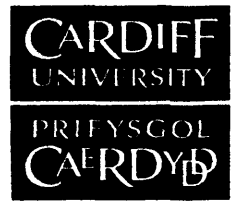
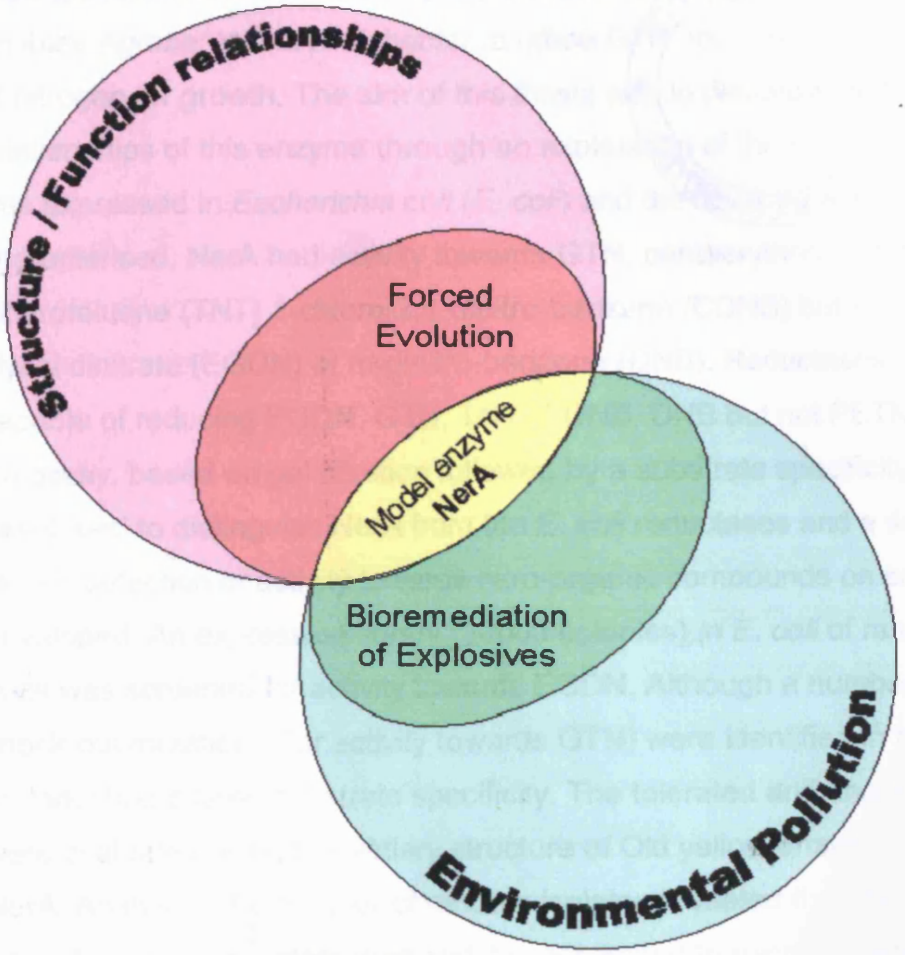


BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax +44 (0)29 20371921
e-mail bindery@cardiff.ac.uk

Navigating the protein landscape around nitroglycerin reductase



The work in this thesis stands at the overlap between the themes "Bioremediation of explosives" and "Forced evolution".

By Michael N.M.Bertels, for the degree of doctor of philosophy
at the University of Wales, July 2004.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Nitro-glycerine reductase (NerA) liberates nitrite from glycerol trinitrate (GTN) and enables *Agrobacterium radiobacter* to utilise GTN and related explosives as sources of nitrogen for growth. The aim of this thesis was to describe the structure / function relationships of this enzyme through an exploration of the protein landscape. *nerA* was expressed in *Escherichia coli* (*E. coli*) and the resulting activity was characterised. NerA had activity towards GTN, pentaerythritol tetranitrate (PETN), trinitrotoluene (TNT) 1-chloro-2,4-dinitro-benzene (CDNB) but not against ethylene glycol dinitrate (EGDN) or m-dinitro-benzene (DNB). Reductases native to *E. coli* capable of reducing EGDN, GTN, TNT, CDNB, DNB but not PETN were also found. An assay, based on gel filtration followed by a substrate specificity assay, was developed to distinguish NerA from the *E. coli* reductases and a screening method for the detection of activity towards nitro-organic compounds on solid medium was developed. An expression library (30000 colonies) in *E. coli* of random mutants of *nerA* was screened for activity towards EGDN. Although a number of tolerated and knock-out mutations (for activity towards GTN) were identified in *nerA*, none of the mutants had altered substrate specificity. The tolerated and knock-out mutations were evaluated using the tertiary structure of Old yellow enzyme, a homologue of NerA. Analyses of a number of random isolates revealed that the ability to liberate nitrite from GTN is widespread and it is not limited to micro-organisms previously exposed to explosives. The statistical parameters involved in the evolutionary cycle that was used to navigate the protein landscape were analysed. It was concluded that mutagenic PCR can generate only 40% of all amino acid exchanges that are possible and that the majority of these exchanges can be achieved only by one particular mutagenic event.

Acknowledgement

I would like to thank Dr. Andrew Lovering (University of Birmingham, UK) for providing me with a sample of purified NfsB and Dr. Andrew Karplus (Oregon State University, USA) for sending me a picture of the FMN environment in Oye. I would also like to thank Dr. Sam Marshall for sharing material and knowledge. A special thanks goes to my supervisors Dr. Graham White and Dr. Andy Morby for their scientific and moral support during my Ph.D. project and their patience. Finally I would like to thank all my colleagues in Lab 315 for the many "Breakkies" we enjoyed together.

Declaration

This work has not previously been accepted in substance for any degree and is not currently submitted in candidature for any degree. This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by citations of explicit references. A bibliography is appended. I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .....

Michael Bertels

Date .....

