manipulation of their electrostatic environment. The magnitude of pK$_a$ shifts vary from +0.1 to about +0.5 pH units depending on the distance and presumably also the environment between the interacting groups. Even larger pK$_a$ shifts are likely to occur on introduction of multiple charge changes closer to the active site.

Although the pK$_a$ shifts were qualitatively as predicted, two of the mutations, DQ36 and ND62, gave anomalous ionisation curves. Asn-62 is conserved in all the known subtilisins and whilst the mutation was designed in the hope that the only effect would be a shift in pK$_a$, it is not really surprising that the effect is more complex. However the Asp at position 36

195

is replaced by Gln in some of the homologous subtilisins. Consequently anomalous effects on the ionisation of the active site histidine were not anticipated. These observation highlight our poor understanding of the global influence of electrostatics on enzyme function and certainly cautions against making "best guesses" when selecting targets for mutagenesis.

With subtilisin-DQ36, the deviation is only seen at low ionic strength and is masked at high ionic strength. It must be presumed therefore that the effect is due to electrostatics. Preparations of the DQ36 enzyme contained much higher levels of autolysis products than any of the other mutants constructed and it is possible that this mutant enzyme is structurally less stable at low pH. Wild-type subtilisin is known to be unstable below pH 5 (Markland and Smith, 1971; M. Bycroft, personal communication) , and it is conceivable that the mutation may have raised the pH at which the enzyme becomes unstable. It might be speculated that the reduced values of k_{cat}/K_m at low pH stems from inactivation of the enzyme.

Closer inspection of the crystal structure of the wild-type enzyme reveals that the aspartate residue introduced at position 62 by site-specific mutagenesis, is in close proximity to two pre-existing Asp residues at positions 60 and 99 (each about 6-7 Å from position 62). This region of the enzyme therefore already has a high negative charge density. Introduction of a further acidic group into this region could result in one or more of the aspartate residues exhibiting a highly perturbed pK$_a$ due to the influence of the local negative charge density stabilising the protonated, neutral form (Fersht, 1985). This phenomenon has been observed in several other enzymes (eg lysozyme and α-lytic protease) where the microenvironment of carboxylate groups can result in a highly perturbed pK$_a$ in the range 6.5-8, compared with a more typical value of between 2 and 5.5.

The values of k_{cat}/K_m for hydrolysis of suAAPFpN by subtilisin-ND62 at high pH fit a single ionisation curve quite well but the low pH data deviate away from the theoretical curve, the values of k_{cat}/K_m being anomalously high. This is exactly what might be expected if at the lower end of the pH range Asp 62 was becoming protonated. The now uncharged residue would cease to influence the ionisation of His 64 and the low pH data points would tend to lie on the wild-type curve giving the observed deviation.

The limiting values of k_{cat}/K_m in each case remain virtually unchanged on mutation. Small changes in these values are probably due to inaccuracies in measuring the concentration of active enzyme by active-site titration, and they should therefore not be interpreted as electrostatic effects on the charged transition state species.

From a practical view point the weakest part of the analysis is probably the accuracy of pH measurements. Careful calibration and rechecking of the meter and electrode during each analysis is necessary to avoid errors. The derived pK_a values are however remarkably consistent, pK_as for the same enzyme derived on separate occasions are very close, for example wild-type subtilisin in phosphate buffer at low ionic strength the derived pK_a values are 7.06 and 7.07, similarly for the mutant DQ36 at low ionic strength in imidazole derived values are 6.76 and 6.78. This allows pK_a shifts to be measured quite accurately and the experimental system for detecting pK_a shifts due to electrostatic effects is quite sensitive. The small pK_a shift of only 0.1 units was easily detected in subtilisin-KT213.

CHAPTER 9: FURTHER APPLICATIONS OF ENGINEERED SUBTILISINS

In addition to the experimental measurement of electrostatic effects on the pH dependence of enzyme catalysis, this system for producing novel mutant forms of subtilisin BPN' has also been applied to other areas including: the study of protein:protein interactions; enzyme immobilisation; the assignment of histidine resonances in the proton NMR spectrum, and also as a means of testing computer algorithms which attempt to model electrostatic effects in proteins.

9.1 Protease : protease inhibitor interactions

Chymotrypsin inhibitor 2 (CI-2) is a small protein inhibitor (83 amino acids, M_r = 9300) of serine proteases isolated from *Hiproly* barley. CI-2 belongs to the potato inhibitor-1 family of serine protease inhibitors (Svendsen *et al.*, 1980), whose general function is probably as insect anti-feedants. Protease inhibitors of this type also have important functions in the regulation of many biological processes; an improved understanding of protease:inhibitor interactions may lead to the development of novel therapeutic molecules.

CI-2 is also a potent inhibitor of subtilisin BPN' and the crystal structure of the subtilisin:CI-2 complex has been solved to high resolution (McPhalen *et al.*, 1985). The inhibitor is approximately wedge shaped with the reactive centre (Met-59 and Glu-60) located within an exposed loop at the thin end of the wedge. CI-2 binds to subtilisin such that the loop containing the active centre behaves as a typical substrate, extending into the substrate binding cleft and making contacts with the enzyme at the S and S' subsites as described earlier (1.2.4). The peptide bond between Met-59 and Glu-60 of the inhibitor is hydrolysed extremely slowly.

198

The structural gene encoding CI-2 has been cloned (Williamson *et al.*, 1987) and expressed in *E. coli* using secretion vectors (A. Campbell, in preparation). The recombinant inhibitor is identical to authentic CI-2 (A. Campbell and C. Longstaff, personal communication) and variant forms have been produced by site-specific mutagenesis of the gene cloned in M13. The factors which contribute to the strength and specificity of the interaction between subtilisin and CI-2 can now be investigated through the effects of specific amino acid replacements in either the inhibitor, the enzyme or both.

