A technique to randomise consecutive codons in a sequence of DNA using MAX oligonucleotides.

Mohammed Ashraf

Doctor of philosophy

Aston University
January 2006

1

Aston University

A technique to randomise consecutive codons in a sequence of DNA using
MAX oligonucleotides.

Mohammed Ashraf
Doctor of Philosophy

Thesis summary

Randomisation of DNA using conventional methodology requires an excess of genes to be cloned, since with randomised codons NNN or $NN^G/_T$ 64 genes or 32 genes must be cloned to encode 20 amino acids respectively. Thus, as the number of randomised codons increases, the number of genes required to encode a full set of proteins increases exponentially.

Various methods have been developed that address the problems associated with excess of genes that occurs due to the degeneracy of the genetic code. These range from chemical methodologies to biological methods. These all involve the replacement, insertion or deletion of codon(s) rather then individual nucleotides. The biological methods are however limited to random insertion/deletion or replacement. Recent work by Hughes *et al.*, (2003) has randomised three binding residues of a zinc finger gene. The drawback with this is the fact that consecutive codons cannot undergo saturation mutagenesis. This problem has been addressed by the chemical methodologies but the synthesis of the trinucleotides coding for individual amino acids is tedious and labour intensive.

This thesis describes the development of a method of saturation mutagenesis that can be used to randomise any number of consecutive codons in a DNA strand. The method makes use of "MAX" oligonucleotides coding for each of the 20 amino acids that are ligated to a conserved sequence of DNA using T4 DNA ligase. The "MAX"oligonucleotides were synthesised in such a way, with an *Mly*I restriction site, that restriction of the oligonucleotides occurred after the three nucleotides coding for the amino acids. This use of the *Mly*I site and the restrict , purify, ligate and amplify method allows the the insertion of "MAX" codons at any position in the DNA. This methodology reduces the number of clones that are required to produce a representative library and has been demonstrated to be effective to 7 amino acid positions.

Keywords: Gene randomisation, gene libraries, saturation mutagenesis, randomised codons

2

<u>DEDICATION</u>

For my long suffering wife

Who stood by me.

younger sister, Shabnam Mahmood who was always on hand to support my wife and look after my children when my wife was ill.

Finally I would like to extend my thanks to all my friends and who are far too numerous to name individually, who have seen me at work in my restaurant and who have pushed me to keep going and complete my studies.

<u>CONTENTS</u>

6

List of Tables

List of Figures

Abbreviations

ATP	Adenosine triphosphate
Bp	Base pair(s)
BSA	Bovine serum albumin
CIP	Calf intestinal alkaline phosphatase
DMF	Dimethylformamide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
DTT	Dithioreitol
EDTA	Ethylenediaminetetra acetic acid
F-moc	Nα-9-fluoremylmethoxycarbonyl group
HCC	Hexamine cobalt chloride
IPTG	Isopropyl β-D thiogalactoside
LB	Luria broth
LMP	Low melting point
Mins	Minutes
NEB	New England Biolabs
OAS (PS)	Oligo affinity support (polystyrene)
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PEG	Polytheleneglycol
PCR	Polymerase chain reaction
PNK	Polynucleotide Kinase
RNA	Ribonucleic acid

RPM	Revolutions per minute
Sec	Second
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tm	Theoretical melting temperature
TRIS	Trishydroxymethylaminomethane
UV	Ultra violet
V/V	volume to volume
W/V	Weight to volume
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

CHAPTER 1

INTRODUCTION

Chapter 1: Introduction

1.1 Background

Mutagenesis is a powerful technique for the generation of proteins with improved properties such as improved or altered enzyme activity or specificity, enhanced binding affinity (eg. for transcriptional factors such as zinc fingers which contact DNA), folding efficiency or enhanced chemical and thermodynamic stability.

Nature has achieved many of these changes through evolution. Gradual adaptation to selectional pressures is nature's way of responding to different environmental factors. The evolutionary principle can be thought of as a repetitive cycle that introduces mutations into the genome and selects for beneficial phenotypes, subsequently passing these selected beneficial mutations on to future generations via vegetative propagation or sexual reproduction (Mossner and Pluckthun, 2001).

Biochemists have been trying to mimic this process by performing mutagenesis on genes of interest and selecting clones with improved properties and repeating this cycle over several rounds. Mutagenesis of a gene can be performed either by introducing mutations that are statistically scattered over the whole genome, by focusing them only to a particular region/gene or by focusing them to specific codons within a gene. Each approach will be addressed below.

1.2 Genome wide mutagenesis

Mutations scattered around the genome are created by the use of mutagens *in vivo*. These are divided into a number of groups: chemical mutagens, base analogue mutagens, alkylators and UV light. For example, when penicillin was first discovered the yield produced by the yeast was very low. But by mutagenising the yeast using mustard gas, UV light and other mutagens, the yield was significantly improved from a few milligrams per litre to over 20 grams per litre (Hugo and Russell, 1998).

Chemical mutagens are compounds that increase the frequency of some types of mutations. Their ability to enter the cell, their reactivity with DNA, their general toxicity and the likelihood that the type of mutation they introduce will be corrected all go towards dictating the potency of the mutagen. For example, nitrous oxide acts directly on DNA to change the bases into chemically distinct structures (Singer and Kusmierek 1982). These structures base pair in a different way to the original and change the genetic code directly.

Base analogue mutagens are chemicals that look like normal bases and as such are recognised as normal by the DNA replication system. Their essential property is that they base pair with two different bases thus making mutations because of their lack of consistency in base pairing. An example of this is 5-Bromo-deoxyuridine that can exist in a keto form (mimicking thymine) or an enol (mimicking cytosine). This slight variation in structure leads to less

accurate base pair formation than normal, causing frequent mistakes during the replication process (Singer and Kusmierek, 1982).

Alkylators are chemicals that react directly with certain bases and thus do not require active DNA synthesis in order to act but still require DNA synthesis in order to be "fixed". An example of an alkylator is methyl-nitrosoguandine that acts like the nitrous oxide and changes the bases into chemically distinct structures (changes guanine into O^6 methyl-guanine which mispairs with thymine) (Demples *et al.*, 1982). The change in base structure leads to mispairing and thus changing the DNA sequence directly. However, some modified chemical structures formed by the action of the alkylating agents cannot pair at all and sometimes the damage completely removes the bases from the DNA backbone. This leads to the breaks in the DNA sugar-phosphate backbone, which coupled with the altered bases, prevents the expression of any genes, whose coding sequence they interrupt.

Exposure to UV light generates a number of photoproducts in DNA; two different lesions that unite adjacent pyrimidines in the same strand have been most strongly correlated with mutagenesis. These lesions are the cyclobutane pyrimidine photodimer and the 6-4 photoproduct formed by the exposure of cells to UVB (280-320-nm) and UVC (200-280nm) (Pfeifer *et al.*, 2005). Several minor products such as purine dimers are also formed (Pfeifer, 1997). These lesions interfere with normal base pairing and prevent DNA replication because DNA polymerases will almost always stop if they cannot make a correct base pair. Recently a new class of DNA polymerases have been

described which unlike the replicative polymerases could synthesise past a variety of DNA lesions *in vitro* and have a very low fidelity (Goodman, 2002). These polymerases, however, cannot bypass the 6-4 photoproduct (Lehmann, 2002). This in *E. coli*, leads to the induction of the SOS system in the cell. This is an emergency response that converts the lesions that block the replication into error prone sites and restore replication and thus prevent cell death (Miller and Low, 1984). The SOS system leads to the production of proteins that assemble into a mutasome (an error-prone replication apparatus) at the lesion and allows the DNA polymerase to replicate past the lesion.

1.3 Gene-focussed randomisation strategies

The methods described above generate diversity along the length of a DNA sequence and do not necessarily affect coding sequence. Mutations in a single gene are more specific than mutations of the whole genome. The use of randomized gene libraries was first reported in the late 1980's (Reidhaar-Olson and Sauer, 1988). To make such randomised gene libraries, the gene of interest is first isolated and then sequenced. Following sequencing, mutations are introduced randomly into the coding sequence of the gene and the phenotypic consequences of these *in vitro* mutations are observed by reinserting the gene into the organism. These methods allow the simultaneous introduction of several mutations. The combination of this strategy with selection methods for the desired variants allows the rapid identification of the functional regions of a protein (Magliery and Regan, 2004). The various approaches to achieve gene-focussed randomisation are summarised below.

1.3.1 Nucleotide substitution: error-prone PCR

Random mutagenesis can be focussed to a specific gene or region, *in vitro*, by using error prone PCR (Shafikani *et al*, 1997). Error prone PCR relies on the amplification of target DNA using low fidelity DNA polymerases. This method is based on *Taq* polymerase, which lacks 3' – 5' exonuclease activity and is therefore unable to proof read copied sequences, which results in a high error rate in PCR amplified DNA (Keohavong and Thilly, 1989). The average number of errors varies from 8.9×10^{-5} errors/bp (Cariello *et al*, 1991) to 1.1×10^{-4} errors/bp (Barnes, 1992). The use of a high fidelity polymerase such as *Pfu* polymerase decreases the error rate to 1.6×10^{-6} errors/bp (Lundberg *et al*, 1991). However, even the relatively low fidelity of the *Taq* polymerase is too high to be useful for the construction of combinatorial libraries under standard amplification standards (Neylon, 2004). Varying the concentrations of individual components of the amplification reactions can be utilised to control the frequency of mutations. One of the most straightforward methods is the use of Mn^{2+} as a cofactor instead of Mg^{2+} coupled with biased concentrations of dNTPs (Cadwell and Joyce, 1994). The presence of Mn^{2+} with an over representation of dNTPs in the amplification reaction leads to error rates of approximately one nucleotide per kilobase in the final library (Cadwell and Joyce, 1994; Cirino *et al*, 2003). The error rate can be increased to one in five bases by the use of analogues of the nucleoside triphosphates (Zaccolo *et al*, 1996).

Even though the mutations are introduced at random locations, the frequency of mutation needs to be experimentally defined. Mutation frequencies that are too high will result in the accumulation of multiple mutations in a single gene resulting in a library bearing numerous mutations. The proportion of functional mutants in such a population will be low (Shafikani *et al.*, 1997). Low mutation frequencies will result in a large background of wild type genes and/or fail to identify mutations that affect the function of the protein being studied.

1.3.2 Insertion or deletion of individual codons

Rather than changing individual amino acids in a protein, an alternative approach to gene-focussed randomisation is the deletion or insertion of single codons. Again, this is a random or semi-random approach, where the changes are introduced along the length of the gene. Each mutation causes the encoded protein to be shortened or lengthen by one amino acid. Mutagenesis is followed by phenotypic screening, which allows the importance of individual amino acids in protein function to be defined rapidly. Insertion and deletions of amino acids can have profound effects on the structure and function of the protein (Grishan, 2001) (see section 1.3.3). However, Pascarella and Argos (1992) found that proteins tend to tolerate short insertions/deletions (1-5 amino acid residues) in general. Experimental approaches towards this type of codon-based mutagenesis are summarised below, with detailed mechanistic explanations of the genetic methodologies given in Appendices 1, and 2.

1.3.2.1 Chemical methodologies for codon deletion

Osuna *et al*, (2004) developed a method of deleting codons throughout a target region of a gene. The method, called codon-based random deletion is able to remove complete codons in a random and combinatorial way. The approach is a powerful automated mutagenesis method that generates deletions through the assembly of synthetic oligonucleotides. The synthesis and subsequent incorporation of the mutant oligonucleotides into structural genes permits the study of random deletions in areas of the proteins that are functionally important.

The codon-based random deletion uses normal phosphoramadite chemistry to synthesise oligonucleotides using mononucleotides protected with DMT protecting groups but arrests some of the growing oligonucleotide at the nucleotide before the codon to be deleted by reacting the solid support with a diluted solution of Fmoc-Cl. The remaining oligonucleotides that are not protected by the Fmoc-Cl would then be subjected to three cycles of ordinary synthesis to incorporate three nucleotides according to the wild type sequence. Once the three nucleotides have been incorporated, the protecting DMT and Fmoc groups are removed by treatment with acid and base and the process of mutagenesis is repeated.

The rate of deletion is controlled by the dilution of the Fmoc-Cl solution. If the Fmoc-Cl solution is dilute then very few oligonucleotide chains would be blocked and the library would contain predominantly wild type sequences or

sequences lacking only a few codons in different locations along the target area. By using more concentrated Fmoc-Cl solution, most of the growing oligonucleotide chains would be blocked, and thus more of the growing oligonucleotide strands would contain deletions and only a few wild type oligonucleotides would remain.

1.3.2.2 Random deletion of codons using genetic methods

A method of removing codons from a sequence of DNA was also developed to test the tolerance of TEM-1 β-lactamase to an amino acid deletion (Jones, 2005). The novel method introduced triplet nucleotide deletions at random positions throughout the target gene. The method exploited the properties of the mini-Mu transposon, which is mobile DNA element first discovered in maize (McClintock, 1950) that can be accurately and efficiently inserted into a target DNA sequence using the MuA transposase *in vitro* (Haapa *et al*, 1999). The mini-Mu transposon was modified at both terminal ends to incorporate the recognition sequence for the type II restriction endonuclease *Mly*I. The reaction itself has a very low target preference allowing the transposon to insert itself at any point in a gene, thus allowing deletion of trinucleotides at random (Haapa *et al.*, 1999). A detailed mechanistic description of the procedure used by Jones (2005) is given in Appendix 1.

A transposon-based method of deleting codons is a simple way of creating libraries of protein variants with single amino acid deletion at random positions. The amino acid deletion mutagenesis strategy complements existing methods of

28

probing protein structure and function. However, the strategy is not without its drawbacks. For the methodology to work, the target DNA or plasmid containing the target gene cannot contain a restriction site for *Mly*I. The recognition site for the endonuclease is only 5 base pairs in length so would occur by chance once every 1024 base pairs. This problem can be overcome by the use of silent mutations to remove the restriction sites from the target DNA as was done by Jones (2005) whose work involved removing 4 *Mly*I sites from the pUC 19 plasmid. The other problem with the methodology is that the insertion of the transposon is random throughout the gene. This random insertion means that the even though three nucleotides will be removed by the restriction with *Mly*I, the three nucleotides may not be in frame. Thus, there may also be a mutation in the codon immediately preceding the deletion.

1.3.3 Random frameshift mutations

In terms of effect on the encoded protein, the most profound of the methodologies for gene-focussed randomisation is the random insertion or deletion of a number of bases. If that number of bases is divisible by three, then one or more codons will be inserted or deleted as described above. Otherwise, the effect of such mutagenesis will be catastrophic to the encoded protein – all encoded amino acids downstream of the mutation will be changed. This drastic effect is a consequence of the fact that mRNA molecules are read in successive blocks of three nucleotides (codons). AUG, the nucleotide sequence coding for methionine, which is found at the amino terminal ends of the polypeptide chain, is known as the start signal for ribosomes to begin reading the mRNA. This

beginning of the polypeptide chain with a formyl-methionine residue aligns the ribosome, allowing the mRNA to be read in the correct reading frame (Streisenger *et al*, 1966). However the insertion/deletion of a single nucleotide (for example) will lead to a frameshift mutation resulting in the complete change in the triplets that code for the amino acids following the insertion/deletion. The experimental methodology to achieve the insertion and deletion of arbitrary number of bases into a gene is best exemplified by Murakami and co workers (2002), who have reported a general method termed random insertion/deletion (RID) mutagenesis. This enables the deletion of an arbitrary number of consecutive bases at random positions and at the same time allow insertion of a specific or random sequences into the same place. Again, a detailed mechanistic explanation of this methodology is given in Appendix 2.

1.4 Site-directed and saturation mutagenesis

The greater the degree of knowledge concerning the structure/function relationship of a particular protein, the less the need to generate random mutations. Rather, the imperative is to focus mutations to codons of key importance, such as those encoding amino acids in the active site of an enzyme. This process is known as site-directed mutagenesis. At its simplest, site-directed mutagenesis involves the substitution of one specified codon for another, by changing 1-3 bases of DNA. It can be achieved by a number of methods, of which the simplest is to prime a single stranded DNA with a mismatched, chemically synthesised oligonucleotide. The synthetic oligonucleotide anneals to the ssDNA during synthesis and is itself incorporated into the resulting

double stranded DNA (Gillam *et al.*, 1980). Many variations of this theme have been developed, to achieve the required mutations with increasing efficiency, but all rely on the same basic principle of extending a mismatched oligonucleotide primer.

However, simple site-directed mutagenesis is very labour-intensive and low-throughput. A new experiment must be undertaken to generate each new mutant. For this reason, codon randomisation (saturation mutagenesis) was introduced. Here, rather than substituting one codon with another, a codon is substituted with a "randomised" codon, which codes for all 20 amino acids. This in practice, codon randomisation, or saturation mutagenesis, leads to a mixture of genes that vary from each other only at the specified codon(s). This mixture is known as a combinatorial library. The use of combinatorial libraries has had enormous impact on protein engineering, since it permits the study of the effects of multiple amino acid substitutions within the target protein. The approach is also employed in other nucleic acid applications including antisense RNA oligonucleotide research (Patzel and Sczakiel, 2000; Hemmmings-Mieszczak *et al.*, 2003), and antisense DNA oligonucleotide research (Ho *et al.*, 1996; Lima *et al.*, 1996). Combinatorial libraries have therefore become one of the leading tools in the discovery of new biologically and pharmacologically important proteins.

1.4.1 Saturation mutagenesis using an 'NNN' codon

To generate a randomised sequence of DNA, a synthetic cassette in which the specified codons are synthesised as NNN is produced using standard phosphoramadite chemistry. This technique relies on the ability to synthesize oligonucleotide mixtures in which all four nucleotides are expected to be represented with equal frequency at a given position. Since each of the four phosphoramadites has a different efficiency of coupling to the growing strand, the generation of truly randomised DNA is unlikely. Most DNA synthesisers deliver equal volumes of each phosphoramadite to the solid support. To generate an approximately equimolar mixture of products, a biased mixture of phosphoramadites in the ratio 1.0T: 1.15G: 1.25C: 1.5A (v/v) is required (Ho *et al.*, 1996).

1.4.2 The genetic code: historical overview

Genes are composed of a sequence of four nucleotides (Adenine, Cytosine, Guanine and Thymine) and as described above, the code is formed from combinations of three of the four nucleotides (Jones and Nirenberg, 1966). The genetic code, (the 64 possible codons) is illustrated in Table 1.1. Since these 64 codons encode only 20 amino acids, the code is termed degenerate. Degeneracy is mostly found in the third nucleotide of the codon, and is termed the "wobble" effect (Crick, 1966). The genetic code was believed to be universal through all living things originally but some variations have been found recently (Knight, Freeland, and Landweber, 2001). These variations are limited to the

reassignment of one or a few codons to another amino acid(s). Only two codons (methionine and tryptophan) are not degenerate. All the other amino acids are coded for by between 2 and six different codons (Table 1.1).

Table 1.1: The genetic code (see appendix 3 for the full names of the amino acids).

UUU } phe UUC } UUA } UUG } leu	UCU } UCC } UCA } ser UCG }	UAU } tyr UAC } UAA } UAG } stop	UGU } cys UGC } UGA Stop UGG trp
CUU } CUC } CUA } leu CUG }	CCU } CCC } CCA } pro CCG }	CAU } his CAC } CAA } CAG } gln	CGU } CGC } CGA } arg CGG }
AUU } AUC } Ile AUA } AUG met	ACU } ACC } ACA } thr ACG }	AAU } asn AAC } AAA } AAG } lys	AGU } ser AGC } AGA } arg AGG }
GUU } GUC } GUA } val GUG }	GCU } GCC } GCA } ala GCG }	GAU } asp GAC } GAA } GAG } glu	GGU } GGC } GGA } gly GGG }

It has been proposed that the genetic code has evolved from an earlier version that was composed of only a few amino acids and the number of amino acids coded in the genetic code increased as the code evolved (Wong, 1975). Several hypotheses have been proposed to explain the evolution of the genetic code to

33

the form that is recognised today (Wong, 1975). One hypothesis strongly favoured is that the code developed to minimise the possibility of errors during transcription and translation (Woese, 1965; Golgberg and Wittes, 1966). The hypothesis has been tested by researchers such as Wong (1980), and Di Giulio, (1989). Recently researchers such as Freeland and Hurst (1998) have tried comparing the natural code with random codes. They concluded that very few of the random codes were better then the natural codes. It was concluded that the natural code was therefore optimised to minimise the effects of translation and transcription on the protein.

1.4.3 Effects of a degenerate genetic code on conventional saturation mutagenesis

Combinatorial cassette mutagenesis (introduction of a double-stranded section of synthetic DNA containing one or more randomised codons) can be used to explore large number of mutations in a protein, but as the number of randomised positions increases, the complexity of the gene library resulting from it becomes too large to be completely constructed and therefore subsequently screened for functional proteins. For example, saturation mutagenesis of 12 codons using NNN triplets results in a DNA complexity of 4^{36} and a protein complexity of 20^{12}. Because different codons may translate to the same amino acid, each protein sequence does not appear equiprobably in the library (Arkin and Youvan, 1992). Arginine is encoded by six codons within NNN randomisation whilst tryptophan is encoded only once. Thus, if the expression of the library is representative of the whole genetic code, there will

be a six-fold higher concentration of proteins containing arginine than methionine or tryptophan. Such concentrations are detrimental to biopanning experiments because the interactions between proteins and ligands are non-linear with respect to concentration (Hughes *et al.*, 2003). The result of this concentration bias would mean that the proteins available in high concentrations would in all likelihood saturate immobilised ligands whilst those found at low concentrations would not irrespective of their individual affinity (Hughes *et al*, 2003). Table 1.2 shows the increasing ratio of proteins containing multiple arginine residues at the randomised positions as opposed to those multiple tryptophan residues.

Table 1.2: Theoretical ratio of proteins containing multiple amino acids (Hughes, *et al*, 2003.)

Aston University

Illustration removed for copyright restrictions

The codons $NN^G/_C$ (Jamieson *et al.*, 1994; Parikh and Guengerich, 1997; Reidhaar-Olson and Sauer, 1988,) and the codons $NN^G/_T$ (Scott and Smith, 1990; Gunneriusson *et al.*; 1999, Petrenko *et al.*, 2002) would be considered intelligent "dopes" because the entire set of amino acids is still encoded but

using only half the codons. The use of the "doped" phosphoramadites would generate 31 sense codons. The phosphoramadite mix can be biased to exclude one or more of the redundant codons. The cost of biasing the phosphoramadite mix is that the exclusion of redundant codons is always at the expense of excluding other desired codons (Huang and Santi, 1994). Moreover, those strategies that obligate the cloning of redundant codons alongside those required to encode the 20 amino acids, only allow the complete exploration of a maximum of four amino acids due to the restrictions imposed by the practical transformation efficiency of 10^7 variants (Del Rio *et al.*, 1994). The technical constraints limiting the number of positions that can be randomised however can be overcome due to recent developments. These techniques involve the generation of subset libraries in which a number of codons are randomised independently of each other within each subset. The subset libraries are then combined using DNA shuffling or recombination techniques and generating more diversity then would possible with one library (Collins *et al.*, 2001; Kitamura *et al.*, 2002; Matsuura *et al.*, 2002).

The $NN^G/_C$ or $NN^G/_T$ strategy does, however, still produce 11 redundant codons and one termination codon. The 11 redundant codons will increase the size of the gene library that is generated, whilst the termination codon leads to truncated proteins. The termination of proteins increases with the number of positions randomised and ultimately prevents the achievement of all the theoretical members of a combinatorial library. The problem of truncated protein generation by the insertion of the termination codon can be overcome by

expression in an amber strain of suppressor strain of bacterial cells (Soderlind *et al*, 1995).

The use of NNN randomisation also has potential problems in expression of proteins. Protein expression hosts such as *E.coli* have a marked bias in codon usage (Kane, 1995; Kurland and Gallant, 1996). Differences in codon usage between prokaryotes and eukaryotes can have a significant impact on the heterologous protein production (Zhou *et al*, 2004). A gene with codon usage that is different from the targeted host often has poor expression of protein in that host. The presence of codons that are rare in the host organism, can lead to the ribosome stalling during protein synthesis and also leading to slow rate of synthesis of proteins and translation errors (Brinkmann *et al.*,1989; Kurland and Gallant, 1996; Roache and Sauer, 1999). The presence of rare codons can in some cases also inhibit the synthesis of proteins and thus cell growth altogether (Zhan, 1996).

Different species also have differing contents of AT and GC. Eukaryotic organisms such as *P. falciparum* often use codons that are rarely expressed in the highly expressed genes of prokaryotic organisms such as *E.coli*. (Saul and Battistutta, 1988; Nakamura *et al.*, 1999). *P. falciparum* has an AT rich genome with an AT content of nearly 80% (Zhang *et al.*, 1991). The arginine coding codons AGA (and AGG) are the least common in *E.coli*, but are relatively common occurrences in *P. falciparum* DNA, owing to the high AT bias of this organism (Kane, 1995; Hoffman *et al.*, 2002). This bias in codon usage hampers the expression of genes in different hosts. Hannig and Makrides

(1998) have proposed two alternative strategies to overcome this problem of codon bias. One strategy is to alter the rare codons in the target gene to the preferred codons of the host without affecting the encoded amino acid sequence. Another strategy is to enlarge the tRNA pool of the host by introducing a plasmid that encodes tRNAs for the codons rarely used in the host.