Table of contents

Abbreviations

ATP	adenosine triphosphate
BSA	bovine serum albumin
CDNB	1-chloro-2,4-dinitro-benzene
CHN	2-cyclohexenone
CTP	cytidine 5' triphosphate
dATP	deoxyadenosine 5' triphosphate
dCTP	deoxycytidine 5' triphosphate
dGTP	deoxyguanosine 5' triphosphate
dITP	deoxyinosine 5' triphosphate
DNA	deoxyribonucleic acid
DNB	m-dinitro-benzene
dNTP	mixture of dATP, dCTP, dGTP and dTTP
dTTP	deoxythymidine 5' triphosphate
EDTA	ethylenediaminetetraacetic acid
EGDN	ethylene glycol dinitrate
FMN	flavin mononucleotide
FPLC	fast protein liquid chromatography
GDN	glycerol dinitrate
GTN	glycerol trinitrate
GTP	guanosine 5' triphosphate
HMX	octahydro-1,3,5,7-tetranitro-1,3,5,7-tetraazocine
HPLC	high pressure liquid chromatography
IPTG	isopropyl b-D-thiogalactopyranoside
NAD(P)H	NADH or NADPH
NAD+	oxidised form of nicotinamide adenine dinucleotide
NADH	reduced form of nicotinamide adenine dinucleotide
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
NC	Nitro cellulose
NerA	nitro-glycerine reductase
NfsB	dihydropteridine reductase or TNT reductase
Onr	Organic nitroester reductase
Oye	Old yellow enzyme
PCR	polymerase chain reaction
PETN	pentaerythritol tetranitrate
PHB	p-hydroxybenzaldehyde
RDX	hexahydro-1,3,5-trinitro-1,3,5-triazine
RNA	ribonucleic acid
SDS PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Tetrazolium violet	2,5-diphenyl-3-a-naphthyltetrazolium chloride
TNT	trinitrotoluene
Topo	Referring to a vector that is associated with the enzyme vaccinia topoisomerase I see Section 2.1.2
TTP	thymidine 5' triphosphate

1 INTRODUCTION

Nitro-organic compounds that find application as explosives are xenobiotic, generally recalcitrant to biodegradation and toxic to humans. Manufacturing, handling and disposing of explosives generates vast waste streams that need to be processed in an ecologically acceptable manner.

A nitro-ester reductase is an enzyme that can reduce and thereby (partially) degrade a certain class of nitro-organic compounds

(For example: Glycerol trinitrate + NADH $\rightarrow$ Glycerol di nitrate + NAD^+ + NO_2^-).

Four similar nitro-ester reductases had been reported at the time this project started. These enzymes have activity towards several commonly used explosives, but cannot totally mineralise any explosive.

This thesis describes how the structure of a model nitro-ester reductase (NerA) was altered in an attempt to obtain an enzyme with new substrate specificity or better catalytic properties. This process was intended to give insight into the reaction mechanism by which NerA reduces nitro-organics and to generate new enzymes that can be applied in the biodegradation of explosives.

1.1 Explosives and Biodegradation

1.1.1 Nitro-organic explosives

The most important characteristic about any energetic mixture or compound is that it can release energy in an auto catalytic fashion because it contains both a reducing and an oxidising agent. Mono-molecular explosives contain both agents in one molecule. The most important class of mono-molecular explosives are organic compounds that contain nitro (-NO_2) groups. These nitro-organics can be subdivided into 3 groups according to the atom to which the nitro group is attached (Yinon, 1993).

1. Nitro compounds The nitro group is attached to a carbon atom.

2,4,6- Trinitrotoluene (TNT, Figure 1.1) is a nitro-compound that is the most widely used military explosive.

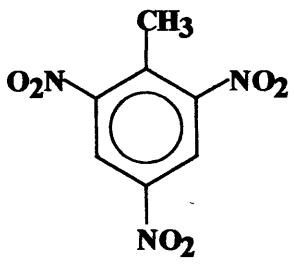


Figure 1.1. 2,4,6-Trinitrotoluene (TNT)

2. Nitrate esters The nitro group is attached to an oxygen. Glycerol trinitrate (GTN, Figure 1.2) is a nitrate ester that is widely used in industrial explosives. Other nitrate esters are: nitro-cellulose (NC), pentaerythritol tetranitrate (PETN) and ethyleneglycol dinitrate (EGDN).

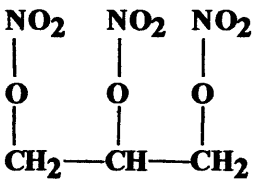


Figure 1.2. Glycerol trinitrate (GTN)

3. Nitramines Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX, Figure 1.3) is a nitramine that is one of the most important military explosives used today. Another nitramine is octahydro-1,3,5,7-tetranitro-1,3,5,7-tetraazocine (HMX).

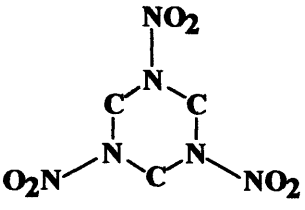


Figure 1.3. Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX)

Toxicity

It has been reported that nitro-organic explosives are toxic to humans although improved industrial hygiene after the 1st world war has reduced the number of poisoning cases among personnel (Yinon, 1990). Various model organisms have been used to determine the toxicity or mutagenicity of nitro organic compounds (Table 1.1). An overview of physiochemical data available for a number of explosive substances can be found in a report by the UK Environment Agency (Fishwick, 2000).

Table 1.1. Models for studying toxicity of nitro organic compounds

Model (Organism)	Reference
Rats	(Reddy, 2000)
Salamander	(Johnson, 2000)
Salmonella histidine reversion assay	(George, 2001)
Earthworm	(Robidoux, 2000,2001)
Seed germination assay	(Siciliano, 2000)

Syntheses, handling and disposal of explosives generate vast waste streams of nitro organic compounds. This poses a threat, as the fate of explosives in the environment is unknown. There are a few complex nitro-compounds of microbial origin that occur in nature (For example chloramphenicol, Turner, 1980). However the nitro-organics that find application as propellants or explosives are considered to be xenobiotic (Ramos, 1996; Shelley, 1996) as they have never been reported to occur naturally in the environment. **More must be learnt about the fate of these xenobiotic and toxic compounds in the environment**. The interaction of a xenobiotic substrate with biological systems is also interesting, from an academic point of view.