9.2 Assigning histidine resonances in the ^{1}H-NMR spectrum of subtilisin

Nuclear magnetic resonance (NMR) provides a direct physical method for measuring dissociation constants of ionising groups in proteins. It could therefore provide an independent method of corroborating the pK_a shifts measured by kinetics.

Exchange of the backbone -N-H protons in D_2O in native subtilisins results in clearly resolved histidine C2-H proton resonances in the aromatic region of the NMR spectrum (Jordan *et al.*, 1985). The pK_a of each histidine can be determined from the variation of the chemical shift with *p*H. Subtilisin BPN' contains six histidines, five of these (including His-64) are solvent accessible and one is buried, however the resonances corresponding to each histidine had not previously been assigned. A series of mutant subtilisins has been constructed in which each of the five non-catalytic histidine residues have been replaced with alanine. The mutant enzymes produced are: subtilisin-HA17; subtilisin-HA39; subtilisin-HA67; subtilisin-HA224; subtilisin-HA238, all of which retain enzyme activity. In each case replacement of a histidine results in the loss of a single resonance in the ^{1}H-NMR spectrum allowing its unequivocal assignment (M. Bycroft, in

preparation). The last remaining resonance is deductively assigned to the active site histidine, since this residue cannot be replaced without destroying enzyme activity. The pK_a as measured by the pH dependence of chemical shift for this resonance is 7.2 which is close to the pK_a derived from kinetics under similar conditions (7.1), suggesting that the assignment is correct.

Unfortunately, it has been found that subtilisin becomes insoluble and precipitates below a salt concentration of about 0.25 M when used at the high protein concentration required for NMR studies. Electrostatic effects will be considerably masked by solvent screening at this ionic strength and it has not therefore been possible to observe small pK_a shifts in single mutants at low ionic strength using NMR. Only some of the double mutants constructed in further studies (eg. subtilisin-DK99EK156) show pK_a shifts which are detectable by NMR (M. Bycroft and A. J. Russell, personal communication).

An earlier attempt to identify the resonance corresponding to His-64 by comparison of the ^{1}H-NMR spectra of subtilisins Carlsberg and BPN' (Jordan et al., 1985) was in error since the subtilisin BPN' obtained from commercial suppliers was actually subtilisin Carlsberg (Russell and Fersht, 1986).

9.3 Predicting electrostatic effects in proteins

At the outset of this study, the magnitude of electrostatic effects on the functional properties of proteins was difficult to predict. Various computer models have been developed which attempt to model electrostatic interactions in a protein based solely on the knowledge of the three dimensional structure of that particular protein. Experimental data with which to test

200

these models has however been scarce.

The data obtained in this and further studies (Russell *et al.*, 1987; Russell and Fersht, 1987) has been used to test the algorithm of Warwicker and Watson (1982). This model takes into account the shape of the molecule and the location of the interacting charges, this is an important advance on simple models since it allows for the greater solvent screening of surface charges than buried charges. Values for the dielectric constant are taken as low for the interior of the protein (3-4), high for the solvent (80) and intermediate for the region at the solvent:protein interface.

Sternberg *et al.*, (1987) have examined whether the Warwicker and Watson algorithm can model the effects of both single and double mutations on the pK_a of His-64 in subtilisin. The atomic coordinates used in the calculation were taken from the CI-2 inhibited form of the enzyme (McPhalen *et al.*, 1985). The algorithm was benchmarked by reproducing an earlier result that modelled the redox potential shift in cytochrome c_{551} on introduction of a buried propionate group as 87 mV (D_{eff} = 20) compared with an experimentally derived value of 65 mV (D_{eff} = 27).

The pK_a shifts predicted by the model were compared with the averages of experimentally determined pK_a shifts for mutant subtilisins at low ionic strength. For subtilisin-DS99, where the distance separating the interacting charges is approximately 13 Å, the calculated pK_a shift is -0.31 compared with a measured shift of -0.4. Similarly, for the mutant subtilisin-DQ36 which is approximately 15 Å from His-64 the predicted pK_a shift is -0.16 which is very close to the measured shift of -0.18. Furthermore, predicted and experimentally determined pK_a shifts for several other single and double mutant are quite close (Sternberg *et al.*, 1987). The results for mutant subtilisins constructed in this study are summarised in Table 9.1.

Table 9.1   Comparison of experimentally determined and predicted

values for effective dielectric constant

Mutant	Distance (Å)	Experimental		Calculated	
		$\Delta p K_a$ (average)	D_{eff}	$\Delta p K_a$	D_{eff}
Asp99→Ser	12.6	-0.40	48	-0.31	62
Asp36→Gln	15.1	-0.18	90	-0.16	101
Lys213→Thr	17.6	+0.08	173	+0.19	73

The only notable discrepancy is for the mutation KT213 where the calculated pK_a shift is +0.19 and the measured shift only +0.08. However in this case the distance separating the charged residues was taken as only 17.6 Å (the distance given by D. Estell was 19±4 Å, it seems likely that the actual distance separating Lys-213 and His-64 is much greater than that assumed in the calculation. Modelling with a structure that has a separation of 23 Å rather than 17.6 Å and with the charge perhaps more exposed to the solvent should yield a predicted pK_a shift which agrees better with the experimental value.