1.5 Elimination of redundancy from saturation mutagenesis

Ideally, saturation mutagenesis would involve cloning only 20 different codons to encode the 20 different amino acids. Moreover, it would be beneficial if those 20 codons reflected the codon preference of the organism in which the gene was to be expressed. Both chemical and biological methodologies have been developed to achieve these goals, as described below.

1.5.1 Chemical methodologies – trinucleotide phosphoramidites

Chemical methodologies centre on the use of di- and tri- nucleotide phosphoramadites. Virnekas *et al.*, (1994) were one of the first groups to work on a strategy to synthesise trinucleotides for use in combinatorial libraries. Their strategy took two factors into account: to make the route to all 20 trinucleotides as convergent as possible and to accommodate codon usage requirements for expression of proteins in *E. coli*. Rather then synthesise trinucleotides coding for the most abundant codons, Virnekas *et al.*, synthesised 20 codons that just avoided rare codons.

The 20 fully protected trinucleotides were generated *via* two protected mononucleosides and seven dinucleotides. The yield of the trinucleotides phosphoramadites was between 25-40%. The trinucleotides were tested, by synthesising various tetranucleotides in duplicate. One set was generated using conventional phosphoramadites and the other using the trinucleotide phosphoramadites to couple the trinucleotides to a support bearing thymine. The products from each set were found to behave identically. Their use in the synthesis of a pair of primers, which were subsequently used to amplify the light chain of an antibody, showed uneven distribution. This difference in distribution was attributed to the relative variations in coupling rates during synthesis. Ono *et al.,* (1995) devised a similar strategy of mutagenesis, but synthesised the anti-sense sequence using a similar principle of seven dinucleotides phosphoramadite blocks. The anti-sense sequences were then transformed into the codons using template-mediated replication.

The use of trinucleotides as synthons for the generation of oligonucleotide libraries shows some coupling bias (Kayushin *et al.,* 1996). The synthons were checked in a DNA synthesis coupling to all four phosphoramadites in test sequences from which the average coupling yield per trinucleotide was calculated. The coupling yield was used to generate a reaction factor using the factor one for the most reactive codon. This reactivity factor was used to ensure that the incorporation of the codons was random by increasing the concentrations of the trinucleotides with a high reaction factor. Thus, a trinucleotide phosphoramadite with the reactivity factor of 2 would be present

in twice as much quantity as the trinucleotide phosphoramadite with a reactivity factor of one.

Sondak and Shortle (1992) also tested the coupling efficiency of the trinucleotides. The coupling was attempted at various concentrations of trinucleotide phosphoramadites. As the trinucleotide phosphoramadite concentration was increased the coupling efficiency also increased. However, the highest concentrations of trinucleotide phosphoramadite led to a coupling efficiency of 4% as opposed to over 99% for the same concentration of the monomer phosphoramadite. The coupling time of the reactions with the trinucleotide phosphoramadites was also varied from 30 seconds to 30 minutes but this increasing and decreasing of the coupling time had no effect on the overall coupling efficiency of the reactions.

Nuener *et al*, (1998) devised a method of generating randomised codons using dinucleotide phosphoramadites as building blocks within a resin splitting procedure. The protocol uses seven dinucleotides that were sufficient to code all 20 naturally occurring amino acids. This method worked by generating a trinucleotide codon sequence with the association of one nucleoside at the 5' end to one dinucleotide at the 3' end. To achieve this, seven dinucleotides were synthesised using the methodology described by Kumar and Poonian (1984). The randomised sequence of the DNA was synthesised by loading equivalent amount of resin on four columns. Four pre-selected mixtures of dinucleotide phosphoramadites were then coupled to respective synthesis resin. Then the mononucleoside phosphoramadite corresponding to the first base of the codon

is coupled to its respective synthesis resin. With each separate pool having five different dinucleotides, the addition of each one of the 5' nucleosides to the dinucleotide phosphoramadites generates 20 differing codons each coding for a different amino acid. The four pools are then mixed and split again into four to achieve the next randomised codon. Using this split and mix strategy; the DNA can be mutagenised at the codon level rather then at the nucleotide level. This mutagenesis method has advantages over the methods based on the synthesis and use of trinucleotide synthons. The dinucleotides are obtained in better yields then trinucleotide synthons (Nuener *et al.*, 1998). The dinucleotide phosphoramadites also show a higher level of coupling efficiency when incorporated in the oligonucleotide synthesis then the trinucleotide synthons (Nuener *et al.*, 1998). The only negative aspect of this methodology is that the number of dinucleotides that need to be synthesised varies with the codons that need to be generated. Seven dinucleotides are the minimum numbers that are required to code the 20 amino acids. However, these 7 dinucleotides are not necessarily the ones that code for the most optimal codons in terms of usage for a particular host strain that the mutagenised DNA will be inserted into. The resin splitting procedure is also quite tedious and laborious as well as being difficult to automate.

Nuener *et al.*, (1998) used eleven dinucleotides to randomise an oligonucleotide at two positions using the dinucleotides strategy. The use of eleven dinucleotides was based on the basis of the codon usage in *E.coli*. Their work suggested that the 11 dinucleotide phosphoramadites were incorporated with

41

comparable frequency although some of them appeared to be either under represented or overrepresented.

The dinucleotides phosphoramadite concept was taken further by the use of Fmoc and DMT protecting groups (Lehmann, *et al.*, 1989; Gaytán *et al.*, 2002). These two protecting groups are mutually orthogonal (Gaytán *et al.*, 1998). The Fmoc protecting group is labile to bases and stable to acids whilst DMT is labile to acids and stable to bases. Gaytán *et al.*, (1998) devised a strategy of combining five Fmoc protected dinucleotides phosphoramadites and four DMT protected monomers to generate 20 trinucleotides, which coded for 18 amino acids. The growing oligonucleotide strand is contaminated with a dilute solution of Fmoc protected dinucleotides and a concentrated solution of a DMT protected mononucleotide at the same time at the point of mutagenesis. The coupling reaction will allow most chains to grow by one nucleotide by reacting with the DMT phosphoramadite and some will grow by two nucleotides by reacting with the Fmoc protected dinucleotides. The wild type oligonucleotide that has been extended by one nucleotide in the coupling reacting can undergo two further coupling reactions to complete the wild type codon. The protecting group is removed by the action of an acid. This action of the acid only affects the DMT protecting group on the monophosphoramadite and not the Fmoc protecting group. Once the wild type codon has been synthesised, then the reaction undergoes removal of the Fmoc protecting group. At this stage the orthogonol nature of the two protecting groups allows the DMT protecting group to stay on the wild type oligonucleotide and prevents its involvement in the next coupling reaction. The mixture of the four monophosphoramadites was

then coupled on to the reaction. This generates oligonucleotide strands that are predominately of the same length, some of which are wild type and some which are mutants. At this stage DMT protecting groups protect all the oligonucleotide strands.

The procedure can then be repeated as many times as is required depending on the number of codon positions being mutagenised. However, only 18 amino acids were encoded by the use of the five Fmoc protected dinucleotides and 4 DMT protected phosphoramadites. Twenty codons were generated of which 19 were sense and one was a nonsense codon. There was also one redundant codon (serine). Two of the amino acids (methionine and tryptophan) were excluded because their inclusion required the addition of two different Fmoc protected dinucleotides. However, with seven different dinucleotides, the redundancy of the codons would have increased, thereby reducing the benefits of the methodology. The reactivity of the dinucleotides was also shown to be unequal and to get a truly randomised codon, the concentrations of the dinucleotides needs to be biased in some way.

This method was modified by Yáñez et al., (2004) in order to eliminate codon redundancy and the lack of a complete set of 20 amino acids available for randomisation. The methodology was similar in principle but instead of using five Fmoc protected dinucleotides like Gaytán et al., (1998), they synthesised 20 Fmoc protected trinucleotide phosphoramadites. The mutagenic oligonucleotides were extended using the Fmoc protected trinucleotide phosphoramadites and the wild type oligonucleotides were then extended using

DMT protected mononucleotide phosphoramadites. The oligonucleotides are synthesised using normal phosphoramadite chemistry up to the point of mutagenesis. At this point in the growing strand a mixture of the Fmoc protected trinucleotides were introduced to the coupling reactions. Some of the Fmoc protected trinucleotide phosphoramadites will react with the growing oligonucleotide. A coupling reaction is then used to add a DMT protected mononucleotide phosphoramadite. The DMT protecting group was removed and the next DMT protected mononucleotide phosphoramadite was coupled on using standard phosphoramadite chemistry. This process was repeated to add the three mononucleotide phosphoramadites coding for the wild type codon. After the addition of the third mononucleotide phosphoramadite, oligonucleotides containing wild type codons protected by DMT group and mutagenised codons protected by Fmoc group are present in the reaction. The reaction is then treated with acid and base to remove the Fmoc and DMT protecting groups and the cycle can then be repeated as required. This method prevents the problems of codon redundancy and also the possibility of truncated proteins caused by termination codons. The coupling reactions can be performed manually or are suited to automation. The results correlate well with the differences in reactivity found in the use of Fmoc protected dinucleotide phosphoramadites (Gaytán et al, 2002) and in the use of DMT protected trinucleotide phosphoramadites (Virnekas et al., 1994; Lytlle et al., 1995). However, the results of the mutagenesis suggest that the distribution of the codons was uneven even though equimolar concentrations of the Fmoc protected trinucleotide phosphoramadites were used.

1.5.2 Biological methodologies – 'MAX' randomisation

MAX randomisation was developed to ensure that one amino acid was encoded using one codon, thus maintaining a 1:1 ratio of codons to amino acids. This eliminates the cloning of redundant codons and termination codons, and thus prevents the formation of truncated proteins. This elimination of the redundancy of the genetic code facilitates the generation of gene libraries in which the gene: protein ratio is maintained at 1:1 irrespective of the number of amino acids being randomised. The MAX randomisation technique also addresses the problem associated with the degeneracy of the genetic code and the bias of certain amino acids within libraries. Each amino acid is encoded only once using the MAX technique, thus all amino acids should be represented in the protein library equally.

The MAX randomisation technique that was developed by Hughes *et al.*, (2003), was based on the codons that most frequently code for the amino acid in *E. coli* proteins. It was developed to randomise the three DNA binding residues of a zinc finger gene from which the differing binding affinities for protein to DNA could be calculated. The method was based around the use of a number of oligonucleotides that were termed selectional oligonucleotides that hybridised on to a template strand. The template strand was synthesised and randomised at various points using conventional means. The template oligonucleotide was hybridised with the 20 selectional oligonucleotides that had been synthesised independently for each position randomised. Selectional oligonucleotides were 9 bp in size and as they hybridise, they seek out their complementary bases in

the template strand. The selectional oligonucleotides would seek out the complementary region of the template strand due to the conserved nucleotide sequence (made up of six nucleotides) at the 5' end of the oligonucleotide. The 3 nucleotides at the 3' end, which code for the individual amino acids will only hybridise to their complement in the randomised sequence. The use of the conserved nucleotides would act to align the oligonucleotide in the correct orientation and at the correct position. Any selectional oligonucleotide that was misaligned could not ligate to the other oligonucleotides and would not partake in the amplification. The high temperature used in the hybridisation stage ensured that any misalignment that could occur in the randomised region was prevented. Furthermore, two other oligonucleotides were hybridised that base pair with a constant region at the beginning and end of the template strand. These constant regions were designed to bind the primers and overlap the template strand to prevent the amplification of the template (Figure 1.1).

The method also is effective in using subsets of amino acids at the randomised position (Hughes *et al.*, 2003). The subsets of amino acids can be added by only using the required selectional oligonucleotides in the hybridisation and ligation reaction. However, elegant as the methodology is, its major flaw is its inability to randomise consecutive codons. With a slight variation of the methodology (see Chapter 6 for more details) the number of contiguous codons that can be randomised can be increased to two. However, that is the maximum for the methodology. The randomisation of DNA using this technology is also constrained by the number of conserved residues that are required to align the selectional oligonucleotides to the template strand. The minimum requirement

Aston University

Illustration removed for copyright restrictions

Figure 1.1: Schematic representation of MAX randomisation. Selectional hybridisation to generate a synthetic cassette for gene randomisation. Here, the template oligonucleotide contains three conventionally randomised codons. The invariant region regions of the template are colour-coded to correspond with the complementary invariant regions of the appropriate selection oligonucleotides. Two additional, unique Constant oligonucleotides are required in the hybridisation mixture to provide primer-binding and restriction sites at the end of the cassette. Primer-binding sites are indicated by broken lines. Their location ensures that only the selection strand is amplified by PCR. The resulting DNA cassette is then digested with restriction enzymes, dephosphorylated (to prevent concatemerisation) and cloned. Taken from the paper by Hughes *et al.*, (2003).

was a nine base oligonucleotide with three of those coding for the randomised amino acid. Thus only every third amino acid can be randomised although the conserved region in between the randomised section can be varied upwards of two codons.

1.6 Zinc fingers – ideal model proteins for developing randomisation techniques

Zinc fingers probably represent the class of proteins most frequently manipulated by saturation mutagenesis. Their robust structure means that manipulation of key residues has little effect on the structural integrity of the protein, meaning that combinatorial libraries, where proteins vary only at key residues can simply be screened for novel proteins with new specificities. These proteins are therefore the topic of the current study and a brief overview is given below.

1.6.1 Protein/DNA interaction of zinc fingers

Proteins that recognise DNA do so by the macromolecular recognition rule. That is, such proteins present a three dimensional shape that is complementary to the surface of the DNA sequence. When the two molecules come into close contact, the numerous atomic interactions that underlie recognition and binding can take place. Nucleotide sequence specific recognition by the protein involves a set of atomic contacts with the bases and sugar phosphate backbone. Hydrogen bonding is critical for the interaction to occur, with amino acid side

chains providing most of the critical contacts with the DNA. DNA studies have shown that most of these DNA binding proteins can be assigned to one of three classes based on their possession of one of three distinctive structural motifs: the helix turn helix (Harrison and Aggarwal, 1990), the zinc finger (Beato, 1989) and the leucine zipper (Kerppola and Curran, 1991).

The zinc finger motif, first discovered in the *Xenopus laevis*, are structural entities involved in protein DNA interactions (Miller *et al.*, 1985). Zinc fingers are involved in many aspects of eukaryotic gene regulation. Homologous zinc fingers occur in proteins induced by differentiation factors and growth signals, in proto-oncogenes, in general transcription factors, in genes that regulate *Drosophila* development and in regulatory genes of eukaryotic organisms (Klug and Rhodes, 1987).

Two-dimensional studies have confirmed that the TFIII-like zinc fingers contain an antiparrallel β-sheet and an α-helix. Two cysteine residues, which are near the turn in the β sheet region, and two histidine residues, which are in the α helix, co-ordinate a central zinc ion and hold these secondary structures together to form a compact globular form (Berg, 1988) (Figure 1.2). Most families of DNA binding proteins use α-helices to make contact with the bases in the major groove, although β-sheets can also make contact (Pavletich and Pabo, 1991). There is however no evidence to suggest that an isolated helix from any known motifs can bind DNA in a site-specific way and no regulatory system has been discovered that uses an isolated helical peptide as a site-specific DNA binding protein. In every reported complex involving α-helical motifs, it appears that the
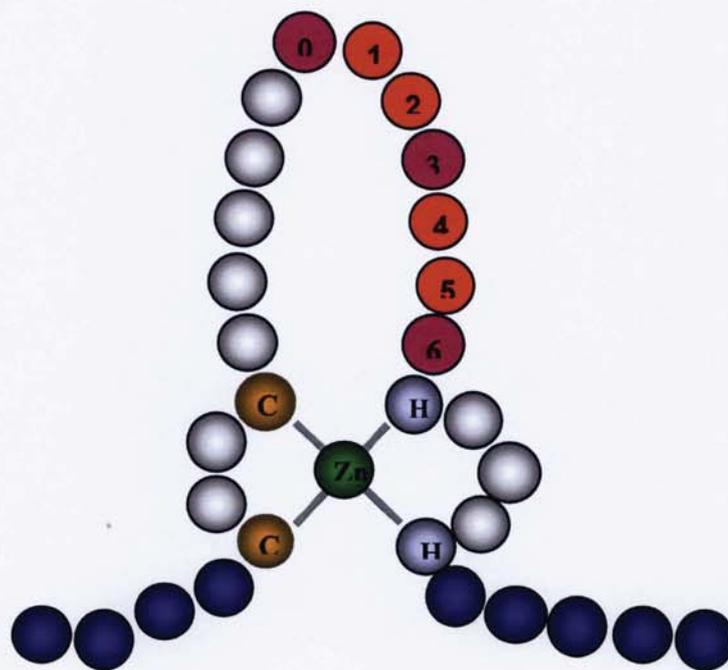
Figure 1.2: A schematic representation of a zinc finger showing the binding residues (0, 3 and 6) and the zinc coordinating residues.
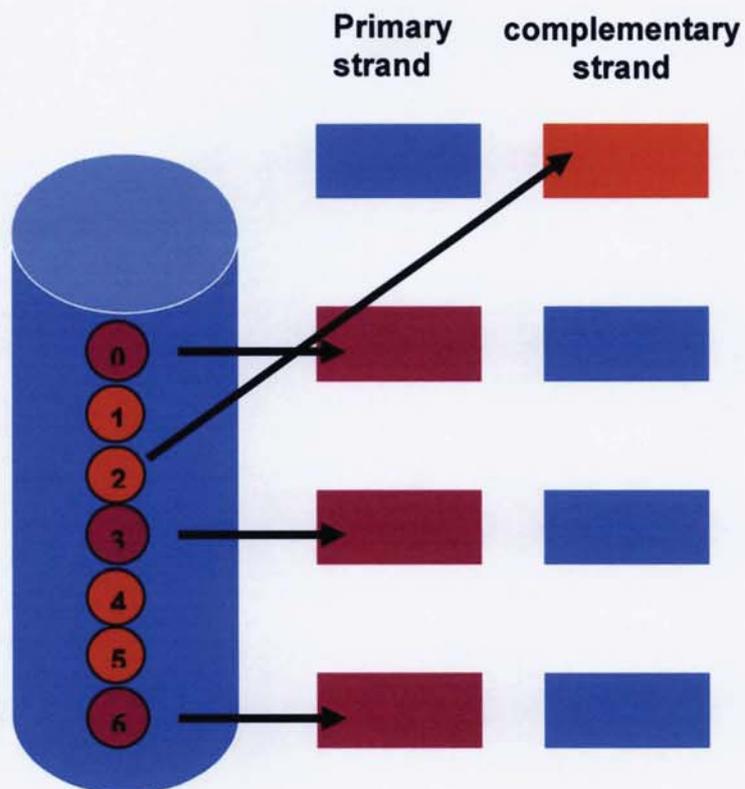
Figure 1.3: A schematic representation of the residues contacts between the zinc finger and the DNA strand.

overall binding specificity results from a set of interactions, with some contacts from a recognition helix and some from surrounding regions of the protein (Pabo and Sauer, 1992) (Figure 1.3).

Contact with the bases plays a crucial role in site-specific binding. Structural studies have shown direct hydrogen bonds between the protein side chains and the bases and occasional hydrogen bonds between the polypeptide backbone and the bases. The major groove appears to be more important for site specific recognition then the minor groove. This is probably due to the fact that the major groove is larger then the minor groove and is thus more accessible, with more potential binding sites for hydrogen bonding and hydrophobic contacts. Analysis of the pattern of hydrogen bonding sites along the edge of the base pairs also suggests that the major groove contacts provide a more reliable basis for sequence specific recognition (Seeman *et al.*, 1976). Furthermore the specific contacts are made essentially to one strand of the DNA. The more heavily contacted strand of DNA lies "anti-parallel" to that, if the protein is aligned from its amino terminus to its carboxyl terminus, of the DNA strand, which runs 3' to 5' (Berg, 1995).

1.7 Aims and objectives of this study

The aim of this study was to address the shortcomings of 'MAX' randomisation: namely to develop a non-redundant, biological method of saturation mutagenesis which could be applied to multiple, contiguous codons. The zinc finger gene described by Hughes *et al.*, (2003) was used as the basis for this study.

CHAPTER 2

MATERIALS AND METHODS

Chapter 2: Materials and methods

2.1 Media recipes

2.1.1 SOB Broth

Bacto Tryptone (Oxoid, Basingstoke),	2% w/v
Yeast Extract (Oxoid, Basingstoke),	0.5% w/v
NaCl	1 % w/v
Potassium Chloride	0.5% w/v
Magnesium Chloride	1% w/v
Magnesium Sulphate	1% w/v

2.1.2 LB Agar

Bacto tryptone (Oxoid, Basingstoke)	1% w/v
Yeast extract (Oxoid, Basingstoke)	0.5% w/v
Bactoagar	1.5% w/v
NaCl	1 % w/v

The media was sterilized by autoclaving for 20 minutes at 121°C. Selective media was prepared by the addition of ampicillin solution to a final concentration of 50 µg/ml after the media had been autoclaved and cooled to approximately 45°C. The media was then plated out into Petri dishes and left to set before use.

2.1.3 LB Broth

Bacto tryptone	1% w/v
Yeast extract	0.5% w/v
NaCl	1 % w/v

The media were sterilised by autoclaving for 20 minutes at 121°C. Selective media was prepared by the addition of filter sterilized ampicillin solution to the media after the media had been autoclaved and cooled to approximately 45°C. The final concentration of ampicillin was 50 μg/ml.

2.2 Buffer recipes

2.2.1 TBE (10 X)

Tris-borate	0.9 M
EDTA pH (8.0)	0.02 M

2.2.2 TAE (50 X)

Tris-acetate	2 M
EDTA (pH8.0)	0.05 M

2.2.3 Loading Buffers

2.2.3.1 Blue/orange loading buffer

Loading buffer was obtained from Promega (UK) and was composed of:

Orange G	0.04%
Bromophenol blue	0.03%
Xylene cynol FF	0.03%
Ficoll$^®$ 400	15%
Tris-HCl (pH 7.5)	10 mM
EDTA (pH 8.0)	50 mM

2.2.3.2 Bromophenol blue loading buffer

Tris-HCl (pH 7.6)	10 mM
Bromophenol blue	0.03%
Xylene cyanol FF	0.03%
Glycerol	60%
EDTA	60 mM

2.2.4 RFB 1 buffer

RbCl	100 mM
$MnCl_2.4H_2O$	50 mM
Potassium acetate	30 mM

CaCl$_2$.2 H$_2$O	10 mM
Glycerol	15% (w/v)
EDTA (pH 8.0)	50 mM

2.2.5 RFB 2 buffer

CaCl$_2$.2 H$_2$O	10 mM
RbCl	10 mM
MOPS	10 mM
Glycerol	15% (w/v)

2.2.6 Ligase buffer (10 X) (obtained from New England Biolabs, U.K)

Tris-HCl (pH 7.5)	500 mM
MgCl$_2$	100 mM
Dithiothreitol	100 mM
ATP	10 mM
Bovine serum albumin	250 µg/ml

2.2.7 CIP buffer (10 X) (obtained from New England Biolabs, U.K)

Calf intestinal Alkaline Phosphatase was used in NEB buffer 2 (2.2.12), buffer 3 (2.2.13) and buffer 4 (2.2.14) respectively.

2.2.8 *Taq* polymerase buffer (10 X) (obtained from Promega, UK)

KCl	500 mM
Tris-HCl (pH 9.0)	100 mM
Triton® X-100	1.0%
$MgCl_2$	1.5 mM

2.2.9 *Pfu* polymerase buffer (10 X) (obtained from Promega, UK)

Tris-HCl (pH 8.8)	200 mM
KCl	100 mM
$(NH4)_2SO_4$	100 mM
Triton® X-100	1.0%
Bovine serum albumin	1 mg/ml

2.2.10 Hybridisation buffer (10 X)

Tris-HCl (pH 7.6)	500 mM
$MgCl_2$	100 mM
PEG 8000	40% (w/v)

2.2.11 NEB buffer 1 (10 X) (obtained from New England Biolabs, UK)

Bis Tris Propane-HCl	100 mM
$MgCl_2$	100 mM

Dithiothreitol (pH 7.0)   10 mM

2.2.12 NEB buffer 2 (10 X) (obtained from New England Biolabs, UK)

Tris-HCl (pH 7.5)   100 mM

NaCl   500 mM

Dithiothreitol (pH 7.9)   10 mM

2.2.13 NEB buffer 3 (10 X) (obtained from New England Biolabs, UK)

Tris-HCl (pH 7.5)   500 mM

$MgCl_2$ 100 mM

Dithiothreitol (pH 7.9)   10 mM

2.2.14 NEB buffer 4 (10 X) (obtained from New England Biolabs, UK)

Tris-acetate   200 mM

Magnesium acetate   100 mM

Potassium acetate   500 mM

Dithiothreitol (pH 7.9)   10 mM

2.2.15 T4 RNA ligase buffer (10 X) (obtained from New England Biolabs, UK)

Tris-HCl (pH 7.8)   500 mM

MgCl$_2$	100 mM
Dithiothreitol	10 mM
ATP	10 mM

2.2.16 T4 PNK buffer (10 X) (obtained from New England Biolabs, UK)

Tris-HCl (pH 7.6)	700 mM
MgCl$_2$	100 mM
Dithiothreitol	50 mM

2.2.17 Single stranded ligation buffer (10 X)

MnCl$_2$	50 mM
HCC	10 mM
ATP	100 μM
MgCl$_2$	100 μM
Bovine serum albumin	100 μg/ml

2.2.18 Single stranded ligase buffer without MnCl$_2$ (10 X)

HCC	10 μM
ATP (100mM)	100 μM
MgCl$_2$ (500mM)	100 μM
Bovine serum albumin	100 μg/ml

2.3 Miscellanous solutions

2.3.1 Sodium Acetate

Sodium acetate was made by dissolving 408.1 g sodium acetate.$3H_2O$ in 800ml of double distilled water. The pH was adjusted to 5.2 with glacial acetic acid and the volume was made up to 1000 mls. The solution was sterilized at $121^{\circ}C$ and stored at room temperature.