1.1.2 Biotransformation of explosives

Bioremediation of xenobiotic substances has some potential advantages over conventional methods like incineration and chemical treatment. Biological methods are more versatile and can be more economical (Chen, 1999).

There are many reports about the biotransformation of various nitro-organics that find application as explosives. Many studies have focused on TNT, probably because it is the most widely used explosive. At the time this Ph.D. project was started (1998) there were only two reports on purified enzymes (line 1 and 2 in Table 1.2) but there were numerous reports on organisms capable of biotransformation of TNT. Many of the groups that reported these organisms before 1998, subsequently detail the enzyme responsible for conversion in these organisms in 2000/01 (labelled with *** in Table 1.3). This indicates that a lot of progress has been made in the field over the recent years and that the biotransformation of explosives remains an important research focus.

The majority of the studies on biotransformation of explosives focus on the way the biofermentor is operated, cosubstrate requirements etc. Some studies actually focus on the composition of the microbial culture involved.

An overview of the studies on enzymes and axenic cultures that can convert TNT is displayed in Table 1.2 and Table 1.3 respectively. In most cases TNT is converted into amino/nitrotoluenes or the protonated dihydride-Meisenheimer complex (Kim, 2000; Vorbeck, 1998). These products are often not further metabolised (Hawari, 2000) and the breakdown products can inhibit the enzyme responsible for conversion (Riefler, 2002). Another problem with biotransformation of TNT is that TNT and especially the breakdown products are effectively immobilised in soil (Achtnich, 2001; Wikstrom, 2000).

Table 1.2. Biotransformation of TNT by enzymes

Overview of literature on microbial biotransformation of TNT. Most groups report formation of mainly 4-amino-2,6-dinitro toluene and its isomers and 2,4-diamino-6-nitrotoluene (A-NT) or the dihydride- Meisenheimer complex. These compounds were not further degraded under the conditions used. *** indicates that it was previous reported that the host organism was capable of converting TNT.

A. Enzymes	Products formed	Reference
1. PETN reductase *Enterobacter cloacae* PB2	Meisenheimer	(French, 1998)
2. Nitrobenzene reductase *Pseudomonas pseudoalcaligenes*	Dihydroxy Nitrotoluene	(Fiorella, 1997)
3. Cytochrome P-450 *Bjerkandera adusta*	Not specified	(Eilers, 1999)
4. Type I nitroreductases soil enterobacteria	Not specified	(Kitts, 2000)
5. Nitro reductase *Pseudomonas aeruginosa*	A-NT	(Oh, 2001) ***
6. Xenobiotic reductase B *Pseudomonas fluorescens* I-C	Meisenheimer / A-NT	(Pak, 2000)
7. Manganese peroxidase white-rot basidiomycete *Phlebia radiata*	None (Total breakdown)	(Van Aken, 1999) ***
8. Glutathione / peroxidase H5 white-rot fungus *Phanerochaete chrysosporium*	None (Total breakdown)	(Van Aken, 2000) ***
9. Respiratory chain *Pseudomonas* sp, strain JLR11	A-NT	(Esteve-Nunez, 2000) ***

Table 1.3 Biotransformation of TNT by micro-organisms

Organisms that were already mentioned in part A (indicated ***) were excluded from this list. There many reports on *Pseudomonas* species that can degrade TNT, in all cases amino nitro toluenes or related compounds were formed:

Bacteria	
Pseudomonas sp. *Vesicularis*	(Davis, 1997)
Pseudomonas sp. Strain KA	(Tope, 2001)
Pseudomonas sp. *Savastanoi*	(Martin, 1997)
Mycobacterium-Vaccael	(Vanderberg, 1995)
Desulfovibrio, desulfuricans strain A	(Costa, 1996)
Clostridium acetobutylicum	(Hughes, 1998)
Serratia marcescens	(Montpas, 1997)
Archea	
Methanococcus Sp (Strain-B)	(Boopathy, 1994)
Fungi	
Phanerochaete sordida, Phlebia brevispora, Cyathus stercoreus	(Donnelly, 1997)
Suillus variegatus	(Meharg, 1997)
Ceratocystis coerulescens, Lentinus lepideus, Trichoderma harzianum	(Samson, 1998)
In a study with seven strains of whiterot fungi Meisenheimer complexes were formed or TNT was mineralised	(Kim, 2000).
C. Plants and Animals	
Datura innoxia (Plant)	(Lucero, 1999)
Parrotfeather (Plant)	(Medina, 2000)
Bromegrass	(Sun, 2000)
Eisenia andrei (Earthworm)	(Renoux, 2000)
Tobacco plant heterologously expressing PETN reductase	(Hannink, 2001)

The majority of the studies on biotransformation of GTN describe total mineralisation by poorly defined mixed microbial cultures. Mineralisation of EGDN by a defined mixed culture has also been reported (a consortium of *Arthrobacter ilicis* and *Agrobacterium radiobacter* (Ramos, 1996)). There are two reports on axenic cultures that can mineralise GTN (Marshall, 2001; Zhang, 1997b), but most studies with axenic cultures report only partial degradation of GTN (Table 1.4B).
Some purified enzymes can reduce nitro groups from GTN whereby nitrite is released (Table 1.4A). Reduced FMN (Flavin Mono Nucleotide, a co-factor present in the majority of the enzymes listed in Table 1.2 and Table 1.4) can remove all three nitro groups from GTN (Meah, 2001).

Table 1.4. Biotransformation of GTN

Overview of literature on biotransformation of GTN. Most groups report that not all nitro groups are removed from GTN.

A. Enzymes

	Reference	
Purified PETN reductase *Enterobacter cloacae* PB2	(French, 1996)	See also Table 1.5
Purified GTN reductase *Agrobacterium radiobacter*	(Snape, 1997)	
Purified nitroester reductases *Pseudomonas putida* *Ps. fluorescens*	(Blehert, 1999)	
Old Yellow Enzyme	(Meah, 2001)	
Xanthine oxidoreductase	(Millar, 1998)	

B. Microbial organisms

Organisms that are also mentioned in Table 1.2 are printed in bold.