Shifts in pK_a can also be expressed in terms of an effective dielectric constant (D_{eff}) between the interacting charges using the equation

$$D_{eff} = 244/\Delta q.r.\Delta pK_a$$

where Δq is the change in charge and r is the separation in Å. Table 9.1 also shows the predicted and calculated values for D_{eff} for the mutants constructed in this study. Whilst the values of D_{eff} are quite close to that of water, these results do not refute the idea that the dielectric constant in the interior of the protein is low. It is also interesting to note that the experimental and predicted values for D_{eff} for the mutant DQ36 are both actually higher than that of water. This may indicate that the environment of Asp-36 is fairly polar.

Accurate prediction of electrostatic effects on catalytic activity and functional properties of proteins is an essential component of protein engineering. The Warwicker and Watson model can predict the magnitude of electrostatic effects on enzyme catalysis with fair accuracy and it will prove useful in the fundamental study and manipulation of the electrostatic properties of proteins.

203

9.4 Application of protein engineering to enzyme immobilisation

Many enzymes used as catalysts in industrial processes are immobilised on solid supports in order to reduce enzyme losses during processes (reveiwed by Rosevear, 1984). In many cases immobilisation also results in improved stability of the enzyme to either chemical or thermal denaturation and hence an increase in the useful life of the enzyme catalyst. Immobilised enzymes have already proved important to several industrial biocatalytic processes such as the production of high fructose corn syrup, resolving amino acid isomers and the production of penicillanic acid, a key intermediate in the production of semi-synthetic antibiotics.

An enzyme may be immobilised by a one of three means, entrapment within a permeable matrix or by direct attatchment to a solid support either by adsorption or by chemical cross linking. In some cases even intact cells containing the enzyme of interest can be immobilised.

Free α- or ϵ-amino groups are the most frequent choice for covalent attatchment of a protein to a solid matrix, but sulphydryl, hydroxyl, imidazole and free carboxyl groups may also be used. One factor reducing the general utility of enzyme immobilisation by chemical cross linking is that a proportion of the bound enzyme may be unavailable for catalysis if it is linked to the support in such a way that the active site is orientated towards the matrix and therefore sterically obstructed.

Protein engineering may prove useful as a general means for tailoring enzymes towards immobilisation methodologies in order to reduce such non-productive binding. There are two possible approaches, groups which facilitate the unwanted cross links might be removed, alternatively specific residues could be introduced at a specific position such that all of the

enzyme is immobilised in a suitable orientation for catalysis. These approaches are being tested using subtilisin BPN' as a model system.

Lysine residues close to the active site of subtilisin BPN' have been replaced with alanine by site-specific mutagenesis of the cloned gene (T. Reinikainen and J. Knowles, personal communication). The target Lys residues are at positions 136, 170 and 213 and the three single mutants, two double mutants and the triple mutant have been constructed and expressed. All the variant enzymes retain proteolytic activity. The mutations are all more than 15 Å from the active site and are fully exposed to the solvent. Based on the mutation KT213 only small pK_a shifts are predicted and the kinetic parameters will probably be unchanged.

The mutants will be immobilised on amino-silanated glass beads using glutaraldehyde via their ϵ-amino groups (Weetal, 1976) and examined to determine whether removal of one or more of the surface lysines correlates with changed immobilisation characteristics. It is anticipated that a greater proportion of the mutant enzymes will be immobilised in a suitable orientation for catalysis than wild-type subtilisin. Unwanted side reactions such as loss of enzyme activity due to protein cross linking should also be reduced.

The second approach, ie the introduction of a specific linking group, could be achieved by mutation of a surface residue distant from the active site of subtilisin to a cystine. This would allow immobilisation onto a suitable thiolated matrix by disulphide interchange. The mutations SC24 and SC87 were produced in order to introduce a disulphide bond into subtilisin and they have no effect on the kinetic properties of the enzyme (Wells and Powers, 1986). Either of these (single) mutants should be suitable for immobilisation.

The use of protein engineering to produce enzymes with improved characteristics for immobilisation may be of significant commercial importance to the food industry and in other biocatalytic processes.

CHAPTER 10: SUMMARY AND CONCLUDING REMARKS

10.1 Summary

The aim of this study was to establish an experimental system for measuring electrostatic effects in proteins. Subtilisin was selected as a suitable model system since this enzyme exhibits a marked pH activity profile which is dependent on the charge state of the active-site histidine. Furthermore, the requirements for an extended investigation by protein engineering could be met: the three-dimensional structure of subtilisin BPN' is known to high resolution; sensitive kinetic methods for analysis are available; it should be straight forward to clone, express and mutate the subtilisin gene.

The structural gene encoding subtilisin BPN' was isolated from a library of *B. amyloliquefaciens* DNA cloned in the plasmid vector pBR322. A subtilisin specific recombinant was identified by hybridisation screening, using a pair of specific oligonucleotide probes complementary to the known sequence of the subtilisin BPN' gene. Severe structural DNA instability problems were encountered during the isolation and characterisation of the subtilisin recombinant. However good fortune allowed the eventual isolation of a stable recombinant which did not appear to have undergone any rearrangements. Nucleotide sequence data confirmed that the recombinant plasmid pPT1 harboured the subtilisin gene.

The gene was recloned in the vector pUB110 and transferred to a protease deficient strain of *B. subtilis* where subtilisin was expressed at high levels. Kinetic analysis of the expressed protease also confirmed that it was subtilisin BPN'. Sufficient quantities of wild-type and mutant subtilisins could be purified easily to allow the effects of the mutations on the pH

207

dependence of the enzyme to be determined.