2.3.2 DTT (Dithiothreitol)

DTT solution was made by dissolving 3.09g DTT in 20ml of 0.01M sodium acetate (pH 5.2) and filter sterilized through a 0.22 micron filter (Nalgene, U.K).

2.3.3 X-Gal (5- Bromo-4-Chloro-3 indolyl-β-D-galactopyranoside)

X-gal (5- Bromo-4-Chloro-3 indolyl-β-D-galactopyranoside) was purchased from Promega (UK) as a 50mg/ml solution in dimethyl formamide (DMF) and diluted to 20mg/ml in DMF. The solution was stored at $-20^{\circ}C$.

2.3.4 Ampicillin solution

Stock solutions were prepared using ampicillin sodium salt and double distilled water. Solutions were sterilised using a 0.2μm syringe filter (Nalgene, U.K) according to manufacturers' instructions.

2.4 Oligonucleotides

2.4.1 Synthesis of oligonucleotides

Commercial oligonucleotides were obtained from MWG-Biotech AG (Anzingerstr, Germany) or Biomers.Net (Sedanstr, Germany). Where in-house production was required, oligonucleotides were synthesised on a Beckman 1000 DNA synthesizer. The synthesis was performed on standard 30nM scale columns supplied by Beckman or else on 1 mg of the oligo affinity support (OAS-PS), which had been placed in an empty 30nM scale column. The synthesis was performed using standard phosphoroamadite chemistry (phosphoramidites were purchased from Glen Research). For randomised bases, a fifth mix was made up by combining 1.0 thymine: 1.15 guanine: 1.25 cytosine: 1.5 adenine (v/v). The variation in concentration of the four bases is to equilibrate the variation in the coupling efficiency of each base (Ho *et al* 1996).

2.4.2 Deprotection of oligonucleotides

In-house synthesized oligonucleotides were cleaved and deprotected using a kit (Beckman) according to manufacturers' instructions and freeze dried using a SpeedVac 100 (Savant).

Oligonucleotides synthesised on the oligonucleotide affinity support (OAS-PS) was deprotected by leaving in concentrated aqueous ammonia for 8 hours. The liquid was decanted and the support washed with water until a neutral pH was achieved. The support was air-dried. The oligonucleotide was not cleaved away from the Oligo affinity support.

2.5 Hybridisation reactions

2.5.1 Hybridisation and ligation of trinucleotides on to on-bead templates

On-bead hybridisation reactions were set up using equal amounts of template and the two hybridisation oligonucleotides (200 picomoles unless otherwise stated), 1 X hybridisation buffer (2.2.10) and 16.5 picomoles each of the 5' phosphorylated "MAX" trinucleotides. The final volume of the reaction was 50μl. This reaction was heated to 95°C and held at that temperature for 2 minutes. The temperature was decreased by 1 degree per minute until 4°C was reached (PTC-0200 programmable thermal controller MJ Research). Then 1 unit of DNA ligase (Gibco) was added along with 1μl of 100 mM ATP and 1μl of 50 mM DTT and left to incubate overnight at 4°C. The reaction was washed

by adding 450 µl of water, vortexing, centrifuging the reaction for 1 minute and then discarding 450 µl of water. This process was repeated 3 times. The reaction was then heated to 95°C for 10 minutes and 45µl of the supernatant removed from the top. The supernatant was used in subsequent amplification reactions.

2.5.2 Hybridisation and ligation of trinucleotides on to standard oligonucleotide templates

Hybridisation reactions consisted of 200 picomoles of the template oligonucleotide, 200 picomoles of each of the complementary outer oligonucleotidedes, 5 picomoles of each of the 5' phosphorylated "MAX" trinucleotides and 1 X NEB ligase buffer in a final volume of 49 µl. The reaction temperature was raised to 95°C for two minutes and then decreased by 1°C per minute until 4°C (PTC-0200 programmable thermal controller, MJ Research). Twenty units of T4 DNA ligase (NEB) were added and the reaction incubated overnight at 4°C.

2.5.3 General oligonucleotide hybridisation

Hybridisation of oligonucleotides was performed by combining equal quantities of complementary oligonucleotides and heating the resultant mix to 95°C for two minutes. The temperature was decreased by 1°C per minute until 4°C was reached (PTC-0200 programmable thermal controller, MJ Research). The hybridised oligonucleotides were then stored at -20°C until required.

2.5.4 Hybridisation and ligation of MAX oligonucleotides on a standard template

Hybridisation of max oligonucleotides was performed by combining 200 picomoles each of template 2, Max1, Max2, MA9 and MA10 oligonucleotides (Figure 6.1) in 1 X NEB ligase buffer in a final volume of 49 μl. The reaction was heated to 95°C for two minutes and allowed to decrease by 1°C per minute until 4°C (PTC-0200 programmable thermal controller, MJ Research). 20 units of T4 DNA ligase (NEB) were then added and the reaction incubated overnight at 4°C.

The second round of hybridisation to add a third MAX position was performed by combining 200 picomoles each of template 3, Max2, MA10 and the 34 bp band of DNA (extracted from PAGE) in 1 X NEB ligase buffer in a final volume of 49 μl (Figure 6.6). The reaction was heated to 95°C for two minutes and allowed to decrease by 1°C per minute until 4°C (PTC-0200 programmable thermal controller, MJ Research). 20 units of T4 DNA ligase (NEB) were then added and the reaction incubated overnight at 4°C.

2.6 Amplifcation of DNA

2.6.1 Polymerase Chain Reaction (PCR)

PCR reactions were made up to a final volume of 100 μl in 200 μl microfuge tubes. Reactions contained one unit of *Pfu* DNA polymerase (Promega), 50

picomoles of each of the forward and reverse PCR primers, 200 µM dNTPs (Amersham Biosciences, Amersham), 1 X *Pfu* polymerase buffer (2.2.9), 0.1 ng of template DNA and double distilled water to make the volume up to 100 µl. The PCR reactions were carried out in a PTC-200 thermal cycler (MJ Research Watertown, MA, USA), employing a standard cycle of 35 cycles at the temperatures listed unless otherwise stated.

92°C for 30 seconds

48°C for 30 seconds

72°C for 60 seconds

Each PCR reaction was started by heating the reaction to 95°C for 2 minutes to denature the oligonucleotides. The final cycle was followed by the reduction of reaction temperature to 4°C. The amplified DNA was then stored at -20°C until required.

2.6.2 Polymerase chain reaction (colony PCR)

Colony screening was carried out by direct amplification of the colonies. A selected colony was picked and touched onto the side of 200 µl PCR reaction tube. A reaction mix was made up and added to the PCR reaction tube. The reaction mix was composed of 1 X *Taq* buffer (2.2.8), 1 unit *Taq* polymerase (Promega), 50 picomoles pUC 19 reverse PCR primer (Appendix 4), 50 picomoles pUC 19 forward PCR primer (appendix 4), 1.5mM $MgCl_2$, 200 µM dNTPs (Amersham Biosciences, Amersham), and double distilled water to 100 µl. The reactions were heated to 95°C for 2 minutes in a PTC-200 thermal

cycler (MJ research Watertown, MA, USA), and the PCR reactions carried out employing the same parameters as Polymerase chain reaction (2.6.1).

2.6.3 Gradient Polymerase Chain Reaction

PCR reactions were made up to a final volume of 130 μl. The reaction consisted of 70 picomoles of each of the two PCR primers, 3.3 units of *Pfu* polymerase (Promega, USA), 200 μM dNTP's (Amersham Biosciences, Amersham), 1 X *Pfu* polymerase buffer (2.2.9), appropriate amount of template and sterile double distilled water to make up the reaction to 130 μl. 10 μl of the PCR mix was aliquoted into twelve 200 μl microfuge tubes. The PCR reactions were carried out in a PTC-200 thermal cycler (MJ Research Watertown, MA, USA), employing a standard cycle of 35 cycles at the temperatures listed unless otherwise stated.

92°C for 30 seconds

$40\text{-}60^{\circ}$C for 30 seconds

72°C for 60 seconds

Each PCR reaction was started by heating the reaction to 95°C for 2 minutes to denature the oligonucleotides. The final cycle was followed by the reduction of reaction temperature to 4°C. The amplified DNA was visualised on agarose gel (2.7.1)

2.7 Visualisation of DNA

2.7.1 Agarose gel electrophoresis

The electrophoresis apparatus was set up according to manufacturer's instructions. A 1%, 2% or 3% agarose gel was prepared by dissolving molecular biology grade agarose (Gibco, Paisley or Flowgen, Rockland) in 50 ml of 1 X TAE buffer (2.2.2). Ethidium bromide was added to a final concentration of 0.5 μg/ml and the gel allowed to set. Once set it was placed in the electrophoresis tank filled with 1 X TAE buffer (2.2.2) and ethidium bromide to a final concentration of 0.5 μg/ml. The comb was removed from the gel and the DNA mixed with loading buffer (2.2.3) was added to the well formed by the comb. A current of 10V/cm was applied for approximately one hour or until the dye front had migrated to an appropriate distance. The gel was then removed, observed under UV light and photographed using a UVP transilluminator (UVP Products UK), or GeneSnap transilluminator photographic systems, (Genesnap UK).

2.7.2 Denaturing polyacrylamide gel electrophoresis

Denaturing page gels were prepared by dissolving 21g of urea in 12.5 ml of 40% acrylamide (Severn Biotech Ltd. UK) (19:1 acrylamide: bisacrylamide ratio), 5ml TBE (10 x) (2.2.1) and water to 50 ml. This was stirred vigorously until the urea had dissolved. Then 600μl of 10% ammonium persulphate (2.3.4) was added followed by 80μl of TEMED (N, N, N', N', tetraethylmethylethylene diamine).

The electrophoresis apparatus was set up as per manufacturers' instruction and the gel poured and set. The samples were then loaded onto the gel and a voltage of 10 v/cm applied for approximately 3 hours. The gel was then removed from the glass plates and placed into 1 X TBE buffer (2.2.1) with ethidium bromide (2.3.3) at a concentration of 2 µg/ml. This was then placed on a shaker at 150 rpm and left for five minutes. The gel was then removed from the buffer and visualised using the UVP transilluminator (UVP products, U.K, or Genesnap transilluminator photographic systems, Genesnap, U.K).

2.7.3 Polyacrylamide gel electophoresis

Polyacrylamide gels were prepared by making up 12.5mls of 40% acrylamide (Severn Biotech Ltd. UK) (19:1 acrylamide: bisacrylamide ratio, 5 ml TBE (10 X) (2.2.1) and water to 50 ml. This was stirred vigorously to mix all the reactants. Then 600 µl of 10% ammonium persulphate (2.3.4) was added followed by 80 µl of TEMED (N, N, N', N', tetraethylmethylethylene diamine).

The electrophoresis apparatus was set up as per manufacturer's instructions and the gel poured and set. The samples were then loaded onto the gel and a voltage of 10 v/cm applied for approximately 3 hours. The gel was then removed from the glass plates and placed into 1 X TBE buffer (2.2.1) with ethidium bromide (2.3.3) at a concentration of 2 µg/ml. This was then placed on a shaker at 150 rpm and left for five minutes. The gel was then removed from the buffer and

visualised using the UVP transilluminator (UVP products, UK, or Genesnap transilluminator photographic systems, Genesnap, UK).

2.8 Gel purification of DNA

2.8.1 Agarose gel purification of DNA (LMP agarose gel)

Agarose gels were prepared using Seaplaque low melting point agarose (Flowgen) according to manufacturers' instructions in 1 X TAE buffer (2.2.2). Ethidium bromide was added to a final concentration of 0.5µg/ml and the gel allowed to set. Once set it was placed in the electrophoresis tank filled with 1 X TAE buffer (2.2.2) and ethidium bromide to a final concentration of 0.5 µg/ml. The comb was removed from the gel and the DNA mixed with loading buffer (2.2.3) was added to the well formed by the comb. A current of 10 V/cm was applied for approximately one hour or until the dye front had migrated to an appropriate distance. The gel was then removed, observed under UV light and photographed using a UVP transilluminator (UVP Products UK), or GeneSnap transilluminator photographic systems, (Genesnap UK).

The required DNA band was excised from the LMP gel and placed into a Promega wizard miniprep column. The Promega wizard miniprep column was then placed into a clean microfuge tube and centrifuged at 6000 rpm for 2 minutes to elute the DNA. The Promega wizard miniprep column was then discarded and the microfuge tube with the eluted DNA was stored at -20°C.

70

2.8.2 Elution of DNA from polyacrylamide gel

The DNA band of interest was visualised under U.V light and excised using a sharp scalpel. The excess acrylamide gel was trimmed and the gel slice transferred to into an Elutatube™ vial. The vial was filled with 1 X TBE (2.2.1) buffer and the lid replaced, ensuring that no air bubbles were formed. The vial was placed in the tray provided according to manufacturer's instructions ensuring that the two windows were perpendicular to the direction of the electric current. 1 X TBE (2.2.1) buffer was added to the electrophoresis gel tank and the vial was immersed in it. A voltage of 8 V/cm was applied for 30 minutes, or until the DNA had eluted out of the gel. The polarity of the current was reversed for 30 seconds and the vial removed from the running buffer. The Elutatube™ was opened and the solution transferred to a clean microfuge tube. The microfuge tube was centrifuged at 14000 rpm for one minute to remove any residual polyacrylamide gel and the solution transferred to a clean microfuge tube. The DNA was concentrated using pellet paint precipitation (2.11) or ethanol precipitation (2.11).

2.9 Quantification of DNA

2.9.1 Quantification of DNA using UV light

DNA was quantified by measuring the UV absorbance at a wavelength of 260nm (Cecil CE3041 spectrometer). The measurement given by the spectrometer is in optical density (O.D) units. An optical unit is defined as the

amount of a substance dissolved in 1ml that will give an absorbance reading of 1.00 in a spectrometer with a 1cm path. Because the average absorbance maximum for the bases is approximately 260nm, optical density for oligo for DNA is commonly reported for this wavelength. The concentration of DNA in a solution after UV absorbance readings have been taken can be determined by the following equation:

Beers law: $A260 = Ecb$

Where b = optical path length in cm

C = concentration in μg/ml

E = extinction coefficient, specifically:

A_{260} 1 = 33μg for single stranded DNA (assuming 1cm optical path length)

A_{260} 1 = 50μg for double stranded DNA (assuming 1cm optical path length).

(Sambrook *et al* 1989).

2.9.2 Quantification of DNA using agarose gel electophoresis

DNA was quantified by comparison with quantitative DNA molecular weight markers using agarose gel electrophoresis (2.7.1). Comparison of relative band intensities was performed using Phoretix ID Advanced Gel Analysis program (Phoretix International, UK).

2.10 Phenol chloroform extraction of DNA

DNA was extracted by adding an equal amount of phenol: chloroform: isoamyl alcohol at a ratio of 25:24:1. The resultant mixture was vortexed and then centrifuged at 14000 rpm in a microfuge for one minute. The upper aqueous phase was removed and the process repeated until no protein was visible at the interface between the organic and inorganic phases. The recovered aqueous phase was mixed with an equal volume of chloroform, vortexed and centrifuged at 14000rpm in an eppendorf microfuge. The upper aqueous phase was removed and the DNA recovered by ethanol precipitation (2.11) before resuspension in an appropriate buffer. The resuspended DNA was stored at -20°C.

2.11 Ethanol precipitation of DNA

DNA was precipitated by the addition of 0.1 volume of 3M sodium acetate (pH 5.5) (2.3.5) and 2 volumes of ice-cold 100% ethanol. The resultant mix was vortexed and placed at -20°C overnight or at -70°C for one hour. The DNA was pelleted by centrifugation at 14000 rpm for 20 minutes in a microfuge. The pellet was dried by careful removal of the liquid. The pelleted DNA was washed in 70% ethanol and the resultant mix centrifuged at 14000 rpm for 10 minutes. The liquid was removed carefully and the pellet allowed to air dry before being resuspended in an appropriate buffer. The DNA was the stored at -20°C. Where required, 2 µl pellet paint (Novogen, UK) was added to the DNA alongside the ethanol, to aid subsequent visualisation of the DNA. The addition of pellet paint (Novagen, UK) alleviates the incubation at the low temperatures.

The precipitation can then be carried out after 5 minutes of incubation at room temperature.

2.12 Enzyme based reactions

2.12.1 Phosphorylation of oligonucleotides

Phosphorylation reactions were set up in a final volume of 50µl unless otherwise specified. The reactions consisted of 1 X Ligase buffer (2.2.6), 10 units T4 PNK (NEB, Hertfordshire), an appropriate amount of DNA and water to a final volume of 50µl (unless specified). The reaction was incubated at 37°C for a length of time varying from one hour to overnight. Alternatively the 1 X ligase buffer (2.2.6) was substituted with 1 X PNK buffer (2.2.15) and supplemented with 1 µM ATP.

2.12.2 Ligations

2.12.2.1 Blunt ended ligation of DNA into pUC 19

Ligation reactions were set up containing 1 X ligase buffer (2.2.6), 20 units T4 DNA ligase, (NEB, Hertfordshire) 10ng phosphatased, *Sma*I restricted pUC19 (MBI Fermantas, Lithuania) and 0.5 picomoles of the insert oligonucleotides. The final volume of the ligation reaction was 20 µl. The ligations were incubated overnight at 4°C unless otherwise stated.

2.12.2.2 Blunt ended ligations of double stranded oligonucleotides

Ligations were set up containing 1 X ligase buffer (2.2.6), 20 units of T4 DNA Ligase (NEB, Hertfordshire), 0.1 to 1 μM of kinased 5' termini, and water to a final volume of 20 μl. Reactions were incubated at 16°C overnight.

2.12.2.3 Single stranded ligations 1

Initial single strand DNA ligation were prepared with 1X RNA ligation buffer (2.2.15), 100picomoles of phosphorylated oligonucleotide, 100picomoles of a second non phosphorylated oligonucleotide and 20 units of RNA ligase (NEB, Hertfordshire) in a 100 μl reaction.

2.12.2.4 Single stranded ligations 2

Single stranded DNA ligations were optimised and prepared with 1 X RNA ligation buffer, (2.2.15), 20 picomoles phosphorylated oligonucleotide, 20 picomoles of a second non phosphorylated oligonucleotide, 20% PEG, and 1 unit of RNA ligase (NEB, Hertfordshire) in a 20 μl reaction.

2.12.3 Restriction digests of DNA

Restriction digests were set up in 1 X appropriate buffer according to manufacturers' instructions, using approximately 3 units of restriction enzyme per ug of DNA in 20 μl final volumes. The restriction digests were incubated at

37°C or at the optimal temperature for the enzyme for two hours. The temperature was then raised to 70°C for twenty minutes to denature the enzyme or where appropriate reaction was extracted with phenol/chloroform (2.10). Double digests were set up in the same manner but using a suitable buffer to allow the best functioning of the two enzymes (according to manufacturer's information).

2.12.4 Removal of 5'phosphate groups from DNA

Phosphatase reactions were set up in a final volume of 20 µl unless otherwise specified. The reactions consisted of 1 X NEB buffer 3 (2.2.13), 1 units CIP (NEB, Hertfordshire), up to 0.5 µg of DNA and water to a final volume of 20 µl (unless specified). The reaction was incubated at 37°C for an hour. The DNA was phenol/chloroform extracted (2.10) to denature and remove the enzyme from the reaction and ethanol precipitated (2.11) to concentrate the DNA. CIP is also active in NEB buffer 2 (2.2.12), NEB buffer 4 (2.2.14) as well the NEB buffers for *Eco*RI (2.2.20) and *Bam*HI (2.2.19)

2.13 Transformation of *E. coli*

2.13.1 Preparation and transformation of *E. coli* (DH5α) (Rubidium chloride method)

A single colony of *E. coli* (DH5α) was selected from a freshly streaked agar plate, inoculated into 10 ml of SOB broth (2.1.1) and incubated overnight at

37°C in a shaker at 250 rpm. The overnight culture (2 ml) was transferred into 800 ml of LB pre-warmed to 37°C. This was incubated at 37°C in a shaker at 250 rpm until the culture was midway through the log phase (OD of 0.45 at 550 nm). The cells were then chilled on ice for 30 minutes and then pelleted at 4°C (2500 rpm for 15 minutes in a Beckman JA14 rotor). The supernatant was then removed and the cells resuspended by gentle pipetting in 264 ml of ice cold RF1 (2.2.4). The resuspended cells were then incubated on ice for one hour. The cells were pelleted again and the supernatant removed (2500 rpm for 15 minutes in a Beckman JA14 rotor). The cells were resuspended in 64 ml RF2 (2.2.5) and incubated on ice for 15 minutes. The prepared cells were then aliquotted into 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and stored at -80°C until required.

The competent *E. coli* (DH5α) were thawed on ice and 100 μl was added to a ligation mix (20 μl), swirled to mix and allowed to incubate on ice for 30 minutes. The cells were heat shocked at 37°C for 45 seconds and returned to ice for a further two minutes. One hundred μl of 2 X LB (2.1.3) was added and the cells were then incubated at 37 °C in a shaker at 250 rpm for 1 hour. Filter sterilised 1 M IPTG (10 μl) and 50 μl of 20 mg/ml X-gal were added to the cells, which were plated out on selective media.

2.13.2 Preparation and transformation of *E. coli* (DH5α) (CaCl₂ method)

A single colony of *E. coli* (DH5α) was selected from a freshly streaked agar plate, inoculated into 10 ml of LB broth (2.1.1) and incubated overnight at 37°C in a shaker at 250 rpm. One hundred µl of the overnight culture was added to 30 mls of sterile pre-warmed LB broth (2.1.3) and incubated at 37°C until the cells had entered logarithmic growth phase, (0.4 OD at 550nm). The cells were pelleted in a cooled Beckman centrifuge (4000 rpm for 5 minutes at 4°C) using a Beckman JA20 rotor. The supernatant was discarded. The cells were resuspended in ice cold sterile 50mM CaCl₂ (2.3.2) (at 20% of the original volume of media) and left on ice for 20 minutes. The cells were again pelleted in a cooled Beckman centrifuge (4000 rpm for 5 minutes at 4°C) using a Beckman JA20 rotor and the supernatant was discarded. The cells were again incubated on ice for 20 minutes. The wash step was repeated for a third time and the cells were resuspended in ice cold CaCl₂ (2.3.2) at 4% of the original volume of media and incubated on ice for 30 minutes.

100 µl of the competent cells were then added to 10 ng of transforming DNA in pre chilled 1.5 ml microfuge tubes. The contents were mixed by flicking the tubes and the tubes placed on ice for 30 minutes. The cells were then heat shocked for one minute in a 37°C water bath before being returned to ice for 2 minutes. 0.9 ml LB broth (2.1.3) was added to the cells and these were incubated at 37°C for 45 minutes. Subsequently 200 µl aliquots of the transformed cells were plated on LB media (2.1.2) containing ampicillin (2.5) at a concentration of 50 µg/ml and incubated at 37°C overnight.

2.14 Purification of plasmid DNA (small scale)

Plasmid DNA was recovered from cells grown in LB broth (2.1.3) using the Promega Wizard™ miniprep kit system (Promega, Madison) according to manufacturer's supplied instructions.

2.15 DNA sequencing

2.15.1 Sequencing of clones (version 3)

Plasmid DNA was prepared for sequencing using Promega wizard miniprep kit (2.14). Plasmid DNA was then electrophoresed on a 1% agarose gel (2.11.1) to estimate the yield of the plasmid. Sequencing reactions were set up according to the Big Dye (version3) protocol according to instructions supplied by Birmingham University and the samples sent for analysis to the genomics lab at Birmingham University.

2.15.2 Sequencing of clones (Plasmid to profile)

Plasmid DNA was prepared for sequencing using Promega wizard miniprep kit (2.14). Plasmid DNA was then electrophoresed on a 1% agarose gel (2.11.1) to estimate the yield of the plasmid. Plasmid DNA (1 ng) was combined with 3.4 picomoles primer in 10 µl with double distilled sterile water and supplied to the

genomic lab at Birmingham University for sequencing reactions using big dye version 3 protocol.

2.16 Reagents suppliers

General chemicals were purchased from Sigma-Aldrich (Poole, UK) unless otherwise stated. Enzymes were purchased from New England Biolabs (UK) or Promega (UK) unless otherwise stated. ATP was purchased from Amersham Biosciences (now GE Healthcare, UK) as a 100 mM solution.