Pseudomonas sp. strain RI-NG1	(White, 1996)
Bacillus thuringiensis, B. cereus *Enterobacter agglomerans*	(Meng, 1995)
Rhodococcus (total mineralisation of GTN)	(Marshall, 2001)
Fungi	
Phanerochaete chrysosporium	(Servent, 1991)
Penicillium corylophilum Dierckx (total mineralisation of GTN)	(Zhang, 1997b)

C. Plants.

Beta vulgaris (Sugar beet)	(Goel, 1997)
Tobacco expressing PETN reductase	(French, 1999)

1.1.3 Mechanisms of Biotransformation

Pathways in micro organisms

Large scale bioremediation processes for example, in a wastewater treatment plant, use undefined mixed cultures (Schlegel, 1990). From a commercial perspective there seem to be no reasons to attempt bioremediation by an axenic culture. However from an academic perspective it can be interesting to study biotransformation by axenic cultures. In a mixed culture, microbial communities can be variable in composition (even during a single experiment) and there can be complex interdependencies. Pure cultures are more suitable to study the actual mechanisms of biotransformation. There are two organisms that appear several times in articles on biotransformation by axenic cultures.

1. The filamentous white rot fungus *Phanerochaete chrysosporium*. The fungus converted GTN into its di- and mononitrate derivatives with the formation of nitric oxide and nitrite (Servent, 1991). The fungus partially degraded TNT (Donnelly, 1997; Fernando, 1991) and it was reported to further degrade amino/nitrotoluenes (Jackson, 1999). There is also evidence that the organism degrades RDX (Fernando, 1991; Bayman, 1995).
2. Aerobic Gram-negative bacteria of the genus *Pseudomonas*. These bacteria converted GTN into its di- and mononitrate derivatives with the formation of nitrite (White, 1996). Nitroester reductases that catalyse the reaction have been isolated from species of *Pseudomonas* (Blehert, 1999; Fiorella, 1997, Table 1.5). There are reports on the partial degradation of TNT (Davis, 1997) and there are claims that the bacteria can further degrade the amino/nitrotoluene-intermediates (Alvarez, 1995; EsteveNunez, 1998; Fiorella, 1997) or that dinitrotoluene is formed instead (Martin, 1997). Considering the number of *Pseudomonas* species that have been reported to be able to convert TNT, it would be of interest to see if there are any *Pseudomonas* species that do <u>not</u> have activity towards TNT.

GTN Reductases

There is one report of cell free extract that can remove the nitro groups from GTN without the addition of cofactors (Meng, 1995) but all the enzymes in Table 1.2 and Table 1.4 convert the nitro-organic substrates via a reductive pathway requiring NAD(P)H. The majority of these enzymes are reported to be flavo-enzymes and four of the flavo enzymes have been well characterised (Table 1.5). They are homologues of the intensively studied Old Yellow Enzyme (Oye) from brewer's bottom yeast (*Saccharomyces cerevisiae*) (Williams, 2002). Indeed, Oye itself has activity towards GTN (Meah, 2001).

Table 1.5. Properties of flavo enzymes capable of reducing GTN

References for these enzymes have already been given in Table 1.4. "Specific activity" lists the specific activity towards GTN (U/mg = μmol of substrate consumed per minute per mg of enzyme). "Major product" lists the major products formed when GTN was reduced. 2-cyclohexenone (CHN) is a model substrate often used in the study of Old Yellow Enzyme homologues.

Organism	*Enterobacter cloacae* PB2	*Agrobacterium radiobacter*	*Pseudomonas putida*	*Pseudomonas fluorescens*
Enzyme name	Onr / PETN reductase	NerA / GTN reductase	XenA	XenB
Specific activity (U/mg)	25	15	124	110
Major product	Unknown	1,3 GDN	1,2 GDN	1,3 GDN
Substrates	PETN, EGDN, TNT CHN	PETN, TNT	CHN, GDN	TNT, GDN
Not Substrates		EGDN	TNT	CHN
Cofactor	NADPH	NADH	NAD(P)H	NAD(P)H

Onr and NerA follow a Ping-Pong reaction mechanism (indicated by competitive inhibition by the cofactor NAD(P)H) when reducing GTN. The FMN-dependent nitrate ester reductases listed here all have a monomeric mass of around 40kDa. They release nitrite from GTN and produce one isomer of glycerol dinitrate (GDN) in excess of the other (See "Preferred products" in column 2 in Table 1.5). They have very poor or no activity towards the GDN that is formed.

The enzymes have been isolated from bacteria that can utilise GTN or PETN as a nitrogen source for growth. These bacteria were obtained via an enrichment strategy, this implies that GTN can cross the cell-membrane. The native functions of the enzymes are unknown.

Substrate specificity

Several widely applied explosives seem susceptible to at least partial degradation. However most degradation studies on TNT and RDX conclude that recalcitrant intermediates are produced. It would be of interest to find or engineer biological systems that can mineralise the intermediates, or systems that can avoid pathways that produce these intermediates. It remains unclear by which exact mechanism the nitroester reductases mentioned in this section reduce GTN. One fascinating aspect of the four reductases is that they all seem to belong to the Oye family. Another fascinating aspect is the difference in substrate specificity:

- Onr and NerA have a comparable activity towards GTN, but a very different activity towards EGDN.
- XenA and XenB can release two moles of nitrite from GTN whereas NerA can release only one. None of the enzymes can reduce all three nitro-ester bonds to produce glycerol.
- Most enzymes have activity towards a nitro-ester (C-O-NO_2 like GTN) *and* a nitro-compound (C-NO_2 in TNT), with exception of XenA.
- XenA and XenB seem to have 5-8 times higher activity towards GTN than Onr or NerA.
- NerA and Onr have a strong preference for NADH and NADPH respectively as cofactor, whereas XenA and XenB accept both cofactors.