Four enzymatically active mutant subtilisins have been constructed by oligonucleotide-directed site-specific mutagenesis of the gene cloned in the single-stranded bacteriophage vector M13. The targets for mutagenesis were identified from knowledge of the three dimensional structure of the enzyme and were designed to alter the local electrostatic environment of the active-site histidine so changing its pK$_a$. The mutations involved insertion or removal of charged amino acid residues at the enzyme surface in the vicinity of the active site.

Two of the mutant enzymes (subtilisins KT213 and ND62) exhibited an increased pK$_a$ relative to wild-type enzyme. The other two mutants (subtilisins DS99 and DQ36) exhibited decreased pK$_a$s. The pK$_a$ shifts in all cases were qualitatively as predicted from a simple electrostatic model where the low pH, protonated form of His-64 (i.e. positively charged) is stabilised or destabilised by alterations to the charge environment, resulting in an increased or decreased pK$_a$ respectively. The pK$_a$ shifts were maximal at infinitesimal ionic strength and masked at high ionic strength confirming that the changes were due to electrostatic effects and not artefacts as a result of structural perturbation. Steady state kinetic analysis of the mutant enzymes also indicated that no gross changes in enzyme structure had occured as a result of mutation since k_{cat} for the hydrolysis of suAAPFpN remains virtually unchanged from the wild-type value. Values of K_m for this substrate are also similar to wild-type in all but one case (ND62) and here the four fold increase in K_m is probably due to direct electrostatic repulsion of the negatively charged substrate by the negative charge introduced into the enzyme.

The conclusion from this study is clear; altering the electrostatic

environment of ionisable catalytic groups in an enzyme will have a significant effect on the *p*H dependence of that enzyme. This approach may be generally useful to tailor the *p*H dependence of enymes used in industrial processes

Thus, the aim of the study has been achieved and the experimental data generated demonstrate the feasability of deliberately altering the *p*H dependence of enzyme catalysis. Furthermore, the results are being used by theoreticians to test various models of electrostatic interactions in proteins (Sternberg *et al.*, 1987). This system will continue to provide useful data to further our understanding of the influence of these long range forces in protein structure and function.

10.2 Concluding remarks

The ability to design protein molecules with predetermined functional properties *ab initio* is perhaps regarded as the "Holy Grail" for enzymologists. Clearly this goal can only be attained with a complete understanding of the principles which govern protein folding and the relationship between structure and activity. From an experimentalists viewpoint protein engineering will undoubtedly have a critical role to play in elucidating these fundamental principles.

The ability to modify pre-existing enzymes to tailor them to specific commercial or biomedical applications may be a more realistic short term goal. For the present, however, protein engineering remains a powerful tool for the functional analysis of protein molecules and will contribute significantly to our comprehension of structure and activity relationships in proteins.

Abbreviations

ATP	adenosine 5'-triphosphate
bp	base pairs
bzVGRpN	benzoyl-L-val-L-gly-L-arg *para*-nitroanilide
ccc	covalently closed circular
CM	carboxymethyl
cpm	counts per minute
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytosine 5'-triphosphate
ddATP	dideoxyadenosine 5'-trophosphate
ddCTP	dideoxycytosine 5'-triphosphate
ddGTP	dideoxyguanosine 5'-triphosphate
ddTTP	dideoxythymidine 5'-triphosphate
DEAE	diethyl amino ethyl
ds	double stranded
dGTP	deoxyguanosine 5'-triphosphate
DMSO	dimethyl sulphoxide
dNTP	deoxynucleotide 5'-triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
IPTG	β-isopropyl-thio galactoside
kb	kilobase pairs
LGT	low gelling temperature
NCI	N-*trans* cinnamoyl imidazole
PEG	polyethylene glycol
PMSF	phenyl methyl sulphonyl fluoride
rpm	revolutions per minute

SDS	sodium dodecyl sulphate
ss	single stranded
SSC	standard saline citrate
suAAPFpN	succinyl-L-ala-L-ala-L-pro-L-phe *para*-nitroanilide
TES	N-tris(hydroxymethyl)methyl-2-aminoethane sulphonic acid
Tris	2-amino-2-(hydroxymethyl) propane-1,3-diol
TTP	thymidine 5'-triphosphate
M_r	relative molecular mass
X-gal	5'-bromo,4'-chloro,3'-indolyl β-D galactoside

Predicting the T_d of oligonucleotide probes

The T_d of an oligonucleotide probe is defined as the temperature at which an oligonucleotide will be 50% dissociated from its perfectly complementary DNA sequence. Under conditions of about 0.9 M Na^+ (approximately 6xSSC) the T_d can be predicted empirically from the number of GC and AT base pairs in the duplex (Suggs *et al.*, 1982).

$$T_d = 4 \ ^\circ C \text{ per GC base pair} + 2 \ ^\circ C \text{ per AT base pair}$$

This relationship gives an accurate prediction of the hybridisation characteristics of oligonucleotides in the 11-20 nucleotide size range. The presence of one or more mis-matches will destabilise the duplex and reduce its T_d, apparently by more than the loss of the contributions of the base pairs involved. This reduction in T_d can be exploited to allow a mutagenic oligonucleotide primer to be used as a hybridisation probe capable of discriminating between a mutant DNA sequence, to which it is perfectly complementary, and the wild-type sequence with which it is mis-matched.

Annealing the probe to DNA (colony lifts, dot blots or plaque lifts) at a low temperature (15-20 $^\circ C$) where the probe hybridises to both mutant and wild-type sequences, followed by washing at incrementally increasing temperatures until the probe has washed off the wild-type sequence but remains sufficiently hybridised to the mutant sequence to give a signal detectable by autoradiography, proves to be a reliable method for detecting mutations.