CHAPTER 3

SINGLE STRANDED LIGATION OF "MAX"

TRINUCLEOTIDES USING T4 RNA LIGASE

Chapter 3: Single stranded ligation of "MAX" trinucleotides using T4 RNA ligase

3.1 Introduction

"MAX" trinucleotides are the preferred codons of usage in *E. coli* (Nakamura *et al.*, 1997). Since the aim of this project was to join contiguous "MAX" trinucleotides together, an obvious approach was to employ RNA ligase to ligate single stranded DNA trinucleotides together (ssDNA is a substrate for RNA ligase) onto a conserved sequence oligonucleotide. The resulting construct could then be amplified and converted to dsDNA by PCR.

3.2 Preliminary Assessment

Before the ligation of the trinucleotides was attempted, a model ligation of two oligonucleotides was attempted, to ensure that RNA ligase could indeed ligate ssDNA. Two oligonucleotides MA1 & MA2 and a PCR primer, Primer1 were obtained from MWG. MA2 was modified at the 3' end with a C_7 amino group to prevent self-ligation in subsequent experiments (Figure 3.1). MA2 was 5' phosphorylated (2.12.1) and a single stranded ligation was set up (2.12.2.3). The results were analysed on a 3% agarose gel (2.7.1). No product of the expected size resulted (data not shown).

The first ligation was performed in a relatively large volume and therefore the experiment was repeated, both in a smaller volume and with different

MA1   MA2

5'   3' 5'   3'

CTATGCCATGACTGAAGCTTTAGTCGCAGCGACGAATTA   TAATTCGTCGCTGCGACTAAAGCTTCAGTCATGGCATAG-NH₂

                                                                                          AGTCAGTACCGTATC

CTATGCCATGACTGA   3'   5'

5'   3'

Primer1   Primer1
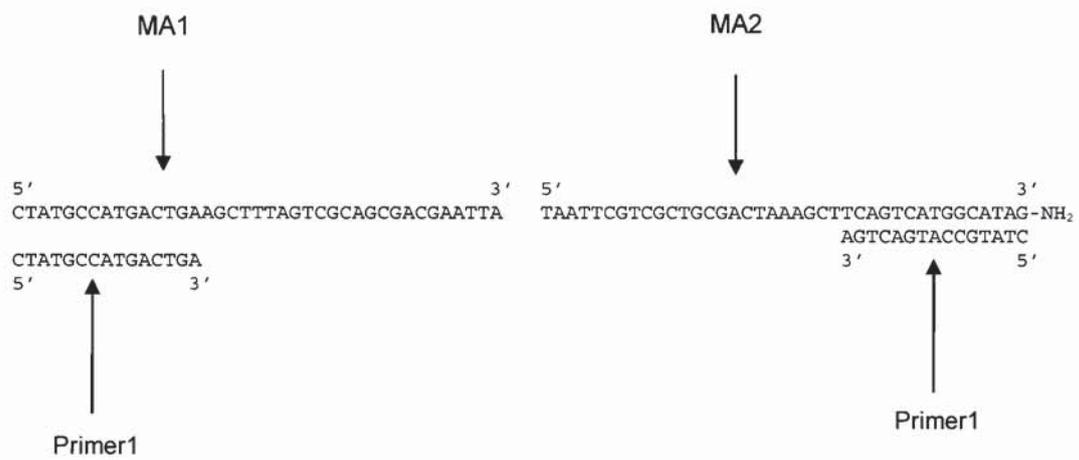
Figure 3.1: Schematic representation of the two single stranded DNA segments to be ligated using RNA ligase.
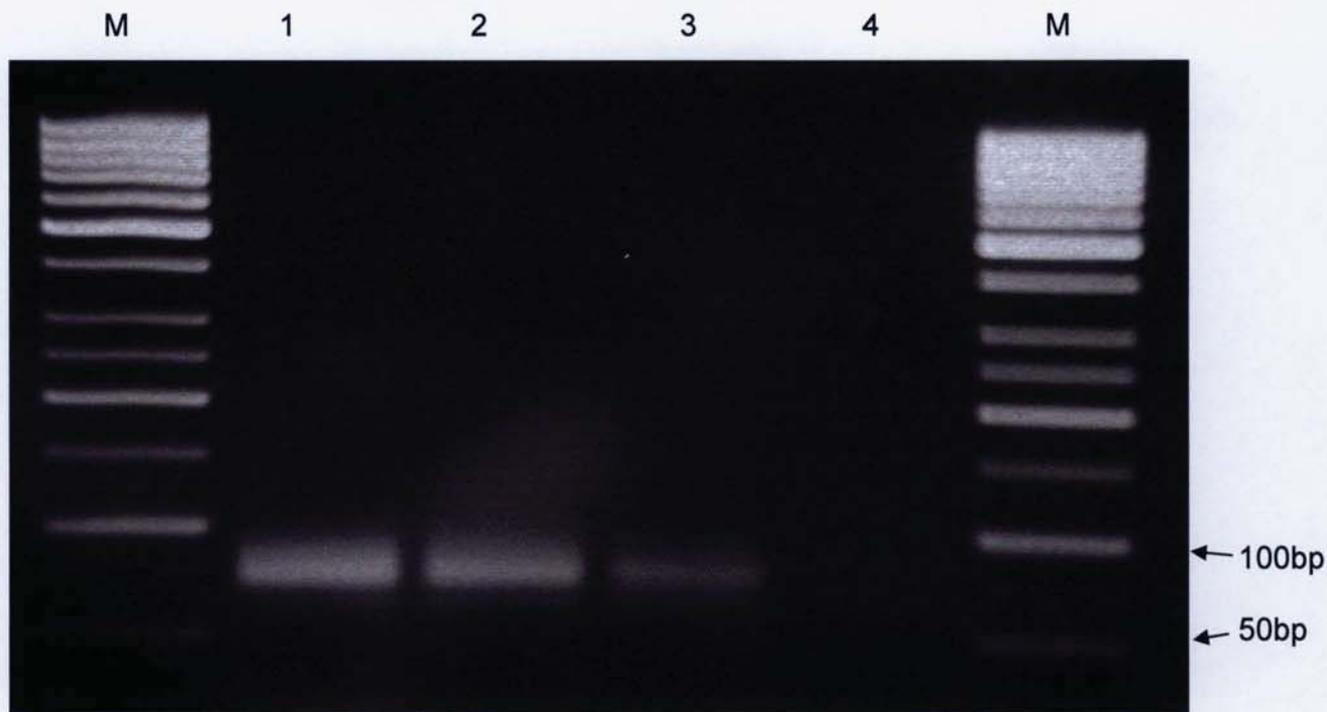
Figure 3.2: Amplification of DNA ligated with RNA ligase at varying temperatures. One µl (5%) of each ligation was amplified and 10 µl of the resulting PCR product (10%) was electrophoresed. The temperatures at which the ligations took place for the template are: Lane 1: 37°C, lane 2: 26 °C, lane 3: 16 °C and lane 4: 4 °C. Lane M: 50 bp marker (Biorad).

temperatures (37°C, 26°C, 16°C, and 4°C; 2.12.2.4), to try to improve the yield of ligated DNA. The products were electrophoresed (2.7.1) on a 3% agarose gel and also amplified by PCR (2.6.2). No product was visible directly from the ligation, but a product of the required size was visible after amplification (Figure 3.2)

The amplification reactions show that the single stranded ligase was working (Figure 3.2). The optimum temperature for RNA ligase was chosen as 37°C since the PCR product from the 37°C ligation showed the brightest band. The amount of product diminished as the temperature at which the ligations were incubated decreased. The RNA ligation at 4°C acted as a negative control and showed that the product generated by the PCR was due to the ligation of the oligonucleotides. The ligation is not, however, at a level that can be visualised on an agarose gel without amplification.

The work of Tessier *et al.*, (1996) showed that the ligation of DNA could be achieved to quantitative levels using RNA ligase. However, the composition of the buffer that was used by Tessier *et al.*, was different to the buffer used in this study which was supplied by NEB with the RNA ligase. Tessier also used acceptor (OH group) to donor (phosphate group) ratio of 5:1 and used PEG at 16% as a molecular chelator.

In an attempt to mirror Tessier's work more closely, new oligonucleotides, MA3 and MA4 and a second primer, Primer 2 were purchased from MWG

Aston University

Illustration removed for copyright restrictions

Figure 3.3: Schematic representation of the two single stranded DNA segments to be ligated using RNA ligase, modified from a method by Tessier *et al* (1996).

(Figure 3.3). These oligonucleotides were similar in design to the oligonucleotides used by Tessier *et al.*, with minor modifications made to the oligonucleotides to incorporate the sequences corresponding to the two PCR primers.

Single stranded ligations in both RNA ligase buffer from NEB (2.2.15) and Tessier's single-stranded ligase buffer, with and without $MnCl_2$ (2.2.17/2.2.18) and with acceptor:donor ratios of both 5:1 and 2.5:1 were attempted at 37°C. In all cases, the ligations contained 16% PEG and 1 units of RNA ligase. None of the ligations yielded a product distinct from the negative (- ligase) control (data not shown).

3.3 Summary

At best, single-stranded ligation with RNA ligase had given poor yields. For "MAX" randomisation, near quantitative yields would be required. Moreover, single-stranded ligation would offer little opportunity for control over the numbers of trinucleotides added, or their identity. The approach was therefore abandoned.

CHAPTER 4

GENERATION OF A RANDOMISED DNA SEQUENCE ON

OLIGO AFFINITY SUPPORT

Chapter 4: Generation of a randomized DNA sequence on oligo affinity support

4.1 Introduction

Oligo affinity support (PS) is a non-shrinking, non-swelling matrix compatible with automated synthesis. It has a low non-specific binding to DNA. The oligo affinity support (PS) has an adenosine residue attached through the exocyclic amino group. The synthesis of DNA on the oligo affinity support (PS) progresses normally on removal of the 5' DMT group. The DNA generated on the oligo affinity support (PS) can be treated with ammonium hydroxide to deprotect it. This action however leaves the DNA still attached to the oligo affinity support (PS). The Oligo affinity support was selected for use in this study, as it would enable the separation of the template strand away from the ligated MAX trinucleotide strand, so facilitating purification of the latter.

4.2 Initial methodology: adding trinucleotides to a template strand

A randomised template oligonucleotide (Fig 4.1) was synthesised in-house on oligo affinity support (PS) (2.4.1), deprotected and washed (2.4.2). The presence of the beads precluded estimation of oligonucleotide concentration by absorbance at 260 nm. Therefore to carry out any experiments using this oligonucleotide, the following assumptions were made. The Oligo affinity support was supplied by Glen Research at a binding capacity of 42 μmol/g. Thus, synthesis of the randomised oligonucleotide was on a 42 nmol scale. The

Table 4.1: Identities and absorbances of the trinucleotides coding for each of the 20 amino acids preferably used by *E. coli* (Nakamura *et* al., 1997).

MA5   MA6

```
5'                    3' 5' 3' 5' 3' 5' 3' 5' 3' 5' 3' 5' 3' 5' 3' 5'                    3'
TGACTGAAGCTTTAGT  MAX  MAX  MAX  MAX  MAX  MAX  MAX  CATCAGCGTACGACGT
ACTGACTTCGAAATCA  NNN  NNN  NNN  NNN  NNN  NNN  NNN  GTAGTCGCATGCTGCA
3'                                                                   5'
                                                      GTAGTCGCATGCTGCA
                                                      3'            5'
```

Oligo affinity   On-bead template   Primer3
support (PS)
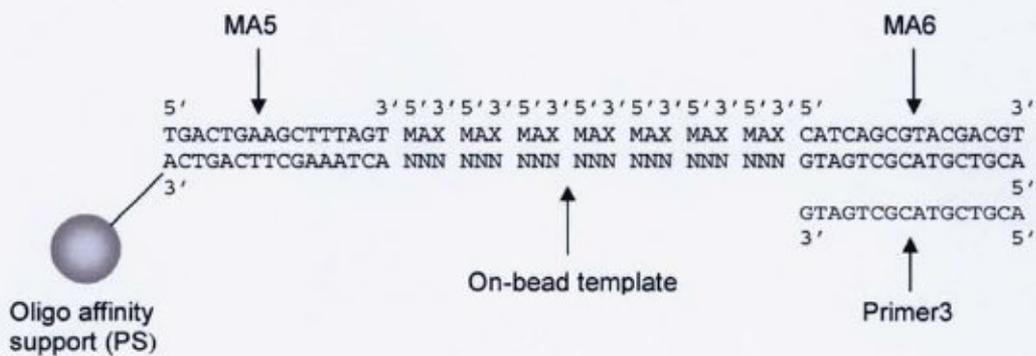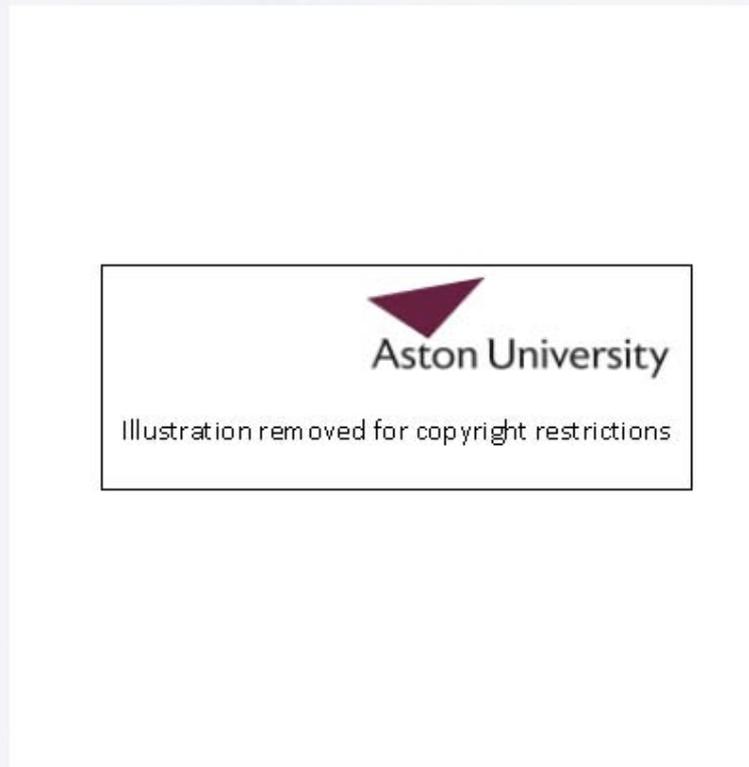
Figure 4.1: Schematic representation of the on-bead template showing the hybridisation oligonucleotides (MA5 and MA6) and reverse primer (Primer3). Note that MA5 also serves as the forward primer.

synthesis of DNA is estimated to be 99% efficient per cycle and after 50 cycles 50% was assumed to have been lost and therefore the final full length template strand was assumed to be at a 21 nmol scale.

"MAX" trinucleotides are the preferred codons of usage in *E. coli* (Nakamura et al., 1997) and were also synthesised in-house (2.4.1), deprotected (2.4.2) and their concentrations estimated by absorbance at 260 nm (2.9.1) (Table 4.1). Each of the "MAX" trinucleotides (40 picomoles) were 5' phosphorylated (2.12.1) in a final combined volume of 100 μl to enable their use in subsequent ligation reactions. The 5' phosphorylated "MAX" trinucleotides were stored at -20°C until required.

Two other oligonucleotides, MA5 and MA6 were synthesised (2.4.1) and deprotected (2.4.2) to complete the conserved region of the DNA cassette (Figure 4.1). A third oligonucleotide, Primer3 was obtained from Dr. M.D. Hughes (Aston University) for use as a PCR primer alongside MA5 (Figure 4.1).

A hybridisation and ligation reaction was set up (2.5.1) to generate the DNA cassette illustrated in Figure 4.1. The resulting product was amplified (2.6.1) with MA5 & Primer3 and using 1 μl of the purified hybridised / ligated product as template. The product was examined by electrophoresis both before and after amplification but could only be visualised after amplification (Figure 4.2).
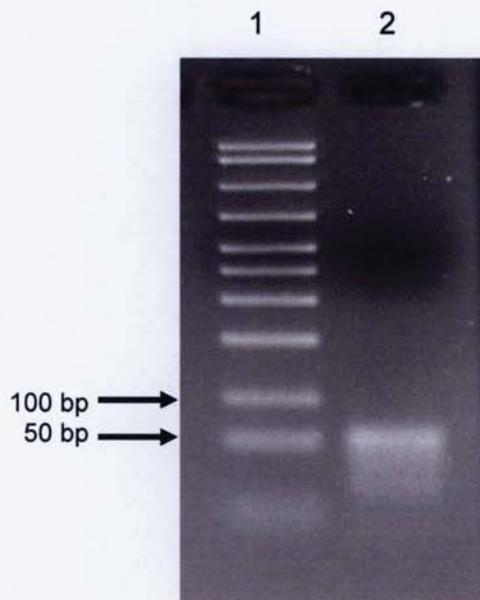
Figure 4.2: Agarose gel (3%) showing lane 1, 50 bp marker (Biorad); 10 µl (10%) of the product of amplification of the hybridised and ligated product shown in Figure 4.1. The expected size of the DNA was 53 bp.

Table 4.2: Results of first cloning experiment. The white colonies found in the product ligation transformation have incorporated the synthetic DNA cassette randomised at 7 positions.

	Blue colonies	White colonies
Product ligation	64	304
Cut pUC 19	19	0
Self ligation	230	0
Uncut pUC 19	Approximately 10000	0
No plasmid	0	0
No amp/no plasmid	Lawn	0

The size of the amplified DNA corresponds to that expected (53bp, Figure 4.2). This product of the PCR reaction was extracted with phenol-chloroform (2.10) and the DNA was ethanol precipitated (2.11). A phosphorylation reaction was set up to add a phosphate group to the 5' ends of the DNA cassette (2.12.1). The precipitated DNA was not resuspended in water, but the phosphorylation reaction was set up in the same test tube so that the entire DNA (approximately 50 picomoles) would be phosphorylated. T4 ligase buffer was used rather then PNK buffer as it contains ATP and is 100% compatible with PNK. This eliminated the phenol chloroform and the ethanol precipitation procedures and the 5' phosphorylated DNA could be used directly in ligations.

Ligations were set up with the 5' phosphorylated PCR product and *Sma*I restricted, phosphatased pUC 19 (2.12.2.1). These ligations were incubated at 4°C overnight and then transformed into competent *E. coli* (DH5α) (2.13.2). The transformed *E. coli* was plated onto nutrient agar plates containing 50 µg/ml ampicillin (2.1.2). A fourth plate was set up and used to plate untransformed *E. coli* on to a nutrient agar plate (2.1.2) with ampicillin and a fifth was set up but without the ampicillin. The colonies on each were counted. The white colonies in the product ligation plate (see Table 4.2) show that the DNA has been incorporated into the plasmid.

Eleven white colonies were selected randomly and the plasmid DNA prepared (2.14). The prepared plasmid DNA was visualised on a 1% agarose gel (2.7.1) to estimate the quantity of DNA recovered. A double digest was subsequently set up using *Bam*HI and *Eco*RI (2.12.3). These two enzymes' restriction sites

are located either side of the *Sma*I site so if the PCR product had been incorporated into the plasmid then the size of the excised DNA fragment will be larger than if the PCR product had not been incorporated. The restriction digest of the plasmids was then visualized by electrophoresis on a 3% agarose gel (2.7.1) (Figure 4.3).

The results shown in Figure 4.3 demonstrate that upon restricting the plasmid DNA with *Bam*HI and *Eco*RI, a DNA fragment is generated that corresponds to a size of between 50 and 200 bp. Specifically, lanes 2, 3, 4, 7, 8, 11 and 12 have a single insert which corresponds to a size of approximately 75bp. Lanes 5 and 9 appear to have incorporated a concatamer made up of two inserts ligated together. Lane 6 and 10 have a concatamer of three of the inserts ligated into the plasmid. If the cassette DNA had not ligated into the *Sma*I site of pUC 19 then the excised fragment would have been expected to have a size of approximately 20 bp.

Sequencing reactions (2.15.1) were carried out and sent for analysis to Birmingham University. The sequencing showed that the "MAX" codons had not been incorporated into the inserted oligonucleotide (data not shown). The only "MAX" codons that were found were at a level lower then would be expected if they had occurred due to random chance. This lack of "MAX" codons in the DNA suggested that a different approach was required.

Figure 4.3: Restriction analysis of plasmid DNA from selected white colonies. Five μl of each plasmid mini-prep (10%) was digested in a 20 μl reaction with both *Eco*RI and *Bam*HI and 50% of each digest was examined by electrophoresis. Lane 1, 50bp marker (BioRad); lanes 2-12, digests of clones 1-11.

Figure 4.4: Schematic representation of the redesigned oligonucleotides in relation to each other. MA7 is longer than MA6 (Figure 4.1) and Primer 4 has a sequence that is not found on the template strand.

4.3 A revised methodology for incorporation of "MAX" trinucleotides into oligonucleotide strand

A likely explanation for the inclusion of such low incorporation of "MAX" trinucleotides into the cloned products was that the template oligonucleotide was leeching from the OAS(PS) support and being amplified by PCR. To circumvent this eventuality, the oligonucleotides were redesigned. Specifically, the distal end of cassette (with respect to the support) was extended so that the new reverse PCR primer, Primer4, could not hybridize to the on-bead template (Figure 4.4).

A hybridisation and ligation reaction was set up (2.5.1) using the redesigned oligonucleotides and the product amplified by PCR (2.6.1) with MA5 and Primer4 as the forward and reverse primers. The PCR reaction was visualised on a 3% agarose gel (2.11.1), which revealed two products (Figure 4.5). The upper band was less intense but was the expected size. However the lower band at 40-45 bp does not correspond to the expected size and may be a primer dimer.

4.3.1 Varying the quantity of template

The low yield of the expected product and the presence of the lower second product may be due to low amount of template. To test this hypothesis and to optimise the quantity of DNA used in future PCR reactions, a number of DNA amplification reactions were set up (2.6.1) with varying quantities of template.
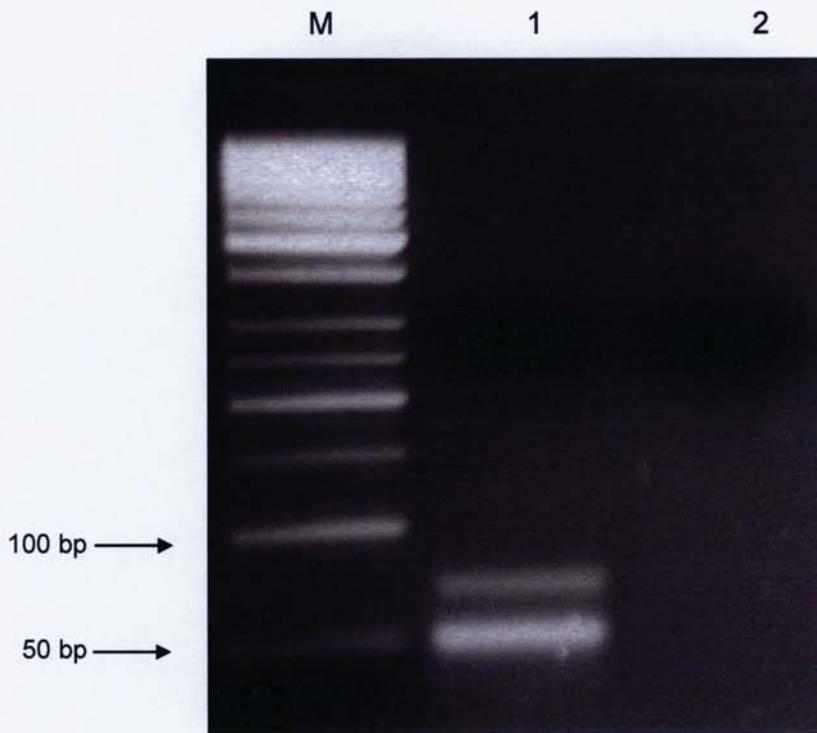
Figure 4.5: Agarose gel (3%) showing the product of a PCR reaction using the new extended oligonucleotide design (see Figure 4.4) in the initial hybridisation. Lane M: 50bp marker (Biorad); Lane 1 contains 10µl of the PCR product; Lane 3 contains 10µl of the PCR product of the negative control. The larger fragment of DNA corresponds to the expected size 69 bp. The smaller and brighter band of approximately 50 bp is a possible primer dimer.

Figure 4.6: Agarose gel (3%) showing the products of amplification reactions (10%) using varying amounts of hybridised and ligated product as template. Lanes 1, 2, 3 and 4 contain 4µl, 6µl, 8µl and 10µl of template respectively; Lanes M: 50bp marker (BioRad).

The results indicate that varying the amount of template made little difference to the intensity of the upper band (Figure 4.6). The intensity of both bands increased with the increasing amount of template used. Therefore future PCR reactions were set up with 2 µl of hybridisation product as template.

4.3.2 Optimisation of annealing temperature of PCR

Varying the amount of template did not alleviate the problem of the two DNA bands being amplified. To optimise the annealing temperature of the PCR, a gradient PCR was performed (2.6.3).

The results of the gradient PCR (Figure 4.7) indicate that the optimum annealing temperature for PCR was 48°C. The gradient PCR also showed that the PCR product diminishes as the annealing temperature exceeds 48°C and there was no product at all at annealing temperatures of 56°C or above. The gradient PCR failed to resolve the matter of the two bands of DNA.