At the time this Ph.D. project was started the sequences of XenA and XenB were not yet available. The closest family member of NerA, both structurally and functionally was Onr. The most distinct difference between Onr and NerA was the contrast in activity towards EGDN. Because of the structural similarity (sequence homology) it was assumed that activity towards EGDN was a likely attribute for a mutant of NerA to obtain. Native NerA had no activity towards EGDN so that a mutant would easily be detected. By relating the mutation to the new phenotype, much could be learnt about the interaction of an Oye-like enzyme with a nitro-ester.

The process of finding the mutation(s) that would allow NerA to gain activity towards EGDN involves exploring the protein landscape around NerA for activity towards EGDN. This concept is introduced in Section 1.2.

NerA in *E. coli*

The *nerA* gene was heterologously expressed in *E. coli*. (Snape, 1997). It gave whole cells of *E. coli* the ability to release nitrite from GTN. Initial experiments suggested that NerA gave *E. coli* the ability to utilize GTN as a source of nitrogen for growth under anaerobic conditions (S.J.Marshall, pers.com.).
It was therefore assumed that enrichment could be used to select for *E. coli* host expressing enhanced versions of NerA.

1.2 Navigating the protein landscape

Structure and Function

There are two important catalytic properties of enzymes: 1.They greatly reduce the activation energy of a conversion. 2.They have high substrate specificity. These two properties have been explained for quite a few enzymes with the help of structure / function analysis. The major goal of the study of structure / function relationships is to be able to customise the catalytic properties of enzymes to perform chemical conversions of interest, in this case the degradation of explosives. The mechanism of the catalytic site stands at the heart of the relationship between structure and function in enzymes.

All amino acids and other components of the structure can contribute to the configuration of the active site, either directly by providing a catalytic residue or indirectly by affecting the folding of the entire enzyme. The configuration of the active site ultimately defines the chemical conversion that can be performed (Schuster, 1997).

Structure / function relationships can be studied in a number of ways.

- A structural molecular model of an enzyme can be generated based on primary sequence similarity to another enzyme, with known structure. This model can then be related to the proposed function of the enzyme and thus a reaction mechanism can be proposed.
- The sequence of proteins can be altered. The change in structure is then correlated to the change in function. Enzymes with enhanced properties can be further altered and again passed through a screen. This iterative process mimics key elements of Darwinian evolution and can be referred to as forced evolution (Arnold, 1996) or the exploration of the protein landscape (Section 1.2.1).
- The structure of proteins with similar function (functional family) or the function of proteins with similar structure (structural family) can be compared. These two comparisons are based on obtaining insight using the "family information" available in nature.
- Entirely new structures (artificial enzymes) can be generated or designed (for example antibodies raised against molecules that resemble transition states (Kamei, 2001)).

1.2.1 Protein landscape

Forced evolution can be visualised as the exploration of a protein (structure / activity) landscape (Kuchner, 1997; Trakulnaleamsai, 1995).
The original structure of a protein stands at the origin of a three dimensional space. Every structural change is a step in the XY plane (the sequence space). The performance of the enzyme can be plotted along the Z-axis, thus creating a landscape. Peaks in the landscape represent structures with high activity. Just like in geographical landscapes, peaks can be single islands or they could be connected by continents, reefs or mountain ranges.

The evolutionary cycle

The evolutionary cycle consists of three essential steps:

1. Generating change
2. Selection for improved function
3. Characterization of improved structures

There are 3 ways to generate change (take a step in the XY plane of the protein landscape).

1. Protein engineering
2. Random mutagenic methods
3. Sexual exchange

The first method represents a rational walk through the protein landscape. Only a hand-full of new structures are generated. The latter two have counterparts in nature. They generate a population of new structures that covers a certain plot in the XY plane. The random approach explores a small patch around the native sequence. The sexual approach explores the reef between the family members. The various ways of generating change are further introduced below.

Rational mutagenesis / Protein engineering

When the structural model of an enzyme is available and the function of various groups within the enzyme has been proposed, the enzyme can be altered deliberately and precisely in order to obtain an enzyme with improved catalytic properties (Harris, 1998). The structure of an enzyme can be altered by site directed mutagenesis of the encoding gene. Indeed protein engineering has been successful for various enzyme families including lipases. Lipases have been altered for use in commercial detergents (Svendsen, 1997).

However enzymological reaction mechanisms are still poorly understood and the "rational design" is actually an educated guess in most cases. There is not enough known about the reaction mechanism involved in conversion of explosives or about the structure of nitro ester reductases to make suggestions for engineered improvements.

1.2.2 Random mutagenic methods

Random mutagenic methods represent single steps to explore the direct environment (plot) around an elevated area in the protein landscape. A known structure is altered randomly to generate an offspring of structures that are only marginally different from the native structure. Preferably all unique members of the offspring are evaluated in order to detect nearby peaks. The fact that random mutagenesis allows the navigation of only a very small plot in the protein landscape is a major limitation. Progress can only be made in a linear fashion (i.e. beneficial mutations accumulate one after the other) and if the original structure is close to a local maximum in the landscape, the small surface that is covered by random mutagenesis will not extend beyond that island of activity. In fact the gorge between two islands can be so deep[1] that even natural evolution does not cross it in 2 billion years (Hall, 1998). However the information obtained from random walks might be beyond any speculation of rational design. There are numerous methods to introduce random changes in a gene but PCR offers the greatest potential in controlling the parameters of mutagenesis.

There is a hypermutagenic PCR method that involves all four transitions to an equal extent (Vartanian, 1996). In fact the enzyme that is most commonly used in PCR reactions, Taq-polymerase has a fairly high intrinsic mutation frequency (Zhao, 1997). The mutagenic PCR protocol used in this study is presented in Section 2.3.6

[1] The meaning of a deep gorge is that two islands are only a few mutations away from each other, but the structure that lies halfway is degenerate and the evolutionary path cannot pass through that degenerate structure.

1.2.3 Family information

Structural versus Functional family

The exploitation of family members to obtain information about structure / function relationships depends primarily on structural and functional family members that are found in nature and that have been generated by natural evolution. There is an important difference between the functional and structural family. Enzymes that are structurally similar to NerA are close to NerA in the XY plane of the protein landscape. Although their structure is similar to NerA, they might not have activity towards GTN. In that case they represent a peak in another Z (activity) dimension. Enzymes that are functionally similar to NerA are close to NerA in the Z (activity) direction. Although their function is similar to NerA, they might not have a structure that resembles NerA. In that case they represent a peak that lies very far away in the XY direction of the protein landscape.
Overlaps and differences in structure and function in family members might give insights into structure / function relationships.