A 15-17 nucleotide oligonucleotide containing a single mis-match to the

wild-type DNA sequence appears to wash off at a temperature about 5 °C below the T_d for the perfect match with the mutant sequence. It should be noted that the relationship between T_d and oligonucleotide length is non-linear, so it is advisable to use oligonucleotide probes of less than 20 bases as probes for single mis-matches. Longer probes may be used to detect more complex mutations.

REFERENCES

Alden, R. A., Wright, C. S. and Kraut, J. (1970) *Phil. Trans. Roy. Soc. Lond. B 257*, 119-124.

Anagnostopoulos, C. and Spizizen, J. (1961) *J. Bacteriol. 81*, 741.

Anfinsen, C. G. and Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205-301.

Bachmann, B. J. (1984) In *Genetic Maps 1984. A compilation of linkage maps of genetically studied organisms. Vol. 3* (ed. S. J. O'Brien) Cold Spring Harbour Laboratory, pp 145-161.

Band, L. and Henner, D. J. (1984) *DNA 3*, 17-21.

Bankier, A. T. and Barrell, B. G. (1983) In *Techniques in the Life Sciences, Biochemistry B5, Techniques in Nucleic Acids Biochemistry B508* (ed. R. A. Flavell). Elsevier Scientific Publishers, Ireland.

Bender, M. L., Begue-Canton, M. L., Blakely, R. L., Brubacher, L. J., Feder, J., Gunter, C. R., Kezdy, F. J., Killheffer, J. V., Marshall, T. H, Miller, C. G., Roeske, R. W. and Stoops, J. K. (1966) *J. Am. Chem. Soc. 88*, 5890-5914.

Bernier, R., Driguez, H. and Desrochers, M. (1983) *Gene 26*, 59-65.

Biggin, M. (1984) Ph.D. Thesis, Cambridge University.

Birnboim, H. C. and Doly, J. (1979) *Nucl. Acids Res. 7*, 1513-1523.

Blow, D. M., Birktoft, J. J. and Hartley, B. S. (1970) *Nature 221*, 337-340.

Bode, W., Papamokos, E., Musil, D., Seemueller, U. and Fritz, H. (1986) *EMBO J. 5*, 813-818.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyncker, H. L., Boyer, H. W. (1977) *Gene 2*, 95-113.

Bolivar, F. (1978) *Gene 4*, 121-136.

Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H.-Y. and Poulos, T. (1986a) *Proc. Natl. Acad. Sci. USA 83*, 3743-3745.

Bryan, P., Rollence, M., Pantoliano, M. W., Wood, J., Finzel, B., Gilliland, G. L. Howard, A. J. and Poulos, T. (1986) *PROTEINS: Structure, Function, and Genetics 1*, 326-334.

Canosi, U., Morelli, G. and Trautner, T. A. (1978) *Mol. Gen. Genet. 166*, 259-267.

Carter, P. (1986) *Biochem. J. 237*, 1-7.

Carter, P., Bedouelle, H. and Winter, G. (1985) *Nucl. Acids Res. 13*, 4431-4443.

Caruthers, M. H., Beaucage, S. L., Becker, C., Efcavitch, W. E., Fisher, E. F., Gulluppi, G., Goldman, R., deHaseth, P., Martin, F., Matteucci, M. and Stabinsky, Y. (1982) In *Genetic Engineering. Principles and Methods. 4*, (ed. J. K. Setlow, A. Hollaender), Plenum, London. pp 1-17.

Chan, V.-L. and Smith, M. (1984) *Nucl. Acids Res. 12*, 2407-2419

Clewell, D. B. (1972) *J. Bacteriol. 110*, 667-676.

Cornelis, P., Digneffe, C. and Willemot, K. (1982) *Mol. Gen. Genet. 186*, 507-511.

Courtney, M., Jallat, S., Tessier, L.-H., Benavente, A. Crystal, R. G., Lecocq, J.-P. (1985) *Nature 313*, 149-151.

DelMar, E. G., Largman, C., Brodrick, J. W. and Geokas, M. C. (1979) *Anal. Biochem. 99*, 316-320.

Docherty, A. J. P. (1980) Ph.D. Thesis, University of Bristol.

Doi, R. H. (1982) In *The Molecular Biology of the Bacilli Vol. 1* (ed. D. Dubnau). Academic Press, London. pp 72-110.

Drenth, J. and Hol, W. G. J. (1967) *J. Mol. Biol. 28*, 543-551.

Drenth, J., Hol, W. G. J., Jansonius, J. N. and Kockock, R. (1972) *Eur. J. Biochem. 26*, 811-816.

Dubnau, D. (1982) In *The Molecular Biology of the Bacilli Vol. 1* (ed. D. Dubnau), Academic Press, pp148-178.

Dubnau, D. and Davidoff-Abelson, R. (1971) *J. Mol. Biol. 56*, 209.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G. and Wells, J. A. (1986) *Science 233*, 659-663.

Estell, D. A., Graycar, T. P. and Wells, J.A. (1985) *J. Biol. Chem. 260*, 6518-6521.

Ferrari, F., Trach, K., Spence, J., Ferrari, E. and Hoch J. A. (1984) In *Genetics and Biotechnology of Bacilli* (ed. A. T. Ganesan and J. A. Hoch) Academic Press, London. pp 323-332.

Fersht, A. R. and Renard, M. (1974) *Biochemistry 13*, 1416-1426.

Fersht, A. R. (1985) In *Enzyme Structure and Mechanism*, Second Edition. W. H. Freeman, New York.