4.3.3 Cloning of excised PCR product

The PCR product was electrophoresed on a 2% Sea Plaque (LMP) gel (2.7.1) and the upper band of DNA excised from the gel and the DNA eluted (2.8.1). Ligation reactions were set up (2.12.2.1) but instead of the PCR product being used in the ligations, the DNA eluted from the LMP gel was ligated into *Sma*I phosphatased pUC19. The ligated DNA was transformed into competent *E.coli* (2.13.2) (Table 4.3). Ten white colonies were selected, the plasmid DNA

99

Figure 4.7: Gradient PCR to optimise the annealing temperature of the PCR. Annealing temperatures: lane 1 40°C, lane 2 40.5°C, lane 3 41.5°C, lane 4 43.2°C, lane 5 45.5°C, lane 6 48.4°C, lane 7 51.7°C, lane 8 54.6°C, lane 9 56.8°C and lane 10 58.4°C. Lanes M: 50bp marker (BioRad)

Table 4.3: Results of cloning experiment to ligate the amplified 67 bp synthetic DNA cassette into pUC19 and clone that into *E. coli.*

	Blue colonies	White colonies
Product ligation	48	1120
Cut pUC 19	10	0
Self ligation	3-5000	0
Uncut pUC 19	5-8000	0
No plasmid	0	0
No amp/no plasmid	Lawn	0

recovered (2.14), and sequencing reactions set up and sent to Birmingham University (2.15.1). The sequencing data indicated that the "MAX" trinucleotides had not been incorporated (data not shown). The lack of "MAX" trinucleotides in the sequencing results again suggested that there was a problem with the methodology.

Re-inspection of the oligonucleotide design revealed that unexpectedly, the template strand might still be amplified by PCR using primers MA5 and Primer4. This can occur after the initial cycle of PCR. Specifically, oligonucleotide MA5 hybridises with the template strand and is extended to form double stranded DNA. Meanwhile, Primer4 hybridises with MA7 and is also extended to form double stranded DNA (Figure 4.8A). The products of these extensions may then hybridise with each other during subsequent cycles (Figure 4.8B) to generate a full length DNA fragment (Figure 4.8C). Once the full-length product has been generated then it may be amplified with the forward and reverse primers MA5 and Primer4 (Figure 4.8D).

The smaller, 50 bp products seen in Figures 4.5, 4.6 and 4.7 may be caused by the hybridisation of MA7 in the randomised region of the on-bead template. Subsequent ligation with MA5 would form a product 48 bp in size (Figure 4.9).

A

MA5

5′   3′
TGACTGAAGCTTTAGT
ACTGACTTCGAAATCA NNN NNN NNN NNN NNN NNN NNN GTAGTCGCATGCTGCA
3′   5′

Oligo affinity
support (PS)

MA7

5′   3′
CATCAGCGTACGACGTCAGCGACCAGATGATG
        GTCGCTGGTCTACTAC
3′   5′

Primer4

B

3′   5′
GTAGTCGCATGCTGCAGTCGCTGGTCTACTAC
CATCAGCGTACGACGTCAGCGACCAGATGATG
5′   3′

3′   5′
ACTGACTTCGAAATCA NNN NNN NNN NNN NNN NNN NNN GTAGTCGCATGCTGCA
TGACTGAAGCTTTAGT NNN NNN NNN NNN NNN NNN NNN CATCAGCGTACGACGT
5′   3′

C

3′   5′
GTAGTCGCATGCTGCAGTCGCTGGTCTACTAC
TGACTGAAGCTTTAGT NNN NNN NNN NNN NNN NNN NNN CATCAGCGTACGACGT
5′   3′

D

MA5

5′   3′
TGACTGAAGCTTTAGT
3′   5′
ACTGACTTCGAAATCA NNN NNN NNN NNN NNN NNN NNN GTAGTCGCATGCTGCAGTCGCTGGTCTACTAC
TGACTGAAGCTTTAGT NNN NNN NNN NNN NNN NNN NNN CATCAGCGTACGACGTCAGCGACCAGATGATG
5′   3′
        GTCGCTGGTCTACTAC
3′   5′

Primer4

Figure 4.8: Schematic representation of the generation of full length product without the incorporation of "MAX" trinucleotides. A, preliminary hybridisation of primers; B amplification products of initial hybridisation; C hybridisation of amplified products; D PCR of full length product.
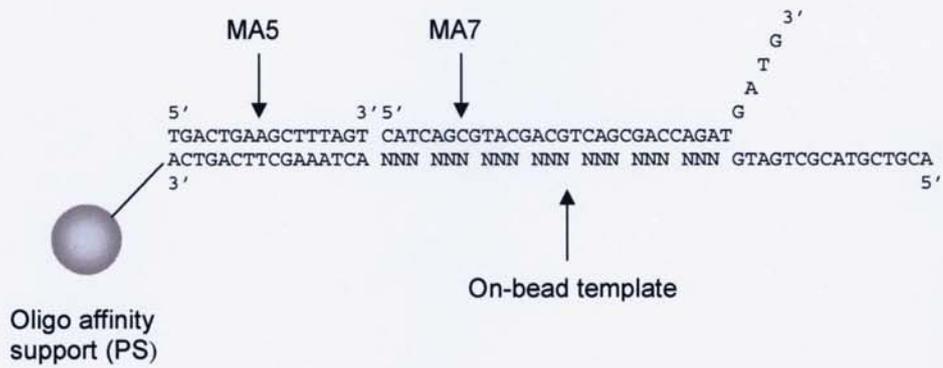
Figure 4.9: Schematic representation of potential hybridisation between MA7 and the randomised region of the template strand.

4.4 Summary

Experiments employing OAS(PS) support had consistently given PCR products of the required length, but without the incorporation of "MAX" trinucleotides. Meanwhile, concurrent work by Dr Marcus Hughes (Aston University) with non-contiguous codons demonstrated conclusively that the template strand was indeed leeching from the support and being carried into subsequent PCR reactions. The oligo affinity support was therefore abandoned.

CHAPTER 5

CLONING OF DNA SEQUENCE RANDOMIZED AT TWO

AMINO ACID POSITIONS

Chapter 5: Cloning of DNA sequence randomized at two amino acid positions

5.1 Introduction

The basic principles described in Chapter 4 were continued but henceforth, all randomisation studies were performed in solution, rather than on a solid support. At the same time, owing to the potential difficulties illustrated in Figure 4.9, the number of amino acid positions randomised was reduced from seven to two. The new methodology envisaged was to employ a template and hybridise a pair of oligonucleotides, which had extensions past the template at both ends (Figure 5.1). The extensions would be complementary to the two primers thus only allowing the amplification of a ligated strand and preventing amplification of the template (Figure 5.1).

5.2 Cassette hybridisation

Hybridisation and ligation reactions were set up (2.5.2) as illustrated in Figure 5.1, with and without "MAX" trinucleotides. The purpose of the latter was to act as a negative control, to examine hybridisation in the absence of the trinucleotides. The results, visualised on 3% agarose (2.7.1), indicate that there is very little difference between the two hybridisation reactions. Both produce bands at the same positions on the agarose gel (Figure 5.2), although the largest band in lane 1 is brighter then the corresponding band in lane 2. The presence of a band of the correct size in lane 2 without "MAX" trinucleotides is possibly

106

```
                    Primer 6                                              MA7
                        │                                                  │
                        ▼                                                  ▼
5'          ▼          3'
CTATGCCATGACTGA
5'                                              3'5'3'5'3'5'                                    3'
CTATGCCATGACTGAAGCTTTAGTCGCAGCGACGAATTA MAX MAX CATCAGCGTACGACGTCAGCGACCAGATGATG
          TCGAAATCAGCGTCGCTGCTTAAT NNN NNN GTAGTCGCATGCTGCA
          3'         ▲                         ▲                    5'
                                                                    GTCGCTGGTCTACTAC
                                                                    3'              5'
                     │                         │                         ▲
                     │                         │                         │
                   MA8                     Template1                  Primer4
```

Figure 5.1: Schematic representation of Template1, which contains two conventionally randomised codons, in relation to MA7 & MA8 and the PCR primers. Note that both PCR primers hybridise outside the template strand and should prevent the amplification of the template strand.
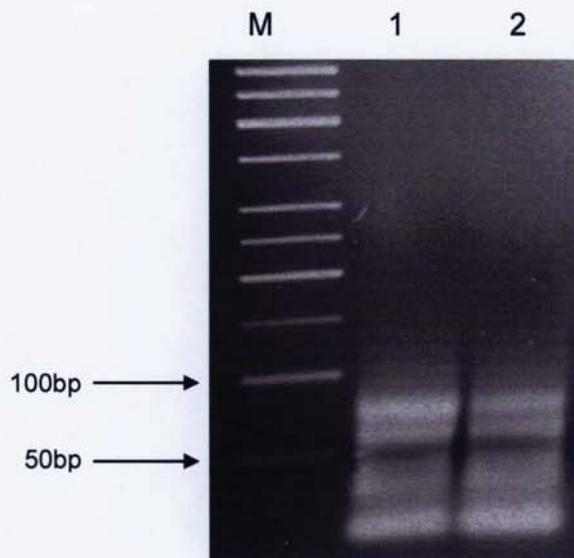
Figure 5.2: Visualisation of the hybridisation reaction illustrated in Figure 5.1. Lane 1, +"MAX" trinucleotides; lane 2, -"MAX" trinucleotides; lane M, 50bp marker (Biorad).

107

due to hybridisation between Template1, MA7 and MA8 in the absence of ligation. Since the gel depicted in Figure 5.2 was electrophoresed at room temperature and the lowest calculated annealing temperature in the hybridisation was 46°C, it is likely that the hybridised product will be stable and appear on the agarose gel as a single band corresponding to the correct product size. Unsurprisingly, this larger band was not eliminated by heating the hybridisation prior to electrophoresis (data not shown).

5.3 Excision, amplification and cloning of the randomised cassette

To maximise the possibility of amplifying the required 77 bp cassette, the band corresponding to 77 bp was excised from lane 1 (Figure 5.2) and the DNA eluted (2.8.1) and amplified (2.6.1) using 1μl of the extracted DNA as template. The required product was amplified successfully (Figure 5.3) although a smaller band corresponding to approximately 50bp was also apparent. The amplified DNA was phenol chloroform extracted (2.10), ethanol precipitated (2.11) and subsequently ligated into pUC 19 (2.12.2.1). The ligations were then transformed into competent *E.coli* (DH5α) (2.13.1) (Table 5.1). Nineteen white colonies were selected and the plasmid DNA was extracted (2.14). Plasmid samples were assessed by electrophoresis and then sequenced (2.15.1).

The sequencing data (Figure 5.4) was encouraging. Of the 19 samples that were sent for sequencing, 17 produced readable data. Of these, two contained concatamers formed by the ligation of two inserts. Including these concatemers, a total of 38 randomised codons were sequenced, comprising 19 "MAX"

Figure 5.3: Amplification of the 77 bp product extracted from lane 1 of the gel illustrated in Figure 5.2.

Table 5.1: Cloning of the 77bp amplified DNA cassette in pUC19.

	Blue colonies	White colonies
Product ligation	84	53
Cut pUC 19	0	0
Self ligation	700	0
Uncut pUC 19	3000	0
No plasmid	0	0
No amp/no plasmid	Lawn	0

```
          Conserved sequence          Randomised    Conserved sequence
   ◄─────────────────────────────►      codons    ◄─────────────────────────────►

TTT AGT CGC AGC GAC GAA TTA  AGC AAA  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  CTG ATG  CAT CAG CGT ACG ACG
NNT AGT CGC AGN GAC GAA TTA  TGG ATT  CAT CAG NGT ACG ACG
TTT A*T CGC AGC GAC GAA TTA  TTT AGC  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  AGT GGC  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  ACC CAG  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  CAT CA*  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  CAN GNC  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC *** CAG  ATG ATG  C*T CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  CAG TGC  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  ACC CAC  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA  NNG ATG  CAT CAG CGT ACG ACG
TTT AGT TGC AGG GNC GAA TTA  CAT GAN  CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA           CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA           CAT CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA           CA* CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA           C*T CAG CGT ACG ACG
TTT AGT CGC AGC GAC GAA TTA           C*T CAG CGT ACG ACG
TTT AGT NGC AGN GAN GAA TTN           CAT CAG CGT ACG ACG
```

Figure 5.4: Sequencing data of cassettes randomised with "MAX" trinucleotides, at two positions. Legend: Red, "MAX" trinucleotides; Purple, ambiguous sequences; Green, non-"MAX" trinucleotides; *, Deletion.

codons, 4 codons with ambiguous sequences, 2 non-"MAX" codons and 1 codon that contained a one base deletion. The remaining 6 clones (12 codons) contained no insert corresponding to the randomised sequence. Rather, their sequence revealed that oligonucleotides MA7 and MA8 had ligated together directly.

If all the sequencing results are considered together, a 42% incorporation of "MAX" trinucleotides was achieved. Conversely, if the clones in which "MAX" trinucleotides were omitted are removed from the analysis, then the percentage of "MAX" codon incorporation rises to nearly 71%. Moreover, if the four ambiguous codons were in fact "MAX" codons" the percentage of "MAX" incorporation could rise as high as 88%. Again omitting the clones that contained no randomised codons, 53% of clones contained "MAX" trinucleotides in both randomized positions.

5.4 Summary

Results obtained from the ligation of trinucleotides were encouraging, but the relatively high proportion of clones containing no randomised codons was problematic for future application of the methodology. As suspected with previous methodology (Chapter 4), extended regions of NNN randomisation in the template strand were enabling proximal annealing of outer, constant sequence oligonucleotides. Thus the direct ligation of oligonucleotides MA7 and MA8 demonstrated that the potential mis-annealing illustrated in Figure 4.9 was in fact occurring. In the current protocol, the low ligation temperature of 4°C would facilitate such annealing considerably. If this occurred, then subsequent ligation would yield a product that differs from the required product by only 6 bp. Since 6 bp is difficult to differentiate on a 3% agarose gel, the two products would likely appear as one band on an agarose gel, as seen in Figure 5.3.

From the high degree of incorporation of "MAX" trinucleotides into clones that contained randomised codons, the methodology clearly showed promise, but would require further revision.

CHAPTER 6

DOUBLE STRANDED LIGATION OF "MAX"

OLIGONUCLEOTIDES

Chapter 6: Double stranded ligation of "MAX" oligonucleotides

6.1 Introduction

The use of trinucleotides in randomising DNA was partially successful. However, to improve the randomisation process further a significant revision was required. The trinucleotide approach utilised very low temperatures, which enabled non-specific hybridisation to take place leading to the ligation of hybridisation oligonucleotides. To overcome this problem, oligonucleotides were redesigned (Figure 6.1). These oligonucleotides were a minimum 9bp in length, which allowed annealing at higher temperatures than was possible with the trinucleotides. The higher annealing temperatures would in theory reduce the probability of mismatches during the hybridisation reactions.

6.2 Randomisation of two consecutive codons

This method is a modification of the MAX randomisation technique described by Hughes and co-workers (2003). Oligonucleotides Max1, Max2 and MA10 were 5' phosphorylated (2.12.1; 300 picomoles of each oligonucleotide in final reaction volumes of 50 µl) to enable their use in subsequent ligation reactions. Two hundred picomoles of the oligonucleotides Template2, Max1, Max2, MA9 and MA10 were hybridised and ligated together, (2.5.4) overnight. Two control ligations, one without ligase and one without Max1 and Max2 were also set up. The resulting constructs were amplified by PCR (2.6.1) with Primer4 and
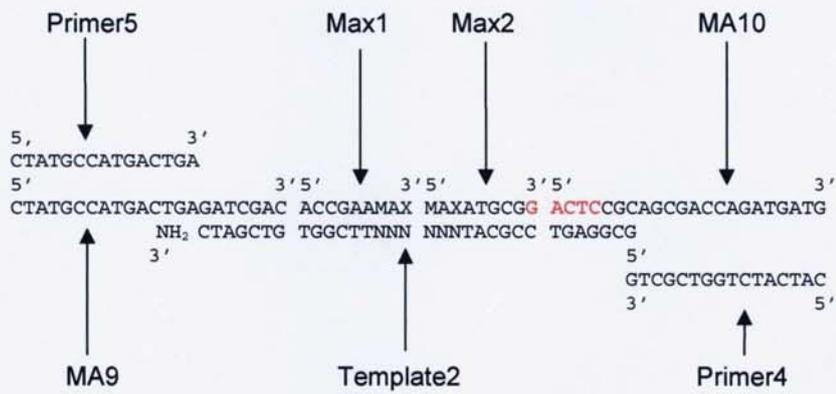
Figure 6.1: Schematic representation and nucleotide sequences of redesigned oligonucleotides. Note that Max2 and MA10 together form an *Mly*I recognition site (shown in red).

Primer5 (Figure 6.1), and 1 µl of the appropriate hybridisation and ligation reactions, both neat and diluted 50 fold, as template. The amplification reactions were visualised on a 3% agarose gel (2.7.1) (Figure 6.2).

The amplification produced a product of the expected size (Figure 6.2, lanes 1 & 2) with no amplification in any of the controls (Figure 6.2, lanes 3-7), confirming that successful amplification is dependent upon ligation of the "MAX"-containing oligonucleotides in the construct.

To investigate the most efficient methodology for subsequent precipitation of amplified DNA, similar samples of PCR products were concentrated by ethanol precipitation both alone and in the presence of pellet paint (2.11) (Figure 6.3). Figure 6.3 clearly demonstrates that both methods of concentrating DNA are efficient, since the bands in lanes 1 and 2 are of similar intensity. Pellet paint was therefore added to all future ethanol precipitations, owing to its ease of use and the shorter incubation time.

The DNA amplified and shown in Figure 6.2 (lanes 1 and 2) was combined, phenol chloroform extracted (2.10), ethanol precipitated (2.11), 5' phosphorylated (2.12.1), ligated into pUC19 plasmid DNA (2.12.2.1) and transformed into *E. coli* (DH5α) (2.13.1). The results of the transformation are shown in Table 4.1. White colonies, which had incorporated the PCR product, were selected. The inserted synthetic DNA in each clone was amplified using pUC19 primers (see Appendix 4). Figure 6.4 shows the amplified DNA on a 3% agarose gel (2.7.1). As controls, 2 blue colonies from the control ligation plate
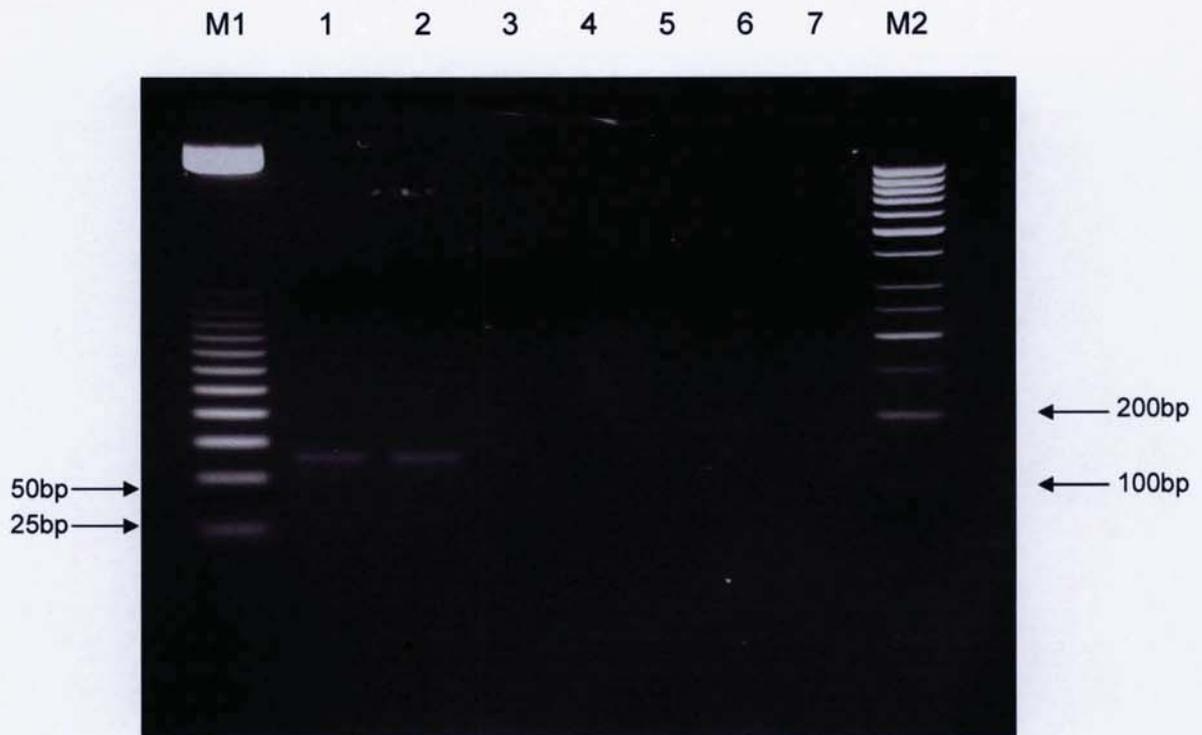
Figure 6.2: Amplification of hybridised and ligated DNA with two randomized codons, as illustrated in Figure 6.1 (10% of PCR product). Lane 1, undiluted template; lane 2 50-fold diluted template; lanes 3 and 4 negative controls lacking ligase (undiluted and diluted respectively); lanes 5 and 6 negative controls lacking Max1 and Max2 (undiluted and diluted respectively); lane 7 negative control lacking template; lane M1 25 bp DNA ladder (Promega); lane M2, 50 bp DNA ladder (MBI Fermentas)
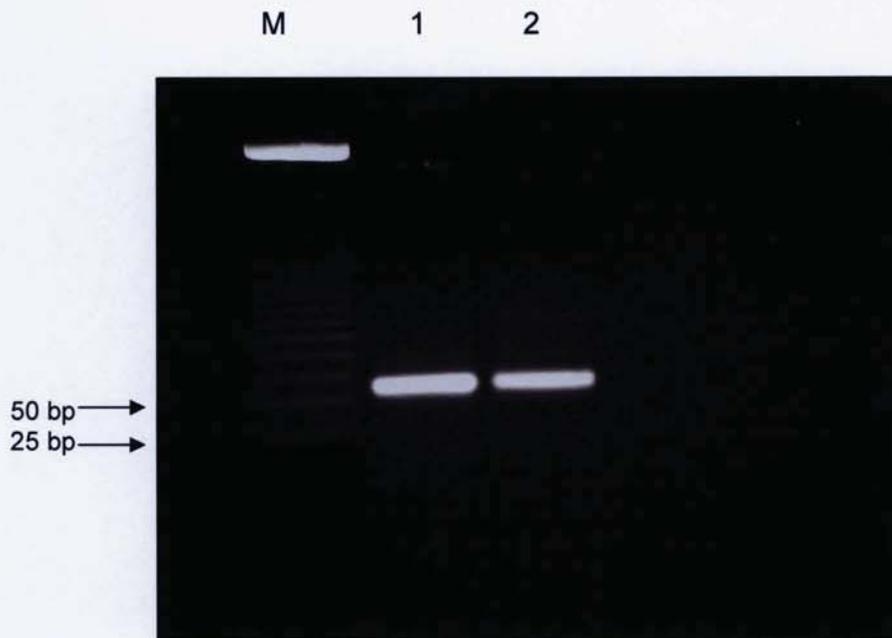
Figure 6.3: The use of pellet paint in concentrating amplified DNA. Lane 1, pellet paint precipitation; lane 2 conventional ethanol precipitation; lane M was 25bp DNA ladder (Promega).

Table 6.1: Transformation results from cloning the cassette containing two randomised positions into pUC19.

	Blue colonies	White colonies
Product ligation	78	33
Self ligation	0	0
Control ligation	430	0
Uncut pUC19	4000	0
No plasmid	0	0
No ampicillin	lawn	

Figure 6.4: PCR of colonies generated by inserting a cassette containing two randomised positions into pUC19. Lanes 1-15, 10% of PCR product from white colonies; Lane 16, negative control; lanes 17 and 18 10% of PCR product from blue colonies selected from the self ligation plate; lanes M, 50 bp marker (Biorad).

were also amplified (Figure 6.4, lanes 17&18). DNA amplified from each white colony (Figure 6.4, lanes 1-15) are all greater in size than that amplified from the blue colonies, suggesting that the synthetic insert has been incorporated into the vector successfully. The observed variation in the sizes of the amplified DNA is due to incorporation of concatemers. The PCR products from all white colonies except 11 and 16, which were of very low concentration, were all sent for sequencing to Birmingham University (2.15.2).

The sequencing data is presented in Figure 6.5 and shows that 12 out of 18 inserts have incorporated 'MAX' at both randomised positions (24 codons), 5 contain one MAX codon, 3 positions (8.3%) are ambiguous, 3 positions (8.3%) are non-"MAX" and 1 randomised codon has been deleted. Thus "MAX" incorporation has been increased to 81%, even if the ambiguous codons are assumed to be non-"MAX".