Gene shuffling

When directing the evolution of a protein much can be learnt from evolution in nature. Sex is a very important factor in natural evolution and it should be incorporated in forced evolution. When genes are shuffled, the parental structures exchange information to generate an offspring that contains new combinations of parental information. A method called "sexual PCR" (gene shuffling) allows in-vitro recombination (Stemmer, 1994a; Stemmer, 1994b). By shuffling single mutants, the synergetic effect of two mutations can be examined. The DNA sequences that are to be shuffled are cut into small segments. The segments are the template for a PCR-like reaction. The segments are melted and reassembled based on their homology. Enzymes that belong to the same family are similar and therefore relatively close in the XY plane. Indeed it has been reported that the shuffling of a family of genes accelerates directed evolution (Crameri, 1998).

It must be noted that the wild-type DNA that encodes the family members might not be homologous enough to allow shuffling, due to differences in codon usage. In that case shuffling must be performed with artificial DNA with a codon table of only 21 codons (McPherson, 1999).

1.2.4 Applications of forced evolution

A classical example of forced evolution is improvement of agricultural crops by selection of preferred varieties. It is a form of forced evolution since the selection process is imposed by man. In the classical examples only the second and third steps of the evolutionary cycle are controlled by man. The rate limiting step in these cases is the generating of change. There are some instances where natural processes can involve quite dramatic changes. An example is the exchange of non-homologous genes among bacteria if strong selective pressure is applied (Chakrabarty, 1992).

The process of forced evolution can be accelerated by increasing the rate of change in the encoding gene. Bacterial strain improvement programs, where industrial strains were exposed to mutagenic substances and then screened for certain phenotypes are an early example of this. In this example both the selection process and the rate of change are artificial. There are numerous reports on the application of forced evolution. An overview of recent literature, organized by theme is given in Table 1.6. The majority of the studies are concerned with the alteration of substrate specificity.

Table 1.6. Applications of forced evolution

A number of applications of forced evolution is listed here. Applications include altered substrate specificity, increased stability or performance or to make enzymes suitable for applications in new fields.

A. Altered substrate specificity	
Hydrolysis of a sterically hindered 3-hydroxy ester in enantiomeric excess	(Bornscheuer, 1998)
Moxalactamase activity from four cephalosporinase genes	(Crameri, 1998)
Lactate dehydrogenase	(Dafforn, 1995)
Mu class glutathione transferases	(Hansson, 1999)
New recognition site for a restriction endonuclease	(Lanio, 1998)
Aspartate aminotransferase	(Oue, 1999; Yano, 1998)
Making a galactosidase into a fucosidase	(Zhang, 1997a)
Increasing activity of a Beta-lactamase towards cefotaxime.	(Stemmer, 1994b; Zaccolo, 1999)
Trypsin inhibitor	(Yokobayashi, 1996)
B. Increased (thermo) stability	
Make a heme peroxidase suitable for use in laundry detergent	(Cherry, 1999)
Stabilise an esterase without compromising its activity	(Giver, 1998)
Thermostability of tumor suppressor p53	(Matsumura, 1999)
Thermostability of catalase I from *Bacillus stearothermophilus* by adding random C terminal.	(Matsuura, 1999)
Stable and functional cysteine-free antibody single-chain fragments (scFv)	(Proba, 1998)
Make *E. coli* resistant to both X rays and ultraviolet photons	(Ewing, 1995)
C. Improved performance	
Green fluorescent protein	(Crameri, 1996)
Arsenate detoxification pathway in *E. coli*	(Crameri, 1997)
Antifreeze protein HPLC-8	(Meijer, 1996)
D. Commercial / Medical applications	
Industrial catalysts in general	(Arnold, 1996; Arnold, 1997)
Optimisation of enzymes for use in organic sythesis	(Bonneau, 1996)
Deprotection of an antibiotic p-nitrobenzyl ester in aqueous-organic solvents	(Moore, 1996)
Subtilisin E from *Bacillus subtilis* with enhanced total activity in aqueous dimethylformamide	(You, 1996)
Ability to phosphorylate zidovudine, application in anti-HIV gene therapy	(Christians, 1999)
Glutathione S-transferases for detoxication	(Gulick, 1995)

Table 1.6 continued E. Bioremediation	
Polychlorinated biphenyls	(Bruhlmann, 1999; Kumamaru, 1998)
Dechlorination of atrazine	(Wackett, 1998)
Organophosphate hydrolase	(Mee-Hie Cho, 2002)

1.2.5 Exploring the protein landscape around NerA

Forced evolution has successfully been applied in the field of bioremediation of chlorinated compounds (Table 1.6E) and it has been successfully used on an Oye-like enzyme (Altamirano, 2000). Therefore forced evolution will be applied on NerA.

Starting point for exploration

At the time this Ph.D. project was started, Onr and NerA (Table 1.5) were the only two enzymes that had been characterized that could reduce GTN. Because of the xenobiotic character of the nitro-ester bond in GTN it was assumed that very few enzymes would have activity towards GTN. This assumption was strengthened by the fact that *onr* and *nerA* had significant sequence similarity. This implied that Onr and NerA occupied a very small XY plot in the protein landscape and that activity towards GTN was a unique feature. This was one of the reasons why NerA was chosen as the starting point for the exploration of the protein landscape. The rationale behind choosing NerA as the starting point for exploration was:

- The assumption that NerA was a rare enzyme and that it would prove difficult to find other enzymes with similar function.
- The fact that NerA was available from a heterologous source and that it had already been characterized in the Cardiff laboratory (Snape, 1997).
- The fact that a screening method was available to select for *E. coli* expressing enhanced versions of NerA (see "NerA in *E. coli*" in Section 1.1.3).

Towards the end of this Ph.D. project it became clear that these assumptions were not all valid, see Chapter 7.