Fersht, A. R., Leatherbarrow, R. J. and Wells, T. N. C. (1986) *Trends Biochem. Sci 11*, 321-325.

Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301-304.

Fisher, S. H., Rosenkrantz, M. S. and Sonenshein, A. L. (1984) *Gene 32*, 427-438.

Fujita, Y. and Fujita, T. (1986) *Nucl. Acids Res. 14*, 1237-1252.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y. and Inouye, M. (1984) *EMBO J. 3*, 2437-2442.

Gitt, M. A., Wang, L.-F. and Doi, R. H. (1985) *J. Biol. Chem. 260*, 7178-7185.

Glover, D. M. (1984) *Gene Cloning: The mechanics of DNA manipulation.* Chapman and Hall Ltd. London and New York.

Goebel, W., Kreft, J., Bernhard, K., Schremph, H. and Weidinger, G. (1978) In *Genetic Engineering* (ed. Boyer and Nicosia) Elsevier/North Holland, pp 47-58.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. and Stormo, G. (1981) *Ann. Rev.* Microbiol. *35*, 365-403.

Gronenborn, B. and Messing, J. (1978) *Nature 272*, 375-377.

Grunstein, M. and Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA 72*, 3961-65.

Gryczan, T. J., Contente, S. and Dubnau, D. (1978) *J. Bacteriol. 134*, 318-329.

Gryczan, T. J., Contente, S. and Dubnau, D. (1980) *Molec. Gen. Genet. 177*, 459-467.

Gryczan, T. J. (1982) In *The Molecular Biology of the Bacilli Vol 1*, (ed. D. Dubnau) Academic Press, London. pp 307-326.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557-580.

Hanahan, D. (1985) In *DNA Cloning: a practical approach. Vol. 1* (ed. D. Glover) IRL Press, Oxford, p 121

Hardy, K. G. (1985) In *DNA Cloning: a practical approach. Vol. 2* (ed. D. M. Glover) IRL Press, Oxford, pp 1-17.

Harris, T. J. R. (1983) In *Genetic Engineering 4,* (ed. R. Williamson) Academic Press, London.

Hirono, S., Akagawa, H., and Itakura, Y. (1984) *J. Mol. Biol 178,* 389-413.

Hollenberg, S. M., Lai, J. S., Weickmann, J. L. and Date, T. (1984) *Anal. Biochem. 143*, 341-349.

Horton, R. D. and Lord, S. T. (1986) *Nucl. Acids Res. 14*, 5112.

Hutchison, C. A. and Edgell, M. H. (1971) *J. Virol. 8*, 181-189.

Ish-Horowicz, D. and Burke, J. F. (1981) *Nucl. Acids Res. 9*, 2989-2998.

Ishiwa, H. and Tsuchida, N. (1984) *Gene 32*, 129-134.

Itakura, K., Rossi, J. J. and Wallace, R. B. (1984) *Ann. Rev. Biochem. 53*, 323-356.

Jalanko, A., Palva, I. and Soderlund, H. (1981) *Gene 14*, 325-328.

Jordan, F., Polgar, L. and Tous, G. (1985) *Biochemistry 24,* 7711-7717.

Jones, T. A. (1978) *J. Appl Crystallog. 11*, 268-272.

Kawamura, F. and Doi, R. H. (1984) *J. Bacteriol. 160*, 442-444.

Kezdy, F. J. and Kaiser, E. T. (1970) *Meth. Enzymol. XIX,* 3-27.

Knowles, J. R. (1987) *Science 236,* 1252-1258.

Kramer, W., Schugart, K. and Fritz, H.-J. (1982) *Nucl. Acids Res. 10*, 6475-6485.

Kramer, B., Kramer, W. and Fritz, H.-J. (1984a) *Cell 38*, 879-887.

Kramer, W., Drutsa, V., Jansen, H.–W., Kramer, B., Pflugfelder, M. and

    Fritz, H.-J. (1984b) *Nucl. Acids Res. 12*, 9441-9456.

Kraut, J. (1971) In *The Enzymes Vol. 3*, (ed. P. D. Boyer)

    Academic Press, London. pp 547-560.

Kraut, J. (1977) *Ann. Rev. Biochem. 46*, 331-358.

Kreft, J., Bernhard, K. and Goebel, W. (1978) *Molec. Gen. Genet. 162*, 59-67.

Kreft, J., Burger, K. J. and Goebel, W. (1983) *Molec. Gen. Genet. 190*,

    384-389.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA. 82*, 488-492.

Lacey, R. W. and Chopra, I. (1974) *J. Med. Microbiol. 7*, 285-297

Leatherbarrow, R. J. (1987) *ENZFITTER*, Elsevier Biosoft, Hills Rd.

    Cambridge.

Leatherbarrow, R. J. and Fersht, A. R. (1986) *Protein Engineering 1*, 7-16.

Lederberg, J. (1960) *Science 131*, 269-276.

Lewin, B. (1985) *Genes*, Second Edition, J. Wiley and Sons Inc.

Linderstrom-Lang, K. and Ottesen, M. (1947) *Nature 159*, 807-808.

McPhalen, C. A. and James, M. N. G. (1987) *Biochemistry 26*, 261-269.

McPhalen, C. A., Svendson, I., Jonassen, I. and James, M. N. G. (1985)

    Proc. Natl. Acad. Sci. USA. 82, 7242-7246.

Maniatis, T., Fritsch, E. F., Sambrook, J. (1982) *Molecular Cloning: a*

    Laboratory Manual, Cold Spring Harbour Laboratory, New York

Markland, F. S. and Smith, E. L. (1971) In *The Enzymes Vol. 3*

    (ed. P. D. Boyer) Academic Press, London. pp 561-608.