6.3 Attempted addition of a third consecutive randomised codon

The methodology described in section 6.2 has been used successfully to combine two consecutive codons. To continue this approach, the *Mly*I restriction site (Figure 6.1) was employed. *Mly*I is a relatively unusual enzyme in that it restricts the DNA outside of its recognition site. The positioning of the restriction site was such that restriction with *Mly*I should cleave the randomised DNA cassette directly after the second "MAX" codon. The anticipated method for introducing a third randomised codon was therefore to restrict the cassette containing two randomised positions with *Mly*I, denature the DNA, hybridise it

Figure 6.5: Sequencing data obtained from insertion of two randomised positions. Legend: Red, "MAX" trinucleotides; Purple, ambiguous sequences; Green, non-"MAX" trinucleotides; *, Deletion.

to a template containing three conventionally-randomised codons (Template3) and also to hybridise Max2 and MA10. The resulting DNA cassette would be amplified and cloned (Figure 6.6).

DNA was amplified from the cassette containing two randomised positions, precipitated (2.11) and digested with *Mly*I (2.12.3). The restricted DNA was then electrophoresed on a 15% PAGE (2.7.3) and stained with ethidium bromide (Figure 6.7). The amplification and subsequent digestion yielded five bands, of which the 34 bp product was excised from the PAGE and the DNA eluted (2.8.2). The eluted DNA was concentrated by ethanol precipitation (2.11) and phosphatased (2.12.4) to remove 5' phosphate groups. The resulting product was quantified.

A hybridisation and ligation reaction was set up (2.5.4) with Template3, as illustrated in Figure 6.6. The reaction was incubated overnight at 26°C and both 1μl of undiluted product a 50-fold dilution were amplified (2.6.2) with Primer4 and Primer5. Although the expected size of the amplification product was 65 bp, electrophoresis of the PCR reaction revealed only a 50 bp product (data not shown), suggesting that the method was not appropriate for incorporation of a third randomised position.

Primer5.   Max2   MA10

```
5,              3'
CTATGCCATGACTGA
5'                                              3'5'     3'5'                        3'
CTATGCCATGACTGAGATCGACACCGAAMAXMAX  MAXATGCGG ACTCCGCAGCGACCAGATGATG
        NH2CTAGCTGTGGCTTNNNNNN NNNTACGCC TGAGGCG
        3'                                          5'
                                            GTCGCTGGTCTACTAC
                                            3'              5'
```

Extracted DNA   Template3.   Primer4

Figure 6.6: Schematic representation of the incorporation of three contiguous "MAX" codons on a template strand containing 3 conventionally randomised codons. The relative positions of the hybridization oligonucleotides and the PCR primers are also shown. The sequence shown in red is the recognition site for *Mly*I.

123

Figure 6.7: Amplification and restriction of DNA cassette containing two randomised positions. Four PCR amplifications of the DNA cassette were combined, ethanol precipitated and restricted with *MlyI*. Lane M, 25bp DNA ladder (Promega); Lane 1, digested product. The smallest two bands (34 bp and 28 bp) correspond to the restricted PCR product. Undigested full length product is 62 bp. The band corresponding approximately 50bp is most likely due to the ligation of MA9 and MA10 with only one MAX containing oligonucleotide (Max1 or Max2), which would generate a of 53 bp product. The origin of the fifth band at approximately 75bp is indeterminate.

6.4 Introduction of a further randomised position without use of a randomised template strand

Re-examination of Figure 6.1 suggested an alternative approach to incorporate a third randomised codon, which would not require further use of a conventionally randomised template. After restriction with *Mly*I, a blunt-ended fragment that ended in two "MAX" codons would be generated. Thus, a double-stranded DNA cassette containing a "MAX" codon at the 5' end could simply be ligated onto the product from incorporation of 2 randomised positions (Figure 6.8).

A DNA cassette containing two "MAX" codons was generated as described previously (section 6.2). The product was restricted with *Mly*I, electrophoresed on a 15% PAGE gel (2.7.2) and the 34 bp fragment extracted (2.8.2). The extracted DNA was phosphatased (2.12.4) to remove any phosphate groups and quantified (2.9.2). Meanwhile, two complementary oligonucleotides containing a single "MAX" codon coding for tryptophan (Figure 6.9) were obtained. Max3-Trp was obtained with a 3' amino group to prevent the formation of concatamers during ligation. Max3-Trp was 5' phosphorylated (2.12.1) and then hybridised with a similar quantity of Max4-TrpRC to generate DNA of a final concentration of 50 pmol/μl (2.5.3). The hybridised product (50 pmol) was then ligated with 50 pmol of the extracted DNA containing 2 randomised positions, in a final volume of 20 μl. Control ligations omitting either extracted DNA or ligase were also set up (2.12.2.2). The ligated DNA products were each amplified (2.6.1) with Primer5 and Primer6, using 1μl of the ligated

Primer 5.

5,   3'
CTATGCCATGACTGA
5'   3'
CTATGCCATGACTGAGATCGACACCGAAMAXMAX
GATACGGTACTGACTCTAGCTGTGGCTTMAXMAX
3'   5,

Max3-Trp

(p) 5'   3'
TGGATGCGGACTCGTCGCTGGTCTACTACNH2
ACCTACGCCTGAGCAGCGACCAGATGATG
3'   5'
               CAGCGACCAGATGATG
               3'   5'

Extracted DNA

Max4-TrpRC   Primer6
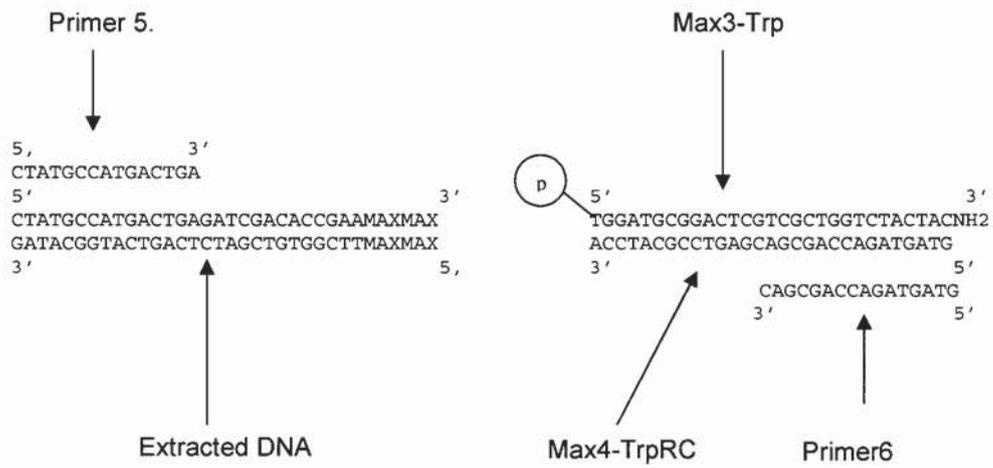
Figure 6.8: Schematic representation of the incorporation of a third randomised codon by blunt-ended ligation.
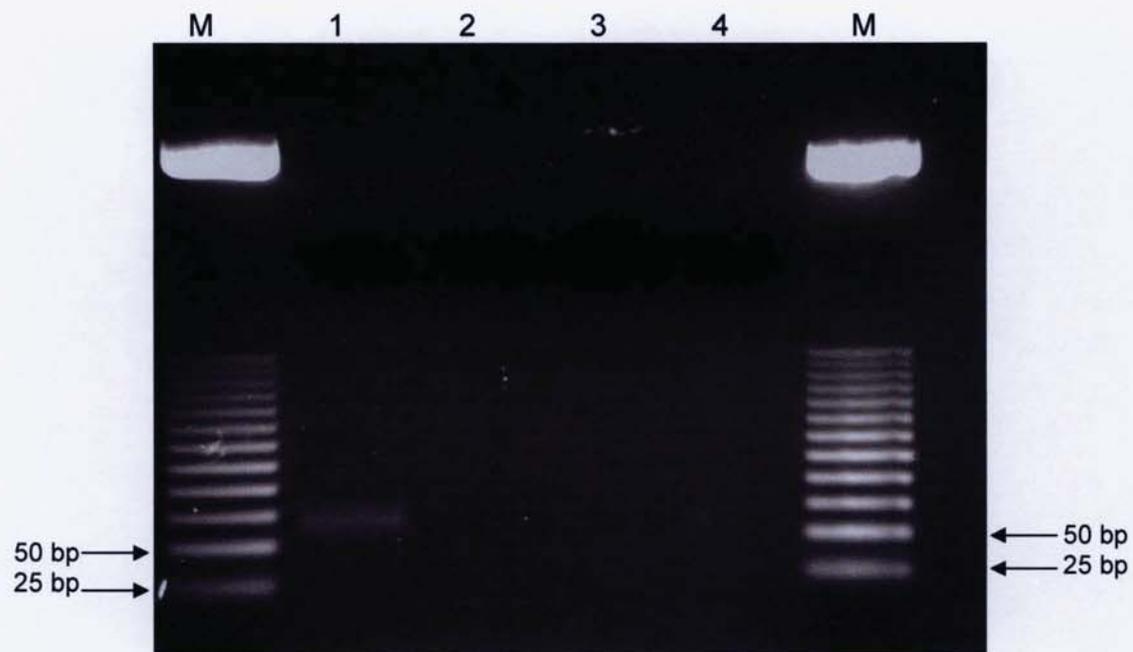
Figure 6.9: Amplification of DNA cassette containing 3 consecutive randomised positions, generated by blunt-end ligation (10% of PCR reaction). Lane 1, cassette; lane 2, -ve control lacking ligase; lane 3 –ve control lacking extracted DNA; lane 4, –ve control lacking template; lane M, 25 bp marker (Promega).

products as template. The results of the amplification were visualised on a 3% agarose (2.7.1) (Figure 6.9). The remaining DNA was then 5' phosphorylated, ligated into pUC19 (2.12.2.1) and transformed into *E. coli* DH5α (2.13.1). Of the resulting white colonies, 14 were amplified by colony PCR (2.6.2) and the DNA sent to the University of Birmingham for sequencing (2.15.2).

The resulting sequencing data demonstrated that the majority of the clones (11/15) have incorporated the third randomised codon, here fixed as TGG (Figure 6.10). Of these, 8 contained two preceding randomised codons and 3 contained only one preceding randomised codon. One clone contained ambiguous sequence and the remaining three clones contained deletions and/or rearrangements.

6.5 Addition of a fourth randomised position by blunt-ended ligation

The successful addition of a third model position was encouraging and to further test the efficacy of the method, a full set of 20 "MAX" oligonucleotides and their reverse complements, was obtained (Figure 6.11). The cassette containing three inserted codons, described in section 6.4, was amplified by PCR (2.6.1), restricted with *Mly*I (2.12.3) and purified through a 15% PAGE gel (2.7.3). The 37 bp fragment was extracted (2.8.2) the DNA phosphatased (2.12.4) and then quantified (2.9.2). The set of 20 oligonucleotides, Max5, which again contained a 3' terminal amino group were individually phosphorylated (2.12.1) and each was hybridised with its reverse complement, Max6-RC (2.5.3). The resulting 20 individual dsDNA oligonucleotides were

```
                Conserved sequence                    Randomised           Conserved sequence
                                                        codons
CTACGCCATGACTGAGATCGACACCGAA  CAT TTT TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  CAT GAT TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  AAC GAT TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  ATG AAC TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  ACC CAT TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  CAT GAT TGGCATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  ACC GAT TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  GAT ACC TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  ACC GA* TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  GAA G** TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  TAT *** TGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  AAC AAC NGG  ATGCGGACTCGTCGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  ATG GAA ***  ***CGGACTCGTAGCTGGTCTACTAC
CTACGCCATGACTGAGATCGACACCGAA  AAT GGA ***           CTCGTCGCTGGTCTACTAC
CTATGCCATGACTGAGATCGACACCGAA  AAA TTT ***  ATGCGG*CTC**CGCAGCGACCGGAT
```

Figure 6.10: Sequencing data obtained from insertion of a third, model randomised position (TGG) by blunt-ended ligation. Legend: Red, "MAX" trinucleotides; Purple, ambiguous sequence; *, deletion; Blue, rearranged sequence.

Primer 5.

Max5

```
5'                    3'
CTATGCCATGACTGA
5'                                                      3'
CTATGCCATGACTGAGATCGACACCGAAMAXMAXTGG
GATACGGTACTGACTCTAGCTGTGGCTTMAXMAXACC
3'                                                      5'
```

Extracted DNA

```
(P)  5'                              3'
     MAXCAGCGACCAGATGATGNH2
     MAXGTCGCTGGTCTACTAC
     3'                          5'
        CAGCGACCAGATGATG
        3'                      5'
```

Max6-RC   Primer4
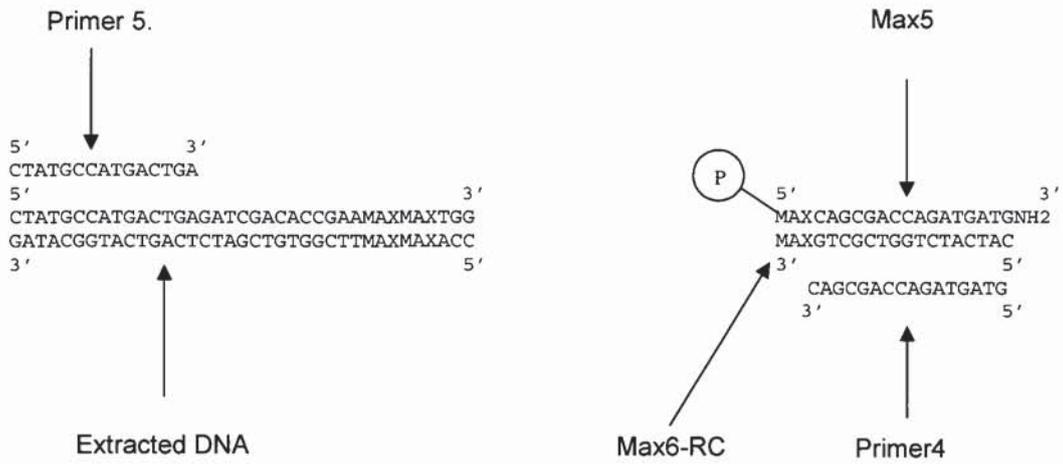
Figure 6.11: Schematic representation of the incorporation of a fourth randomised codon by blunt-ended ligation

combined to a final concentration of 50 pmol/μl total DNA. This dsDNA (50 pmol) was then ligated to 50 pmol of the DNA extracted from the PAGE gel in a final volume of 20 μl. Control ligations, as described in section 6.4, were also set up. The ligated products (1 μl) were then amplified, both neat and in 50-fold dilution (2.6.1) with Primer4 and Primer5. The resulting DNA was visualised on a 3% agarose gel (Figure 6.12). A single product resulted of the required size (58 bp) was produced as predicted, whilst all negative controls gave no product. The product visualised in lanes 1-4 was ethanol precipitated (2.11), ligated into pUC 19 (2.12.2.1) and transformed into *E. coli* DH5α (2.13.1).

Resulting white colonies were selected and amplified by colony PCR (2.6.2). A blue colony was also amplified as a control. Figure 6.13 shows the results of the colony PCR. The white colonies show a range of sizes due to the presence of concatenated inserts.

All PCR products from white colonies were sent to Birmingham University for sequencing (2.15.2). The resulting data is illustrated in Figure 6.14 and demonstrates that the fourth randomised position has been incorporated successfully. The majority of cassettes (23/25) have incorporated both TGG at the third position and a 4th randomised codon. Of these, 19 also have two randomised position inserted successfully by hybridisation. Remaining clones contain deletions and/or single base insertions, probably owing to the exposure to ethidium bromide and UV light during the electrophoresis and excision of the DNA band after visualisation on a PAGE gel or have a sequence of ambiguity.

Figure 6.12: Amplification of DNA cassette containing 4 consecutive randomised positions, generated by blunt-end ligation (10% of PCR reaction). Lane 1-2,; lane 3-4, cassette using neat template; lanes 5-6 -ve control lacking ligase (neat and dilute); lanes 7-8 –ve control lacking extracted DNA (neat and dilute); lane 4, –ve control lacking template; lane M, 25 bp marker (Promega).

Figure 6.13: Colony PCR of the DNA cassette randomised at four positions. Lanes 1-10 and 12-21 white colonies; lanes 11 and 22, blue control colonies; lanes M, 25 bp DNA ladders (Promega).

```
Conserved sequence              Randomised          Conserved sequence
                                  codons
◄─────────────────────►                      ◄─────────────────────►
TGACTGAGATCGACACCGAA    ATG AAC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    GAT ACC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TTT GAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    CAG GAT TGG CCG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TAT ACC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    AAA GAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    ATG GAT TGG GGC   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    AAC ACC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    ATG AAC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    CAT AAC TGG GGC   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TAT ACC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    AAA GAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    AAC CAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    AGC GAT TGG CGC   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TTT GAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    GAT AAC TGG CAG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TAT ACC TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    ACC AAC TGG GTG   CAGCGACCAGATGATG
GGACTGAGATCGACACCGAA    TAT GAT TGG CTG   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    TTN GAN TGG CAT   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    ATG *** TGG TTT   CAGCGACCAGATGATG
TGACTGAGATCGACACCGAA    ATAAGAT TGG GTG   CAGCGACCAGATGATG
TGACTGAGATCNACACCGAA    ATTTGAT TGG GGC   CAGCGNCCAGATGATG
TGACTGA********CCGAA    ACC GAA TT* *CG   CAGCGACCAGAATGATG
TGACTGAGATCGANNNCNNTNNTGNTTGGCCTGN        ANACGACCAGATGATG
```

Figure 6.14: Sequencing data obtained from insertion of a forth, randomised position by blunt-ended ligation. Legend: Red, "MAX" trinucleotides; Purple, ambiguous sequence; *, deletion; Blue, single base insertions.

6.6 Summary

Use of the hybridise and ligate method has improved the efficiency of the randomisation and reduced the frequency of non-specific hybridisation and ligation. The initial sequencing from the cloning of the cassette DNA generated using the randomised template showed that the non specific hybridisation and ligation was limited to only one clone (5.5%) (Clone 15, Figure 6.5). This indicated that the increase in temperature and the use of the MAX oligonucleotides instead of the trinucleotides was acting in the way it had been envisaged. The addition of the third model and fourth randomised position were successful in adding the codons. The addition of the third and fourth MAX codon using the restrict, purify and ligate method shows that the template strand is not necessary for the randomisation and the methodology can be simplified to eliminate the template strand completely.

CHAPTER 7

RANDOMISATION OF A ZINC FINGER GENE

Chapter 7: Randomisation of a zinc finger gene

7.1 Introduction

The success of the "restrict, ligate and purify" methodology described in Chapter 6 (sections 6.4-6.5) suggested that there was no need to use a conventionally randomised template at all in the saturation mutagenesis of contiguous codons. In theory, simple reiteration of the "restrict, ligate and purify" technique over seven cycles would generate a sequence of seven, consecutive randomised codons.

7.2 Restrict purify and ligate

7.2.1 Hybridisation and ligation of oligonucleotides

Oligonucleotides were obtained (MAX6, MAX6-RC, MAX7 and MAX7-RC) as shown in Figure 7.1. MAX6 and MAX6-RC were hybridised together (2.5.3). MAX7 were a set of 20 oligonucleotides, modified at the 3' terminus with an amino group, which varied from each other by three nucleotides at the 5' terminus. The differing three nucleotides in each oligonucleotide coded for each of the 20 amino acids. MAX7-RC were also a set of 20 oligonucleotides, complementary to MAX7. MAX7 oligonucleotides were individually 5'phosphorylated (2.12.1). The individually phosphorylated oligonucleotides were hybridised (2.5.3) with their reverse complement oligonucleotides
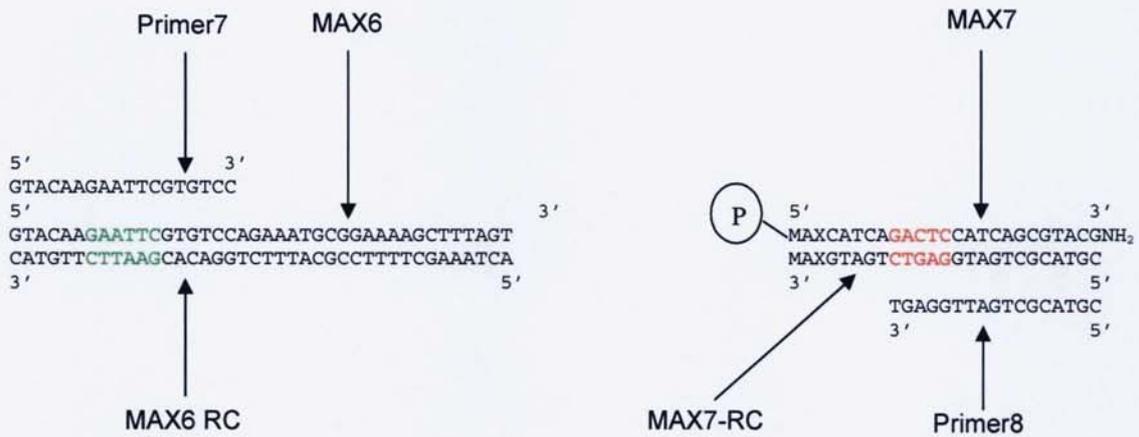
Figure 7.1: Schematic representation of the oligonucleotides in relationship with their primers. The sequence shown in red is the recognition site for *Mly*I. The sequence shown in green is the recognition sequence for *Eco*RI.

(MAX7-RC). The individual hybridised oligonucleotides were then combined so that a mix was generated containing equal concentrations (5 pm/µl) of all the 20 hybridised oligonucleotides.

Ligations were set up (2.12.2.2) between the two sets of hybridised oligonucleotides as illustrated in Figure 7.1. A control ligation was also set up that contained all the reactants as the ligation reaction but lacked the DNA ligase. Figure 7.2 shows the results of the ligation visualized on a 3% agarose gel (2.7.1). Three bands were expected in lane 2 with ligase and only 2 bands in lane 1 without the ligase. However, the presence of excess concentration of oligonucleotides has lead to the lower two bands of DNA merging into one. The gel visualised in Figure 7.2 shows clearly the presence of an extra band that is present in lane 2 (positive ligation) but is absent in lane 3 (negative ligation). The expected size of the ligated product was 65 bp and the product seen in Figure 7.2 corresponds to that size.

7.2.2 Amplification of ligation product

Two oligonucleotides were obtained from MWG and used as forward and reverse primers in the (Primer7 and Primer8, Figure 7.1). The primers were 5' phosphorylated (2.12.2.1) and amplification reactions were set up (2.6.1). The ligation reaction was diluted 50 fold and 1 µl of the diluted ligation mix used as a template. Assuming 100% efficiency, 50 picomoles of template would be generated. Figure 7.2, however, suggests that the efficiency of the ligation is probably approximately 30%. On that basis the amount of template used in each

139
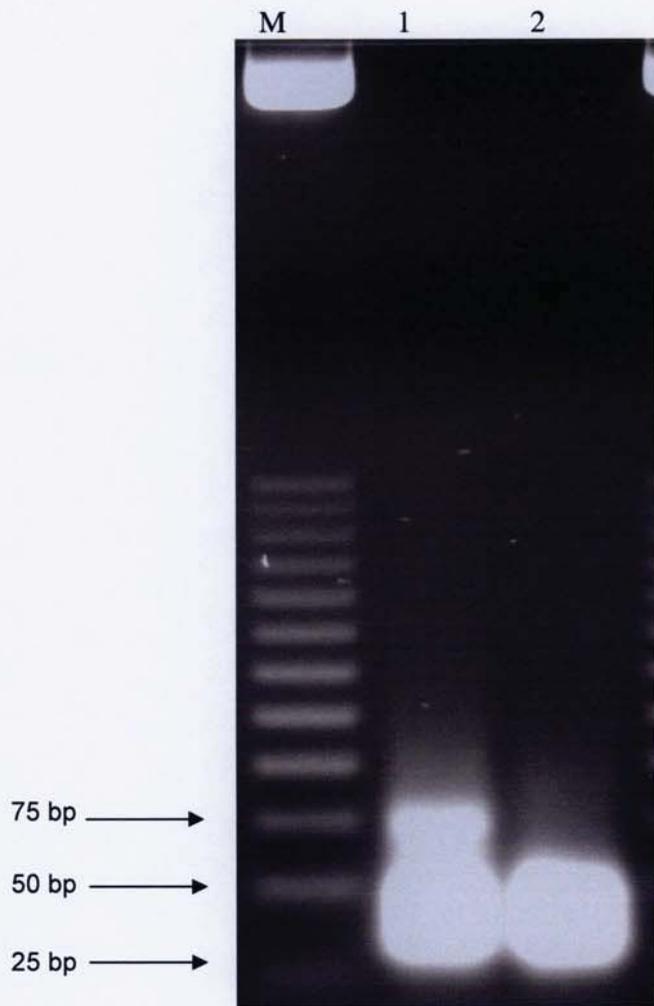
Figure 7.2: Ligation of the ZF for oligonucleotide with ZF rev oligonucleotides visualised on a 3% agarose gel. Lane M: 25bp DNA marker Promega). Lane 1: Positive ligation of oligonucleotides. Lane 2: Negative ligation of oligonucleotides.

reaction was approximately 0.015 picomoles. The results of the amplification were visualised on a 3% agarose gel (2.7.1) (Figure 7.3). Lane 1 in figure 7.3 shows a product between 50 and 75 base pairs. The expected DNA fragment was 65 base pairs. The PCR in lane 2 with the negative ligation as a template shows no band of DNA. The negative control in lane 3 also does not have any bands of DNA suggesting that the presence of the DNA in lane 1 was due to the formation of a product during the ligation reaction, which was acting as a template in the amplification reaction.