Pathways that lead away from NerA

The starting point for the exploration of protein space was in this case NerA. It was the reference point against which to measure improvement or deterioration in activity.

Several pathways that lead away from this starting point must be explored. Pathways that go uphill lead to structures with improved activity (according to the protein landscape model, Section 1.2.1). Three ways were proposed to walk away from the starting point that represents NerA.

1. Based on single mutagenesis of the backbone (Ph.D. project proposal, White pers.com.). This is represented in the protein landscape as exploration of the plot directly surrounding NerA.
2. Based on enzymes structurally related to NerA, that would be suitable for shuffling. This is represented in the protein landscape as exploration of the reef that connects two known peaks.
3. Based on enzymes functionally related to NerA. This is represented in the protein landscape as hopping from peak to peak.

Protein engineering was not pursued because too little was known about the catalytic mechanism of reduction of GTN by NerA.

Required tools

The tools that were required for the exploration of the protein landscape around NerA seemed to be available at the start of this Ph.D. project.

- A tool to generate change. A protocol to introduce random mutations into *nerA* would be based on a protocol described by (Cadwell, 1992) and (Vartanian, 1996).
- A tool to generate an expression library. *nerA* would be cloned and expressed using conventional cloning techniques.
- A screening method would be based on an enrichment strategy ("NerA in *E. coli*" in Section 1.1.3, assuming *E. coli* could utilise the nitrite that was liberated from GTN by NerA.

During the progress of the Ph.D. project it became clear that for example the enrichment tool did not perform as well as expected, see Section 5.1.

Aim

The **aim of this study** was to explore the protein landscape around NerA to understand more about structure / function relationships in biocatalysts in general and specifically to learn more about the interaction of nitro-organic substrates with NerA. Exploration was expected to have the potential to generate new enzymes that could be applied in the bioremediation of explosives. Forced evolution was used to study possibilities for bioremediation of explosives. This study stands at the overlap of these two themes (see figure on title page).

2 METHODS

2.1 Materials

2.1.1 Chemicals, Reagents and Laboratory consumables

General laboratory chemicals were obtained from Sigma-Aldrich (Gillingham, UK) or Fisher Scientific (Loughborough, UK) and were of analytical grade or higher. Plastic ware and general lab consumables were obtained from Greiner (Stonehouse, UK) and Alpha Laboratories (Eastleigh, UK). Other reagents and consumable suppliers are listed in the text as required. The compositions of all media and buffers are listed in Appendix B. LB- and SOB medium, in capsule-form, and Noble agar were supplied by Anachem (Luton, UK). LB agar was obtained from Gibco BRL (Now Invitrogen, Paisley, UK) in a premixed form. All growth media were made up according to the manufacturer's instructions. SOC medium was obtained by supplementing SOB medium with glucose (25mM).

2.1.2 Vector, Host and Antibiotic selection

The cloning and expression vector used in this research was the pTrcHis2-Topo vector. The pTrcHis2 TOPO TA Cloning® Kit is designed for rapid PCR cloning followed by high-level expression in *E. coli*. The vector, pTrcHis2-TOPO, is activated by topoisomerase for direct cloning of a PCR product. Taq polymerase has a terminal transferase that adds a single deoxyadenosine to the 3' end of PCR products. The terminal A targets the PCR product to the T overhang of the vector. Topoisomerase will catalyse the ligation of the insert into the vector (Invitrogen, 1999). Topo *lacZ* was supplied by the manufacturer and Topo∅ was obtained as described in Figure 2.1A. The *nerA* gene was subcloned in the Topo vector using the Topo cloning protocol (Figure 2.1B). The recommended strain, *E. coli* Top10[1] was used to host the plasmid.

---

[1] *E. coli* Top10, derivative of Strain K12, genotype F - *mcr*A _(*mrr-hsd*RMS-*mcr*BC) _80*lacZ*_M15 _*lac*X74 *deo*R *rec*A1 *ara*D139 _(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str ^{R}) *end*A1 *nup*G.

Competent Top10 cells were purchased (Invitrogen, 1999) or prepared as described in Section 2.3.1. LB-agar plates containing ampicillin (50 mg/l) were used to select for host with plasmid. Liquid media were also supplemented with ampicillin (50 mg/l).