Matsubara, H. Kasper, C. B., Brown, D. M. and Smith, E. L. (1965)

    J. Biol. Chem. 240, 1125-1130.

Matthew, J. B. (1985) *Ann. Rev. Biophys. Biochem. 14*, 387-417.

Matthews, B. W., Sigler, P. B. Henderson, R. and Blow, D. M. (1967)

    Nature 214, 652-656.

Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA. 74*, 560.

Maxam, A. M. and Gilbert, W. (1980) *Meth. Enzymol. 65*, 499-560.

Messing, J. (1983) *Meth. Enzymol. 101*, 20-79.

Messing, J., Crea, R. and Seeburg, P. H. (1981) *Nucl. Acids Res. 9*, 309-321.

Messing, J. and Vieira, J. (1982) *Gene 19*, 269-276.

Montell, C., Fisher, E. F., Caruthers, M. H. and Berk, A. J. (1982) *Nature 295*, 380-384.

Morihara, K. Oka, T. and Tsuzuki, H. (1969) *Biochem. Biophys. Res. Comm. 35*, 210-214.

Morinaga, Y., Franchescini, T., Inouye, S. and Inouye, M. (1984) *Biotechnology*, (July). 636-639.

Murray, C. L. and Rabinowitz, J. C. (1982) *Proc. Natl. Acad. Sci. USA. 257*, 1053-1062.

Norris, K., Norris, F., Christiansen, L. and Fiil, N. (1983) *Nucl. Acids Res. 11*, 5103-5112.

Old, R. W. and Primrose, S. B. (1985) *Principles of Gene Manipulation; An Introduction to Genetic Engineering*, Third Edition. Blackwell Scientific Publications, Oxford.

Osroff, G. R. and Pene, J. J. (1984) In, *Genetics and Biotechnology of Bacilli* (ed. A. T. Ganesan and J. A. Hoch) Academic Press, London. pp 285-295.

Ottesen, M. and Svendsen, I. (1970) *Meth. Enzymol. XIX*, 199-215.

Palva, I. Lehtovaara, P. Kaariainen, L. Sibakov, M. Cantell, K. Schein, C. H., Kashiwagi, K. and Eeissmann, C. (1983) *Gene 22*, 229-235.

Palva, I., Sarvas, M., Lehtovaara, P., Sibakov, M. and Kaariainen, L. (1982) *Proc. Natl. Acad. Sci. USA. 1979*, 5582-5586.

Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F. and Poulos, T. L. (1987) *Biochemistry 26*, 2077-2082.

Perry, L. J. and Wetzel, R. (1984) *Science 226*, 555-557.

Perutz, M. F. (1978) *Science 201*, 1187-1191.

Philipp, M. and Bender, M. L. (1983) *Mol. Cell. Biochem. 51*, 5-32.

Pielak, G. J., Mauk, A. G. and Smith, M. (1985) *Nature 313*, 152-154.

Pigott, P. J. and Hoch, J. A. (1985) *Microbiological Reviews 49*, 158-179.

Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J. and Kraut, J. (1976) *J. Biol. Chem. 251*, 1097-1093.

Power, S. D., Adams, R. M. and Wells, J. A. (1986) *Proc. Natl. Acad. Sci. USA. 83*, 3096-3100.

Priest, F. G. (1977) *Bacteriological Reviews 41*, 711-753.

Rastetter, W. H. (1983) *Trends in Biotechnol. 1*, 80-84.

Rees, D. C. (1980) *J. Mol. Biol. 173*, 323-326.

Ricchelli, F., Jori, G., Filippi, B., Boteva, R., Shopova, M. and Genov, N. (1982) *Biochem. J. 207*, 201-205.

Roberts, R. (1986) *Bionet.*

Robertus, J. D., Kraut, J., Alden, R. A. and Birktoft, J. J. (1972) *Biochemistry 11*, 4293-4303.

Rogers, N. K., Moore, G. R. and Sternberg, M. J. E. (1985) *J. Mol. Biol. 182*, 613-616.

Rogers, N. K. (1986) *Prog. Biophys. Molec. Biol. 48*, 37-66.

Rosenberg, S., Barr, P. J., Najarian, R. C. and Hallewell, R. A. (1984) *Nature 312*, 77-80.

Rosevear, A. (1984) *J. Chem. Tech. Biotechnol. 34 B*, 127-150.

Russell, A. J. and Fersht, A. R. (1986) *Nature 321*, 733.

Russell, A. J. and Fersht, A. R. (1987) *Nature 328*, 496-500.

Russell, A. J., Thomas, P. G. and Fersht, A. R. (1987) *J. Mol. Biol. 193*, 803-813.

Sanger, F. and Coulson, A.R. (1978) *FEBS Letters 87*, 107-110.

Sanger,F, Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) *J. Mol. Biol. 143*, 161-178.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA. 74*, 5463-5467.

Scheer-Abramowitz, J., Gryczan, T. J. and Dubnau, D. (1981) *Plasmid 6*, 67.

Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Comm.* 27, 157.

Shaw, W. V. (1987) *Biochem. J. 246*, 1-17.

Shine, J. and Dalgarno, L. (1975) *Nature 254*, 34-38.

Schonbaum, G., Zerner, B. and Bender, M. L. (1961) *J. Biol. Chem. 236*, 2930-2935.

Shortle, D., DiMaio, D. and Nathans, D. (1981) *Ann. Rev. Biochem. 15*, 265-294.

Sibakov, M. and Palva, I. (1984) *Eur. J. Biochem. 145*, 567-572.