7.2.3 Restriction digest of the PCR product

The PCR product shown in lane 1 of figure 7.3 was ethanol precipitated (2.11) to concentrate the DNA. A restriction digest was set up (2.12.3) with the precipitated DNA and *Mly*I. The restriction by *Mly*I produces a blunt ended product. The position of cleavage is 5 bases pairs upstream of the recognition site. The oligonucleotides had been designed such that the *Mly*I would cleave the DNA just after the MAX codon, as described previously in Chapter 6. The restriction digest with *Mly*I generates two fragments, one 43 base pairs and the other 22 base pairs. The product of the restriction digest was visualised on a 12% denaturing PAGE (Figure 7.4).

Figure 7.4 shows three bands of DNA. The band marked 65bp is the unrestricted DNA. The restriction digest wasn't quantitative and thus unrestricted DNA remains. The other two bands are the products of the restriction digest producing a 43mer and a 22mer.
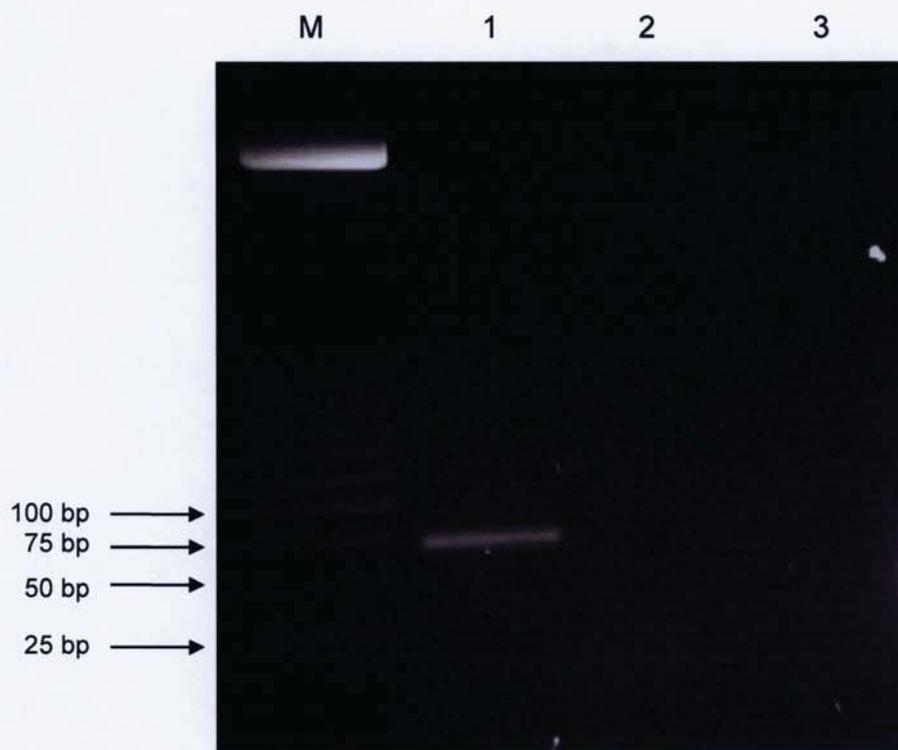
Figure 7.3: Amplification of DNA cassette containing one consecutive randomised position, generated by blunt-end ligation (10% of PCR reaction). Lane 1, amplification using 1µl of positive ligation as template; Lane 2, amplification using 1µl of negative ligation as template; Lane 3, PCR negative control; lane M, 25 bp marker (Promega).
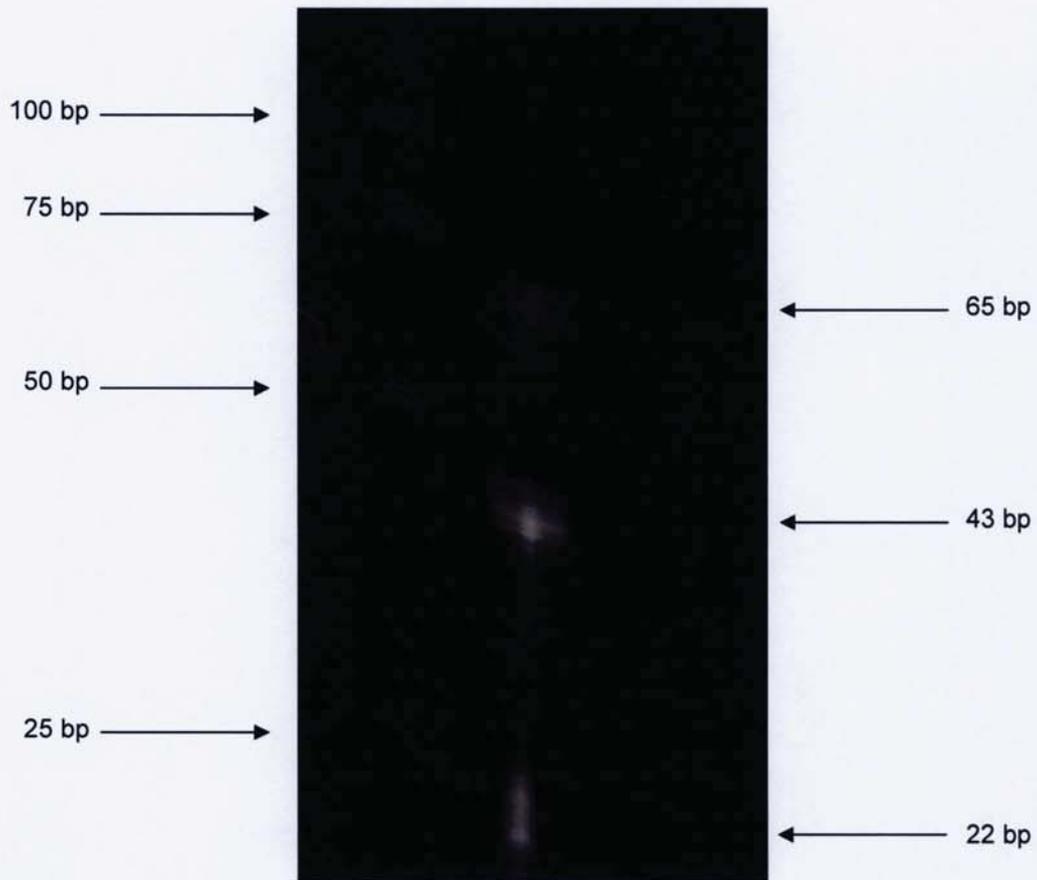
Figure 7.4: Restriction digest with *MlyI*. Lane1: 25 base pair ladder (Promega). Lane 2: Ethanol precipitated PCR reaction (50 picomoles) restricted with *MlyI*.

7.2.4 Reiteration of the restrict, ligate and purify process

The 43mer was excised from the denaturing PAGE and the DNA eluted (2.8.2). The DNA was quantified (2.9.2) and ligations were set up (2.12.2.2) with the extracted DNA and MAX7/MAX7-RC (Figure 7.5). The whole process was repeated six more times to add a total of 7 randomised trinucleotides (21 bases) to the 3' end of the original 40mer producing a DNA fragment 83 bases in length. Figure 7.6 shows the gradual increase in size of the DNA fragment from 65 bases with one position randomised to 83 bases with 7 positions randomised. Faint bands corresponding to 50 base pairs were also seen in some of the lanes. These bands correspond to the self-ligation of MAX7/MAX7-RC, which can be amplified by the Primer8 on its own. The 50 base pair product generated by the amplification has two *Mly*I sites and is broken down into two fragments of 22 bases and 28 bases.

The DNA at each stage of randomisation was ethanol precipitated and ligated into pUC19 (2.12.2.1) and transformed into *E.coli* DH5α (2.12.2.1). White colonies were selected from the transformants and sent for sequencing to the University of Birmingham (2.15.2). The results of the sequencing are shown in Figure 7.7.

Primer7   Extracted DNA   MAX7

```
5'                    3'
GTACAAGAATTCGTGTCC
5'                                              3'
GTACAAGAATTCGTGTCCAGAAATGCGGAAAAGCTTTAGTMAX
CATGTTCTTAAGCACAGGTCTTTACGCCTTTTCGAAATCAMAX
3'                                              5'
```

```
5'                              3'
MAXCATCAGACTCCATCAGCGTACGNH2
MAXGTAGTCTGAGGTAGTCGCATGC
3'                              5'
                    TGAGGTTAGTCGCATGC
                    3'              5'
```

Primer8
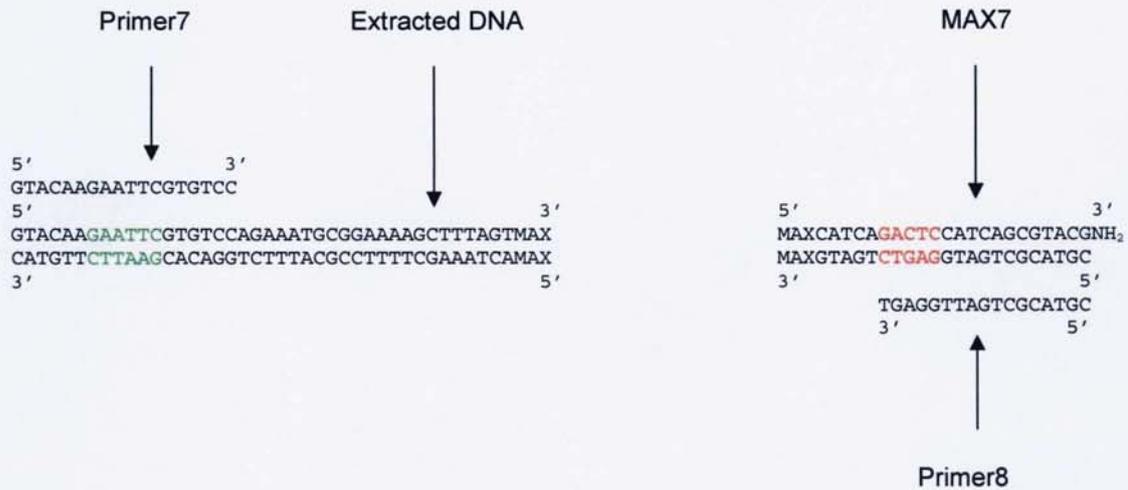
Figure 7.5: Schematic representation of the ligation of the extracted DNA with MAX7/MAX7-RC. The sequence shown in red is the recognition site for *Mly*I. The sequence shown in green is the recognition sequence for *EcoR*I.
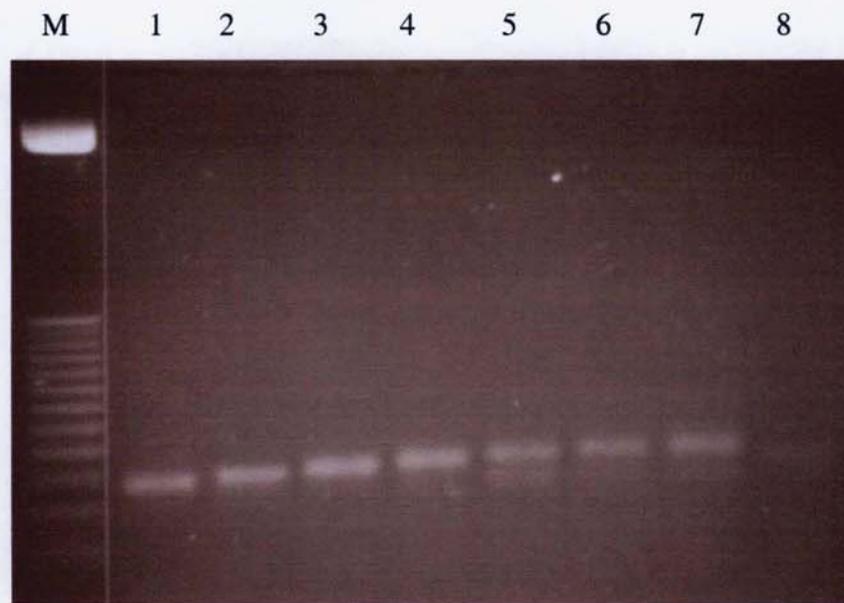
Figure 7.6: Agarose gel showing the gradual increase in size of the DNA as "MAX" trinucleotides are added.using the ligate restrict and purify method. Lane 1; DNA randomised with one "MAX" trinucleotide; lane 2, DNA randomised with two "MAX" trinucleotides; lane 3, DNA randomised with three "MAX" trinucleotides; lane 4, DNA randomised with four "MAX" trinucleotide; lane 5, DNA randomised with five "MAX" trinucleotides; lane 6, DNA randomised with six "MAX" trinucleotides; lane 7, DNA randomised with seven "MAX" trinucleotide; lane 8, PCR –ve control

Conserved sequence   Conserved sequence

```
GGGCATAAGGAAGCTTTAGT  TGC GCG  CATATGTGCATCT
GGGCATAAGGAAGCTTTAGT  GTG ACC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC CGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  ACC TAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG CAG  CATCAGACTCCAT
```

Conserved sequence   Conserved sequence

```
GGGCATAAGGAAGCTTTAGT  TAT CTG CAT  CATCAGACTCCAT
GAGCATAAGGAAGCTTTAGT  CAT CAT GCG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG CGC CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC CGC CAG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG TAT ATT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC TTT AAC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CGC AGC CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGG AAC GAA  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG GCG CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC CAT CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG TGG TTT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGT GGC CGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT TGG GTG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAG TTT TAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT TAT CGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT TGC AAA  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC CAT ATG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  ATT CAG CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGG TGC TTT  CATCAGACTCCAT
GGGNATAAGGAANCTTTAGT  TGG TAT GCG  CATCAGACTCCAT
```

Conserved sequence   Conserved sequence

```
GGGCATAAGGAAGCTTTAGT  TAT ACC TGG TGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TAT TAT AGC CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GGC CGT CAT GGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CGC CAT TGG GTG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TAT CGC TGG CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CGC CGC CTG CAT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  ATT GCC TGG TTT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  ATT ATA TTT GAT  CATCAGACTCCAT
```

Conserved sequence   Conserved sequence

```
GGGCATAAGGAAGCTTTAGT  CGC TTT TGC ACC CAT  CATCAGACTCCAT
GGGCATAAGGAAG TTTAGT   AGC ACC GTG CTG GCG  CATCGGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT TGG CAT CTG GCG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  TGC CGC AGC GAT TGC  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT CGC CGC TGC ATT  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAG GGC CAT CTG TGG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG TTT GTG TGC CAG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  GTG CAT TGC GCG GTG  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT  CAT ACC CTG ATT GGC  CATCAGACTCCAT
```

```
GGGCATAAGGAAGCTTTAGT GTG AGC CAT GGC CTG CATCAGACTCCAT
TCGGTACCCGAAGCTTTAGT GTG AGC TAT GTG TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAT TTT CGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CAT TGC ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TAT TGC CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC GGC ATT TGC CATCAGACTCCAT
GGGCATAANGAAGCTTTAGT CAT ACC CGC CGC GAA CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGG GCG CAT CAT ATG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC CAT GCG GCC CATCAGACTCCAT
```

```
GGGCATAAGGAAGCTTTAGT CGG ACC TTT GTG CAG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CGC TAT CAT ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT AAC ACC AGC CTG TGC CAG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CTG TGG ACC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAT CTG GCG AGC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT ATT CAG TGG ACC AGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAG CGC CGC ACC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TTT GCG ACC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TTT TGC CAG TGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAG CGC TGC CAT TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC CGC CGC CTG ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GGC CGC TAT GTG ATG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CAT TGC ACC ACC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TGC ATT ACC GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC CGC TAT CGC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CGC CGC TGC TAT TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAT CTG CGC ATT TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT TGC CGC CAT ATG CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC CGC CAT CTG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CG* GGC CGC ATT ACC TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG ACC G*C CGC CAT GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ATT G*C ACC TAT GAA CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG GCG CAT GTG TAC GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TTG TAT TGG GCC CTG CATCAGACTCCAT
```

```
GGGCATAAGGAAGCTTTAGT CGC AGC CGC ATT ATT TGC CGC CATCANACTCCAT
GGGCATAAGGAAGCTTTAGT TGC AGC TAT CGC CAT TGC CAG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TAT CTG GCG TTT ACC ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAG CAT AAC CGC TAT GCG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG ACC CTG ACC GTG TAT AGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC TAT GTG TTT AGC CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAG TAT CGC ATT CAT GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC TAT CGC CGC TTG AAA TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC AAC TGC CGC TTT TGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC TGC ACC CAT TTT TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TTT CGC ACC TAT ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TAT GTG CGC CAT TTT ATG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TTT CTG CGC TGG TTT ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC CGC CTG AGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT ACC ACC TGG ATG TAT CTG CAT CATCAGACTCCAT
```

```
GGGCATAAGGAAGCTTTAGT GTG TGC TAT GTG ACC CAG TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TTT ACC TTT AAA GAA CATCAGACTCCAT
NANCATAAGGAAGCTTTAGT GTG CGC TAT GTG TGC GTG CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CGC CGC CAT GAA GCG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CTG TGG GAA CGC TGC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CAT CAT CAG ATG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GGC CAT CTG GTG CAT ATT ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC ACC AGC TAT CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGG ACC CTG AAA ACC TAT TGC CATCAGACTCCNT
GGGCATAAGGAAGCTTTAGT TGC AGC CGC TGC CAT ATG TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CGC CAT TGC TTT TAT CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGG TGC CGC ATT TGG ATG GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TAT ACC CTG ACC ATT TGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC TTT CTG GCG AAA TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC TAT CTG TAT CGC GTG TTT CATCAGACTCCAT
GGGCATAAGGAAGCTATAGT CAT ACC CTG ACC CTG GGC CAG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC GTG ATT TGC TAT TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CAT TGG TTT TGC ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAG CTG ATT ACC CAT TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG ACC CTG GCG GGC TTT TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT AGC CAT TGC CAT CAG AGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAT CTG GCG ACC ACC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC TTT CGC CTG CTG AAA CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GGC CAT CTG GTG CAT ATT ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT ACC TGC ATT CGC TAT TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC CGC AGC CAT AAA ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GGC CGC TAT GAA GTG TGC CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT TAT GCG TTT TTT GAA AGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CAG ATT GGC ATT TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC CAT CAT CAG ATG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC ACC TGC TGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CTG TGC TAT CAT GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC TTT TGC CTG TTT CGC TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CGC CGC CGC CGC TGC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAT GTG TGG ACC ACC CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGG TTT ATG CAG ACC ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CTG TTT AAA ACC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC CGC TGG GAT AAA CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC CGC ACC ACC CAG ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC TAT ACC AAC CAT CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CTG CAT CTG TGG AAA TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAG TGC CGC CAT ACC GCG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG ACC ATT GGC TTT TGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GAT ACC CGC CGC GTG ATT CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC CGC TAT CGC AAA CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAT TGC CTG CTG AGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT GCG CGC ATG AGC ACC CAG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CGC TTT ACC CGC CAT CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CGC CGC CTG ACC CGC TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC TGC CTG AGC CAT TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CTG CGC CAT ATT TAT TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC ACC CGC CGC TAT CTG AAA CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC CAT TGG TAT GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT AGC CGC CAT CGC CAT TTT CATCAGACTCCAT
```

148

Conserved sequence   Conserved sequence

```
GGGCATAAGGAAGCTTTAGT GTG ACC CGT CGC ACC AAA CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAT TGC CAT GTG GAA GTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAT CGC GTG ACC GTG ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT ATT CGC TAT TGG ACC CTG AAC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT ATT CGC TAT TGG ACC CTG AAC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG ACC ATT GGC TTT TGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC CAG TGC CGC CAT ACC GCG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CTG CAT CTG TGG ATG TAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAT TGC TGC ATT TAT TGG CATCAGACTGGGT
GGGCATAAGGAAGCTTTAGT CAT CGC GTG AAC GTG GCG CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT TG* GTG CGC CGC CTG AGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CGC *TT ATT CTG AGC ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC ACC TTT CC* ACC CTG TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGT CTG TGC TGT ATT GGC CGT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGT ACT GTG GAT GTG AGC ATT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT AGT ACC AGT AAA GTG CTG CATTCATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT ATT GCC TGC CAT GTG CAT ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT GCC CAT ACC ACC CGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG GCC GC* TAT TGC CAT TGG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT GCC GC* CTG ACC GGC CTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CGC ACC CTG GCC CAT CTG CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT CAG GCG GCC TAT CAG ACC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAT TTT GCC N*T GTG CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG CAG CGC GCC GCC CAT TGC CATCAGACTCCAT
GGGCATAAGGAANCTTTAGT TGC TTT TGG ACC GAC GAC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT GGC TTT TGG GCC ACC CAT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT CAT ACC CTG CCG CGA GTG TGT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC TAT TGC TGC CAT ACG CAG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC ACC TGC CTG CGT CGC CATNANAATNCNT
GGGCATAAGGAAGCTTTAGT GTG CAT GTG ACC TAT GCC TGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG TGC TAT GTG TG* TGC CGC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CAG TTT ACC ATG GCC CTG CATCAGACTCCAT

GGGCATAAGGAAGCTTTAGT TTT GNC TG                  CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TAT NNCCN GTG CAG NNGCCNC CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT GTG GCC CGC ANC CTG NCC TTT CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CNCNTGNCCACNNNNNTG CATCAGACTCCAT
GGGCATAAGGAAGCTTTAGT TGC CGC CGC   NGNCAANTNAACNGNCANACTACTCCAT
```

Figure 7.7: Sequencing data obtained from insertion of two, three, four, five, six and seven, randomised position by blunt-ended ligation. Legend: Red, "MAX" trinucleotides; Purple, ambiguous sequence; *, deletion; Blue, single base insertions, Green, Non-MAX trinucleotides. The last four sequences were not included in the data analysis due to the large number of ambiguous nucleotides.

Table 7.1: Frequency of the MAX trinucleotides represented in pUC 19 after ligation with the randomised synthetic DNA.

Codon Single letter code	Codon Triple letter code	Frequency	Percentage	Expected percentage
A	Ala	25	2.40	5
C	Cys	111	10.64	5
D	Asp	5	0.48	5
E	Glu	32	3.07	5
F	Phe	59	5.66	5
G	Gly	20	1.92	5
H	His	124	11.89	5
I	Ile	40	3.84	5
K	Lys	13	1.25	5
L	Leu	65	6.23	5
M	Met	14	1.34	5
N	Asn	9	0.86	5
P	Pro	1	0.10	5
Q	Gln	9	0.86	5
R	Arg	152	14.50	5
S	Ser	27	2.59	5
T	Thr	90	8.63	5
V	Val	93	8.92	5
W	Trp	41	3.93	5
Y	Tyr	67	6.42	5
Non MAX		36	3.45	0
Deletions		10	0.96	0
Total		1043	100	100

7.3 Sequencing Data

The sequencing data shows the incorporation of MAX trinucleotides was at a level greater then 95 % (Table7.1). Every single amino acid was represented but the level of representation varied for each amino acid. The variation ranged from 14.5% for arginine (the most popular amino acid coded) to 0.1% for proline. The non-MAX codons accounted for 3.45% of the total analysed sequences with less then one percent of the codons having a single point mutation.

The one proline residue that has appeared in the sequencing was probably due to a mutation rather then due to the mutagenesis. On analysing the data to understand if there was a flaw in the methodology that generated only one proline residue, it was discovered that the MAX oligonucleotides that coded for proline had invariantly been synthesised incorrectly. Instead of the MAX region being coded for by 5'CCG3' it was coded for by 5'GCC3'. The reverse complement strand was correctly generated for the proline codon. This lack of the correct sequence of DNA meant that the two complementary strands could not hybridise together and the sequence for proline did not exit.

Figure 7.9 shows the actual sequence used for proline. Because the 5'GCC3' was in the coding strand, this explains the high presence of this particular sequence in the sequencing data. The presence would probably have been greater if the reverse complement had been complementary to the actual codon used rather then complementary to proline.

5' GCCCATCAGACTCCATCAGCGTACG3'

3' GGCGTAGTCTGAGGTAGTCGCATGC5'

Figure 7.9: The sequence of the proline oligonucleotide. The sequence in red is the sequence that would be ligated to the growing sense strand and for proline should have read 5'CCG3'. The sequence in red is the correct sequence for the reverse complement to proline sequence but is not complementary to the sequence actually used.

7.4 Summary

In conjunction, Figures 7.6 and 7.7 demonstrate that reiteration of the "restrict, ligate and purify" methodology represents a suitable strategy for the non-redundant randomisation of multiple, consecutive codons.

CHAPTER 8

DISCUSSION

Chapter 8: Discussion

8.1 Introduction

The aim of this study, to develop a non-redundant, biological method of saturation mutagenesis which could be applied to multiple, contiguous codons has been achieved successfully. The research has achieved the goal of addition of contiguous trinucleotides using enzymatic action as well addressed some of the problems associated with saturation mutagenesis of DNA.