Smith, E. L., Markland, F. S., Kasper, C. B., DeLange, R. J., Landon, M. and Evans, W. H. (1966) *J. Biol. Chem. 241*, 5974-5976.

Smith, M. (1983) In *Methods of DNA and RNA Sequencing*, (ed. S. M. Weissman) Praeger, New York, 23-68.

Smith, M. (1985) *Ann. Rev. Genet. 19*, 423-462.

Spizizen, J. (1958) *Proc. Natl. Acad. Sci. USA. 89*, 288-293.

Sproat, B. S. and Gait, M. J. (1984) In *Oligonucleotide Synthesis: a practical approach* (ed. M. J. Gait) IRL Press, Oxford. pp 83-115.

Stahl, M. and Ferrari, E. (1984) *J. Bacteriol. 158*, 411-418.

Stauffer, C. E. and Etson, D. (1969) *J. Biol. Chem. 244*, 5333-5338.

Stephens, M. A., Ortlepp, S. A., Ollington, J. F. and McConnell, D. J. (1984) *J. Bacteriol. 158*, 369-372.

Sternberg, M. J. E., Hayes, F. R. F., Russell, A. J., Thomas, P. G. and Fersht, A. R. (1987) *Nature 330*, 86-88.

Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J. Itakura, K. and Wallace, R. B. (1981) In *Developmental Biology Using Purified Genes*. ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 23 (ed. D. D. Brown and D. F. Fox) Academic Press, New York. pp 683-693.

Svedsen, I. (1976) *Compt. Rend. Trav. Lab. Carlsberg 41*, 237-291.

Svendsen, I., Jonasses, I. Hejgaard, J. and Boisen, S. (1980) *Carlsberg Res. Comm. 45*, 389-395.

Thomas, P. G., Russell, A. J. and Fersht, A. R. (1985) *Nature 318*, 375-376.

Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261-287.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973) *Nature New Biology 246*, 40-41.

Tsukagoshi, N. Ihara, H. and Udaka, S. (1984) *Mol. Gen. Genet. 193*, 58-63.

Twigg, A. J. and Sherratt, D. (1980) *Nature 283*, 216-218.

Uhlen, M., Flock, J.-I., and Philipson, L. (1981) *Plasmid 5*, 161-169.

Ulmer, K. M. (1983) *Science 219*, 666-670.

Valenzuela, P. and Bender, M. L. (1971) *Biochim. Biophys. Acta. 250*, 538-548.

Vasantha, N., Thompson , L. D., Rhodes, C., Banner, C., Nagle, J., and Filpula, D. (1984) *J. Bacteriol. 159*, 811-819

Vasantha, N. amd Thompson, L. D. (1986) *J. Bacteriol. V?*, 837-842

Villafranca, J. E., Howell, E. E., Voet, D. H., Strubel, M. S., Ogden, R. C., Abelson, J. N. and Kraut, J. (1983) *Science 222*, 782-788.

Vold, B. S. and Green, C. J. (1984) In *Genetics and Biotechnology of Bacilli* (Ed. A. T. Ganesan and J. A. Hoch) Academic Press, London. pp 35-50.

Volkin, D. B. and Klibanov A. M. (1987) *J. Biol. Chem. 262*, 2945-2950.

Ward, W. H. J., Jones, D. H. and Fersht, A. R. (1987) *Biochemistry 26*, 4131-4138.

Ward, W. H. J. and Fersht, A. R. (1988) *Biochemistry* (in press).

Warwicker, J. and Watson, H. C. J. (1982) *J. Mol. Biol. 157*, 671-679.

Watson, M. E. E. (1984) *Nucl. Acids Res. 12*, 5145-5164.

Weetal, H. H. (1976) *Meth. Enzymol. XLIV*, 134-148.

Weisbeek, P. J. and van de Pol, J. H. (1970) *Biochem. Biophys. Acta 224*, 328-338.

Wells, J. A., Cunningham, B. C., Graycar, T. P. and Estell, D. A. (1986) *Phil. Trans. Roy. Soc. Lond. A 317*, 415-423.

Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. and Chan, E. Y. (1983) *Nucl. Acids Res. 11*, 7911-7925.

Wells, J. A. and Powers, D. B. (1986) *J. Biol. Chem. 261*, 6564-6570.

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. and Estell, D. A. (1987) *Proc. Natl. Acad. Sci. USA 84*, 1219-1223.

Wells, J. A., Vasser, M. and Powers, D. B. (1985) *Gene 34*, 315-323.

Wikinson, A. J., Fersht, A. R., Blow, D. M., Carter, P. and Winter, G. (1984) *Nature 307*, 187-188.

Williamson, M., Forde, J., Buxton, B. and Kreis, M. (1987) *Eur. J. Biochem. 165*, 99-106.

Wong, S.-L. and Doi, R. H. (1984) *J. Biol. Chem. 259*, 9762-9767.

Wong, S.-L., Price, C. W., Goldfarb, D. S. and Doi, R. H., (1984) *Proc. Natl. Acad. Sci. 81*, 1184-1188.

Wong, S.-L. and Doi, R. H. (1986) *J. Biol. Chem. 261*, 10176-10181.

Wright, C. S., Alden, R. A. and Kraut, J. (1969) *Nature 221*, 235-242.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene 33*, 103-119.

Yang, M., Galizzi, A. and Henner, D. (1983) *Nucl. Acids Res. 11*, 237-249.

Yudkin, M. D. (1986) *Mol. Gen. Genet. 202*, 55-57.

Zoller, M. J. and Smith, M. (1983) *Meth. Enzymol. 100*, 468-500.