8.2 Single stranded ligation of "MAX" trinucleotides using RNA ligase

The use of RNA ligase in the randomisation technique did not achieve the results that had been anticipated. The single stranded ligation was anticipated to ligate trinucleotides in between two oligonucleotides, which would then be used to amplify the ligated DNA. It was also anticipated that the number of codons that could ligate in between the two conserved regions would be unlimited. The numbers of trinucleotides that would be ligated into the conserved region would then be separated from each other by the use of PAGE. The differing bands would indicate the differing number of trinucleotides that had been ligated. The different bands of DNA could then be extracted from the PAGE gel, amplified and then ligated into a vector. Tessier and co workers (1986) achieved quantitative levels of ligation of oligonucleotides. Unfortunately my work failed to achieve the same levels of ligation as *Tessier et al.*, We have shown that the T4 RNA ligase does work in ligating DNA but the ligation was far from quantitative. The best that we achieved using T4 RNA ligase was enough to

amplify using PCR but not enough to be visualised by itself on a agarose gel. The research tried to emulate the work of Tessier *et. al.* by using the same buffers and by changing the sequence of the DNA strands. Our research used an amino group at the 3' end to prevent the formation of concatamers whereas Tessier used terminal transferase to add dideoxynucleotides to the 3' terminus of the donor strand to prevent concatemerisation.

8.3 Generation of a randomized DNA sequence on oligo affinity support

The oligo affinity support was used as a base to generate a template strand onto which trinucleotides coding for the 20 amino acids could be hybridised and ligated. The oligo affinity support was to be used to remove the template strand away from the hybridised and ligated strand. The amplification of the DNA suggested that the hybridisation and ligation was taking place.

The sequencing data showed that the presence of the trinucleotides was at a level far lower then would be expected even if the sequence of DNA had been randomly generated. The low level of MAX trinucleotides suggested that the amplification was occurring of the template strand rather then the ligated strand that we required. The hybridisation oligonucleotides were redesigned to circumvent the problem associated with the leeched template strand. This redesign of the hybridisation did not resolve the problem of the non-appearance of MAX codons in the sequencing data.

Work done by Hughes *et. al* (2003*)*. on a zinc finger gene showed that the amplification was occurring due to the presence of the template strand leeching through into the amplification reaction. Hughes and co-workers showed that a product would appear even when the trinucleotides were not added to the ligation reaction. The template strand was able to partake in the amplification reaction. It had been anticipated that the oligo affinity support could be used to eliminate the template strand from the amplification. When it was discovered that the template strand was leeching through and being amplified, it prevented the selection of the strand ligated with trinucleotides as the DNA of choice. The method was not further developed to analyse how efficient the ligation of trinucleotides was because of the unability to prevent the amplification of the template strand.

8.4 Cloning of DNA sequence randomized at two amino acid positions

The oligo affinity support was abandoned when it was realized that it was not possible to separate the template strand on the oligo affinity support from the ligated. The design of the template and hybridisation oligonucleotides was also changed such that the template strand was randomised at only two positions. The hybridisation oligonucleotides were also extended such that the PCR primers could not hybridise to the template and thus were not able to amplify it. The methodology showed promise in that trinucleotides were found in the randomised region but due to the use of low tempretures during hybridisation and ligation, non specific ligation was occurring.

8.5 Double stranded ligation of "MAX" oligonucleotides

The modification of the methodology to prevent non specific hybridisation was also a success but was limited to randomising two contiguous positions. The blunt ended ligation of MAX codons was also a success which was used to add another two codons contiguously making a total of four.

8.6 Randomisation of a zinc finger gene

During the methodology described in chapter 6, it was realised that the presence of the template strand was not necessary. The blunt ended ligation was developed and was used to successfully randomise 7 contiguous positions of randomisation in a synthetic zinc finger as shown in figures 7.6 and 7.7.

8.7 Where the research will lead

The methodology has been proved to work in saturation mutagenesis of 7 contiguous codons and can be used to mutagenised any number of codons within the ability of the constraints laid down by the problems with cloning. The one minor drawback that needs to be addressed is the problem of representation. This is a problem that is not just limited to this methodology but was a problem in the chemical methodologies using trinucleotides. Further application of the methodology described therein will determine if the same

codon representation occurs or whether it changes with the use of new oligonucleotides.

The amino groups in the oligonucleotides were used in this study to prevent concatamer formation in the initial ligations between the oligonucleotides (Figure 7.1). However these are not the only groups that can be used. Tessier *et al.*, (1986) used dideoxynucleotides to prevent the formation of concatamers. Thiol groups and biotin labelling can also be used to achieve the same results. The extension of either MAX7 or MAX7-RC (Figure 7.1) is also postulated to prevent the formation of concatmers due to the presence of a region of single stranded DNA at the point of ligation.

The work could be taken further by using the cassette generated and cloning into the zinc finger gene and then assessing the binding affinity of the protein generated by the saturation mutagenesis of the binding residues and the amino acids in between and see if these play any supporting role in the protein DNA interaction. The DNA cassette was inserted into pUC 19 using blunt ended cloning. This worked to our advantage at this point because the insertion of multiple cassettes into one clone would let us analyse more sequences then if only one cassette had been inserted. The insertion of multiple cassettes in an expression vector would lead to detrimental consequences. To overcome this minor problem, restriction sites can be employed as demonstrated by Hughes *et al.*, (2003), to ensure that only a single cassette will be inserted and it would insert itself in the correct orientation. Finally, like MAX randomisation, it

should be possible to use the methodology described herein to generate required

subsets of amino acids at specified positions.

REFERENCES

References

Arkin, A. P. Youvan, D. C., (1992). "Optimising Nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis". *Biotechnology*, **10**, 297-300

Barnes, W. M., (1992). "The fidelity of *Taq* polymerase catalysing PCR is improved by an N-terminal deletion". *Gene*, **112**, 1, 29-35

Beato, M., (1989). "Gene regulation by steroid hormones". *The Cell*, **56**: 335-344

Berg, J. M., (1988). "Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins". *Proc. Natl. Acad. Sci. (USA).* **85(1)**:99-102

Berg, J. M. (1995). "Zinc finger domains: from prediction to design". *Acc. Chem. Res*, **28**. 14-19

.

Braunagel, M., Little, M., (1997). "Construction of a semisynthetic antibody library using trinucleotide oligos". *Nucleic Acids Research.* **Vol. 25. No. 22:** 4690-4691

Brinkmann, U., Mattes, R. E., Buckel, P., (1989). "High level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the DNA Y gene product". *Gene.* **85:** 109-114

Cadwell, R. C., Joyce, G. F., (1994). "Mutagenic PCR". *PCR methods appl.* **3:** s136-s40

Cariello, N. F., Swenberg, J. A. and Skopek, T. R., (1991). "Fidelity of Thermococcus Litoralis DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis". *Nucleic acids research.* **19**: 4193-4198

Cirino, P. C., Mayer, K. M., Umeno, D., (2003). "Generating mutant libraries using error-prone PCR". *Methods Mol. Biol.* **231**: 3-9

Collins, J., Horn, N., Wadenback, J., Szardenings, M., (2001). " Three conserved transcriptional repressor domains are a defining feature of the TIEG subfamily of SpI – like zinc finger proteins". *Journal of Biological Chemistry.* **274:** 29500-29504

Craig, N. I., (1995). "Unity in transposition reactions". *Science.* **270:** 253-254

Crick, F. H. C., (1966). "Codon-anticodon pairing: the wobble hypothesis". *J. Mol. Biol.* **19:** 820 – 826

Del Rio, G., Osuna, J. and Soberon, X., (1994) "Combinatorial libraries of proteins: analysis of efficiency of mutagenesis techniques". *Biotechniques.* **17**: 1132-1139

Demple B, Jacobsson A, Olsson M, Robins P, Lindahl, T. (1982). "Repair of alkylated DNA in Escherichia coli. Physical properties of O6-methylguanine-DNA methyltransferase". *J Biol Chem.* **257(22)**:13776-80

Di Giulio, M., (1989). "Some aspects of the organization and evolution of the genetic code". *J. Mol. Evol.* **29(3)**:191-201

Freeland, S. J., Hurst, L. D., (1998). "The genetic code is one in a million". *J. Mol. Evo.l* **47**: 238-248

Gaytán, P., Yanez, J., Sanchez, F., Mackie, H. and Soboron, X, (1998). "Combination of DMT-mononucleotide and Fmoc- trinucleotide phosphoramidites in oligonucleotide synthesis affords an automated codon-level mutagenesis method". *Chemistry and Biology.* **5 (9)**: 519-527

Gaytán, P., Osuna. J., Soberon, X., (2002). "Noval ceftazidime-resistance β-lactamases generated by codon-based mutagenesis method and selection". *Nucleic acids research,* **Vol. 30 No. 16**: e84

Gaytán, P., Yanez, J., Sanchez, F. and Soboron, X., (2001) Orthogonal combinatorial mutagenesis: a codon- level combinatorial mutagenesis method useful for multiplicity and amino acid-scanning protocols, *Nucl. Acids. Res.*, **29 (3):** e9

Gillam, S., Astell, C. R. and Smith, M., (1980). "Site-specific mutagenesis using oligodeoxyribonucleotides: isolation of a phenotypically silent φX174 mutant, with a specific nucleotide deletion, at a very high efficiency". *Gene*, **12:** 129-137

Goldberg, A. L., Wittes, R. E., (1966). "Genetic code: aspects of organization". *Science,* **153:** 420-424

Grishan, N. V., (2001). "Fold change in evolution of protein structures". *J. Struct. Biol.,* **134:** 167-185

Goodman, M. F. (2002). "Error prone repair DNA polymerasesin prokarotes and eukaryotes". *Annu. Rev. Biochem.* **71:** 17-50

Gunneriusson, E., Nord, K., Uhlen, N., Nygren, P. A., (1999). " Affinity maturation of a *Taq* polymarase specific antibody by helix shuffling". *Protein Engineering.* **12:** 873-878

Hannig, G., Makrides S. C., (1998). "Strategies for optimizing heterologous protein expression in Escherichia coli". *Trends Biotechnol.* **16(2):**54-60

Harrison, S.C., Aggarwal, A.K., (1990). "DNA recognition by proteins with the helix-turn-helix motif". *Ann. Rev. Biochem.* **59:** 933-969

Haapa, S., Taira, S., Heikkinen, E., Savilahti, H., (1999). "An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications". *Nucleic Acids Research.* **27:** 2777-2784

Huang, W., Santi, D.V., (1994) "Identification of biologically active mutants by combinatorial cassette mutagenesis: Exclusion of wild-type codons from degenerate codons." *Anal. Biochem.* **218:** 454-457

Hemmings-Mieszczak, M., Dorn, G., Natt, F. J., Hall, J., Wishart, W. L., (2003). "Independent combinatorial effects of antisense oligonucleotids and RNAi-mediated specific inhibition of the recombinant rat P2X$_3$ receptor". *Nucleic Acids Research.* **31 (8):** 2117-2126

Hoffman, S. L., Subramanian, G. M., Collins, F. H., Venter, (2002). "Plasmodium, human and Anopheles genomics and malaria". *Nature.* **415:** 702-709

Ho, S. P., Brittons, D. H. O., Stone, B. A., Behrens, D. L., Leffet, L. M., Hobbs, F. W., Miller, J. A. and Trainer, G. L., (1996). "Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA are rationally selected by mapping RNA-accessible sites with oligonucleotide libraries". *Nucliec Acids Research.* **24:** 1901-1907

Hughes, M. D., Nagel, D. A., Santos, A. F., Sutherland, A. J. and Hine, A. V., (2003). "Removing the redundancy from randomised gene libraries". *Journal of Molecular Biology.* **331:** 973-979

Hugo, W.B., Russel, A. D, "Pharmaceutical Microbiology", *Blackwell Science*, London p453

Jamieson, A. C., Kim, S. H, Wells, J. A., (1994). "In vitro selection of zinc fingers with altered DNA-binding specificity". *Biochemistry.* **33**: 5689-5695

Jones, D., D., (2005). "Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β-lactamase to an amino acid deletion". *Nucleic Acids Research*, **33(9):** e80

Jones, O. W., Nirenberg, M. W., (1966). "Degeneracy in the amino acid code." *Biochim. Biophys. Acta.* **119:** 400-406

Kane, J, (1995). "Effects of rare codon cluster on high level expression of heterologous protein in *Escherichia coli*". *Curr. Opin. Biotechnol.* **6:** 494-493

Kayushin, A. L., korosteleva, M. D., Miroshnikov, A. I, Kosch, W., Zubov, D. and Piel, N. (1996). "A convenient approach to the synthesis of trinucleotide phosphoroamadites-synthons for the generation of oligonucleotide/peptide libraries". *Nucleic Acids Research.* **24(19):** 3748-3755

Keohavong, P. and Thilly, W. G., (1989). "Fidelity of DNA polymerases in DNA amplification". *Proc. Natl. Acad. Sci.* (USA). **86 (23):** 9253-9257

Kerppola, T. K. and Curran, T., (1991). "Transcription factor interactions: basics on zippers". *Current opinions in structural biology.* **1**: 71-79

Kitamura, K., Kinoshita, Y., Nemoto, N., Husimi, Y., Nishigaki, K., (2002). "Construction of a block-shuffled libraries of DNA for evolutionary protein engineering: Y-ligation-based block shuffling". *Protein Engineering.* **15**: 843-853

Klug, A. and Rhodes, D., (1987) "Zinc finger: A novel protein motif for nucleic acid recognition". *Trends in biochemical science.* **12**: 464-469

Knight, R. D., Freeland S. J., Landweber, L. F., (2001). "Rewiring the keyboard: evolability of the genetic code". *Nat. Rev. Genet.* **2**: 49-58

Kumar, G., Poonian, M. S. (1984). "Improvements in oligodeoxyribonucleotide synthsis: methyl *N, N*-dialkylphosphoramadite dimmer units for solid support phosphate methodology". *J. Org. Chem.* **49**: 4905-4912

Kurland, C. Gallant, J. (1996). "Errors of heterologous protein expression". *Curr. Opin. Biotechnol.* **7:** 489-493

Lehmann, A. R., (2002). "Replication of damaged DNA in mammalian cells: new solutions to old problems". *Mutant. Res.* **509:** 23-34

Lehmann, C., Xu, Y., Christodoulou, C., Tan, Z., and Gait, M., (1989). "Solid-phase synthesis of oligoribonucleotides using 9-flourenylmethoxycarbonyl (fmoc) for 5'–hydroxyl protection". *Nucleic Acids Research.* **17(7):** 2379-2390

Lima, W. F., Brown-Driver, V., Fox, M., Hanecak, R., Bruice, T. W., (1996). "Combinatorial screening and rational optimisation for hybridisation to folded hepatitas C cirus RNA of oligonucleotides with biological antisense activity". *Journal of Molecular Chemistry.* **272 (1):** 626-638

Lundberg, K. S., Shoemaker, D. D., Adams, M.W., Short, J.M., Sorge, J. A. and Mathur, E.J., (1991). "High- fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*". *Gene.* **108 (1):** 1-6

Lyttle, M. H., Napolitano, E. W., calio, B. l., Kauvar, L. M., (1995). "Mutagenesis using trinucleotide beta-cyanoethyl phosphoramadites". *Biotechniques.* **19:** 274-281

Magliery, T. J., Regan, L., (2004). "Combinatorial approaches to protein stability and structure". *Eur. J. Biochem.* **271:** 1595-1608

Matsuura, T., Ernst, A., Pluckthun, A., (2002). "Construction and characterisation of protein libraries composed of secondary structure modules". *Protein science.* **11:** 2631-2643

McClintock, B., (1950). "The Origin and Behavior of Mutable Loci in Maize". *Proc. Natl. Acad. Sci.* (USA) **36(6):** 344–355

.

Miller, J. H., Low, K. B., (1984), "Specificity of Mutagenesis resulting from the induction of the SOS system in the absence of mutagenic treatment". *Cell.* **37:** 675-682

Miller, J., McLachlan, A. and Klug, A., (1985). " Repetitive zinc-finger binding domains in the protein transcription factor IIIA from *Xenopus* oocytes". *Journal of embroyology.* **4:** 1609-1614

Mizuuchi, K. (1992). "Transpositional recombination: mechanistic insights from studies of Mu and other elements". *Annu. Rev. Biochem.* **61:** 253-254

Mossner, E. and Pluckthun, A., (2001). "Directed evolution with fast and efficient technologies". *CHIMIA.* **55 (4):** 324-328

Murakami, H., Hohsaka, T. and Sisdo, M., (2002). "Random insertion and deletion of arbitary number of bases for codon-based random mutation of DNAs". *Nat. Biotch.* **20**: 76-81

Nakamura, Y., Gojobori, T. and Ikemura, T. (1997). "Codon usage tabulated from the international DNA sequence database". *Nucleic Acids Research.* **222:** 851-856

Nakamura, Y., Gojobori, T. and Ikemura, T. (1999). "Codon usage tabulated from the international DNA sequence databases; its status 1999". *Nucleic Acids Research.* **27**: 292

Neylon, C., (2004). "Chemical and biochemical stratagies for the randomisation of protein encoding DNA sequences: library construction methods for directed evolution". *Nucleic Acids Research.* **32 (4):** 1448-1459

Neuner, P., Cortese, R., Monaci, P., (1998). "Codon-based mutagenesis using dimmer-phosphoramidites". *Nucleic Acids Research.* **26 (5):** 1223-1227

Ono, A., Matsuda, A., Zhao, J., Santi, D. V., (1995). "The synthesis of blocked triplet-phosphoramadites and their use in mutagenesis". *Nucleic Acids Research.* **23 (22):** 4677-4682

Osuna, J., Yáñez, J., Soberón, X., and Gaytán, P., (2004). "Protein evolution by codon-based random deletions". *Nucleic Acids Research*.: e132(17): 36

Pabo, C. O. and Sauer, R. T., (1992). "Transcription factors: structural families and principles of DNA recognition". *Ann. Rev. Biochem.* **61**: 1053-1095

Parikh, A., Guengerich, F. P., (1997). "Random mutagenesis by whole plasmid PCR amplification". *Biotechniques.* **24:** 428-431

Pascarella, S., Argos, P., (1992). "Analysis of insertion/deletions in protein structures". *J. Mol. Biol.* **224:** 461-471

Patzel V, Sczakiel G, (2000). "*In vitro* selection supports the view of a kinetic control of antisense RNA-mediated inhibition of gene expression in mammalian cells". *Nucleic Acids Research.* **28(13):** 2462-6

Pavletich, N. P. and Pabo, C. O., (1991). "Zinc finger DNA recognition – crystal structure of a Zif268-DNA complex at 2.1A". *Science.* **252**: 809-817

Petrenko,V. A., Smith, G. P., Mazooji, M. M., Quinn, T., (2002). "α-helically constrained phage display library". *Protein Engineering.* **15:** 943-950

Pfeifer, G. P., (1997). "Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment". *Photochem Photobiol.* **65 (2) :** 270-83

Pfeifer, G. P., You, Y. H. and Besaratinia. A., (2005) "Mutations induced by ultraviolet light". *Mutation Research,* **571**: 19-31

Reidhaar-Olson, J. F., Sauer, R. T., (1988). "Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences". *Science.* **241(4861):** 53-7

Roache, R. D., Sauer, R. T., (1999). "SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity". *Journal of embroyology.* **18**: 4579-4589

Sambrook, J., Frisch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.p11.21

Saul, A., Battistutta, D., (1988). "codon usage in *Plasmodium falciparium*". *Mol. Biochem. Parasitol.* **27**: 35-42

Scott, J. K., & Smith, G. P., (1990). "Searching for peptide ligands with an epitope library". *Science.* **249**: 386-390

Seeman, N. C., Rosenberg, J. M. and Rich, A., (1976). "Sequence specific recognition of double helical nucleic acids by proteins". *Proc. Natl. Acad. Sci. (USA)*. **73**: 804-808

Shafikhani S, Siegel RA, Ferrari E, Schellenberger V (1997). "Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization".*Biotechniques*. **23 (2):** 304-10

Singer, B., Kusmeirek, J. T., (1982), "Chemical mutagensesis". *Ann. Rev. Biochem.* **52**: 655-693

Sonderlind, E., Vergeles, M., Borrebaeck, C. A. K., (1995). "Domain libraries: synthetic diversity for de novo design of antibody V-regions". *Gene* **160**: 269-272

Sondek, J. and Shortle, D. (1992). "A general strategy for random insertion and substitution mutagenesis: Substoichiometric coupling of trinucleotide phosphoroamadites". *Proc. Natl. Acad. Sci.(USA)*. **89**: 3581-3585

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., Inouye, M., (1966). " Frameshift mutations and the genetic code". *Cold Spring Harb. Symp. Quant. Biol.* **31**:77-84

Tessier. D. C., Brousseau, R. and Vernet, T. (1986). "Ligation of single stranded oligodeoxyribonucleotides by T4 RNA ligase". *Anal. Biochem.* **158**: 171-178

Virnekas, B., Ge, L., Pluckthun, A., Schneider, K. C., Wellnhofer, G. and Moroney, S. E. (1994). "Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis". *Nucleic Acids Research.* **22(25):** 5600-5607

Woese, C. R., (1965). "On the evolution of the genetic code". *Proc. Natl. Acad. Sci. (USA)* **54:** 1546-1552

Wong, J. T., 1975. "A evolutionary theory of the genetic code". *Proc. Natl. Acad. Sci. (USA).* **72:**1909-1912

Wong, J. T., 1980. "Role of minimisation of chemical distances between amino acids in the evolution of the genetic code". *Proc. Natl. Acad. Sci. (USA).* **77:**1083-1086

Yáñez, J., Arguello, M., Osuna, J., Soberon, X., Gaytán, P., (2004). "Combinatorial codon-based amino acid substitutions". *Nucleic Acids Research.* **32 (20):** e158

Zaccolo, M., Williams, D. M., Brown, D. M., Gheradi, E., (1996). "An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues". *J. Mol. Biol.* **255**: 589-603

Zhang, S., Zubay, G., Goldman, E., (1991). "Low usage in Escherichia coli, yeast, fruit fly, and primates". *Gene.* **105**: 61-72
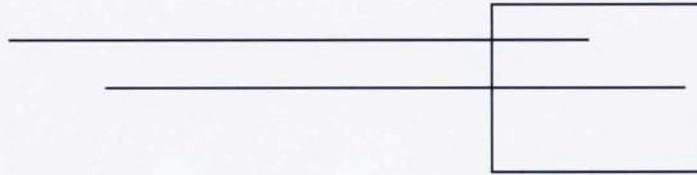
Zhan, K. (1996)."overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth". *Journal of Bacteriology.* **178**: 2926-2933

Zhou, Z., Schnake, P., Xiao, L., Lal, A. A., (2004). "Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimisation". *Protein expression and purification.* **34**: 87-94

APPENDICES

A: Transposon

5' <u>TGCGCC</u>GAGTCA3'
3' <u>ACGCGG</u>CTCAGTCTAG5'

B : Target DNA

......$N_1N_2N_3N_4N_5$....
......$N_1N_2N_3N_4N_5$....

C: insertion of transposon

......   TGACTC----GAGTCA$N_1N_2N_3N_4N_5$......
......$N_1N_2N_3N_4N_5$ACTGAG----CTCAGT   ......

D: Gap filling

......$N_1N_2N_3N_4N_5$TGACTC----GAGTCA$N_1N_2N_3N_4N_5$....
......$N_1N_2N_3N_4N_5$ACTGAG----CTCAGT$N_1N_2N_3N_4N_5$....

The position of the restriction site for MlyI (shown in red). The sequence underlined is the conserved sequence of DNA responsible for the MuA binding (Craig, 1995). The 4 base pair overhangs are removed on insertion into the plasmid. Figure A shows the DNA sequence at one end of the transposon with the 4 base pair overhang. Figure B shows the schematic representation of the target DNA which is restricted to provide sticky ends. The transposon then inserts itself into the restricted DNA (C), and the gaps repaired (D). The insertion of the transposon results in a 5 base pair duplication after the gaps have been repaired (Mizuuchi, 1992).

A generalised scheme of the random insertion/ deletion (RID) mutagenesis for the construction of a library of mutant genes. The procedure consists of eight major steps as follows:

1. The fragment obtained by digesting the original gene with *EcoRi* and *HindIII* is ligated to a linker.
2. The gene fragment is cyclized with T4 DNA ligase to make a circular dsDNA with a nick in the antisense chain.
3. The circular dsDNA is treated with T4 DNA polymerase to produce a circular ssDNA.
4. The circular ssDNA is randomly cleaved at single positions by treating with Ce(iv)-EDTA complex.
5. The linear ssDNAs, which have unknown sequences at both ends, are ligated to the 5'-anchor and the 3'-anchor, respectively.
6. The DNAs that are linked to the two anchors at both ends are amplified by PCR.
7. The PCR products are treated with *BciVI*, Leaving several bases from the 5'-anchor, at the 5'-end. The *BciVI* treatment also deletes a specific number of bases at the 3'-end.
8. The digested products are treated with Klenow fragment to make blunt ends and cyclized again with T4 DNA ligase. The products are treated with *EcoRI* and *Hindiii*, and the fragments are cloned into an *EcoRI-Hindiii* site of modified Puc18 (pUM).

Taken from the paper by Murakami *et al.*, (2002).

Appendix 3

AMINO ACID CODES

Amino acid	Triple letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

APPENDIX 4

pUC 19 reverse PCR primer.

5' CGGGCCTCTTCGCTATTACGCC3'

pUC 19 forward PCR primer.

5' CCTGCGTTATCCCCTGATTCTG3'