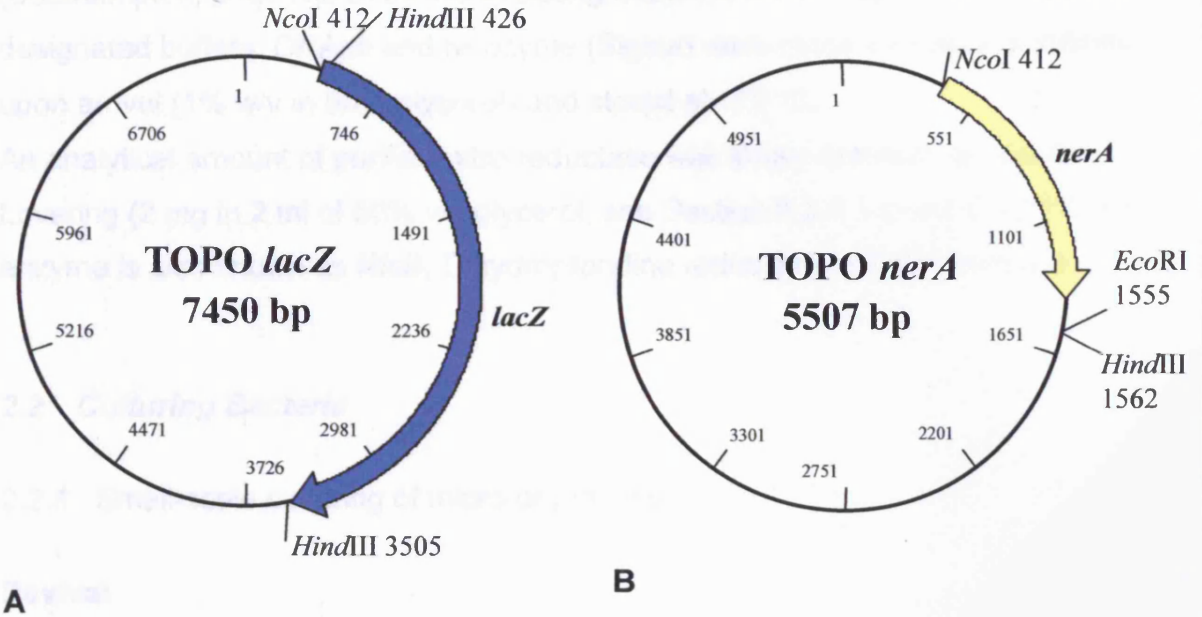


Figure 2.1 Topo plasmids

(A) *lacZ* was removed from Topo *lacZ* using the Hind III sites up- and down-stream of the *lacZ* insert. The re-ligated vector was called Topo ∅ and was used as a receiving vector in a classical cloning approach. (B) Topo *nerA* was constructed as described in Section 4.2. It was used as a template for mutagenic amplification of *nerA*. The resulting PCR product carrying the *nerA* insert could be cloned back into Topo ∅ using the *Nco I* and *Hind III* restriction sites.

Topo *nerA* was used as a template for generating mutants of *nerA* (See "Mutagenic PCR" in Section 2.3.6. Regions flanking the insert were co-amplified in the PCR reaction. These flanking regions contained the *Nco I* and *Hind III* restriction sites that allowed re-introduction into the Topo ∅ vector.

2.1.3 Protein- and DNA-ladders and Enzymes.

Protein- and DNA-ladder, restriction endonucleases, dNTPs, T4 DNA ligase and T4 polynucleotide kinase were purchased from New England Biolabs (Hitchin, UK). *Taq* DNA polymerase and *Pfu* DNA polymerase were purchased from Promega (Southampton, UK). The DNA-manipulating enzymes were supplied and used with designated buffers. DNAse and lysozyme (Sigma) were made into stock solutions upon arrival (1% w/v in 50% glycerol) and stored at –20 °C.
An analytical amount of purified nitro reductase was kindly donated by Andrew Lovering (2 mg in 2 ml of 50% v/v glycerol, see Section 6.3.1, Lovering, 2001). This enzyme is also known as NfsB, Dihydropteridine reductase or TNT reductase.

2.2 Culturing Bacteria

2.2.1 Small-scale culturing of micro-organisms

Revival

Colonies of interest were picked from LB plates or small aliquots (10 µl) of glycerol stock were revived by transferring them to LB-Amp[2] medium (5 ml, in 25 ml airtight Sterilin tubes) and incubation (37 °C, shaking, 18 h).

Growth

LB medium (100 ml in 250 ml Erlenmeyer) was supplemented with ampicillin (5 % w/v stock solution, 100 µl), inoculated with revived culture (5 % v/v) and incubated (shaking 37 °C, 18 h).

Induction

When heterologous expression was required, the LB medium (100 ml) was also supplemented with isopropyl β-D-thiogalactopyranoside (IPTG, 1 M stock solution, 100 µl, final concentration 1 mM) and incubated until the appropriate growth phase was reached (shaking 37 °C, 3 h, attenuance $D_{600} > 1$).

---

[2] When growing environmental isolates, ampicillin was omitted, unless there were indications that the isolate was not sensitive to ampicillin.

Glycerol stock

Glycerol aliquots (1 ml) were introduced into screw cap Eppendorf tubes (2 ml) and the tubes were autoclaved. Microbes were preserved by transferring microbial culture (1 ml, Section above) to the autoclaved glycerol. The glycerol stock was stored (-20 °C) until required.

2.2.2 Large scale culturing of *E. coli*

Medium (7 litres, 2 x LB, Appendix B) was prepared in a 12-litre fermentation vessel. The amount of phosphate solution (1 M K_2HPO_4) that was required to bring the pH of the medium to neutral (pH = 7) was calculated by titrating a small aliquot (100 ml) of the medium. The required volume of phosphate solution (around 130 ml, final concentration in medium < 20 mM) was measured and autoclaved separately.
The fermentor vessel was autoclaved (120 °C, 30 mins) and allowed to cool overnight. The vessel was made airtight by placing a rubber "O"-ring between the vessel and the lid, and was connected to the fermenter (Microferm, New Brunswick Scientific, St Albans, UK). The medium was stirred (200 rpm), aerated (10 l/min) and kept at constant temperature (36 °C ± 1° C). The medium was supplemented with the phosphate solution, ampicillin (5 %w/v, 7 ml) and IPTG (1 M, 7 ml) and inoculated (250 ml overnight culture in 2 x LB-Amp medium, of appropriate *E. coli* strain). At regular intervals, samples were taken and the attenuance of the culture was monitored at 600nm.
When the appropriate growth stage was reached (D_{600} = 1-1.5 after 3-14 h) the vessel was pressurised by blocking the air-outlet and the culture was transferred to centrifuge buckets (6 x 1 litre) using the liquid-outlet. The cells were harvested by centrifugation (6000 x g), resuspended in ice-cold KP50 (total volume < 120 ml) and kept on ice. The supernatant and remaining culture were sterilised before disposal. Temperature sensitive parts of the fermenter were removed and soaked in disinfectant (Decon 90) and the fermentation vessel (including lid) was autoclaved and cleaned.

2.2.3 Culture density

The density of a microbial culture was estimated by reading the attenuance at a wavelength of 600 nm in a bench top spectrophotometer (Pharmacia, Knowlhill, UK. 1 ml sample in polystyrene cuvette) or plate reader (Thermo max, Molecular Devices, Wokingham, UK, 100 μl sample in polystyrene micro-well plate).

2.2.4 Enrichment

An enrichment medium was prepared by supplementing a nitrogen-free medium (Appendix B) with either EGDN, GTN, TNT, or RDX and used to isolate micro-organisms that can utilise explosives as nitrogen source under aerobic conditions. It has not been tested whether or not any of the nitro-organic compounds can withstand autoclaving and therefore the nitro-organic compound was added after autoclaving. Stocks of nitro-organics were sterile by virtue of being prepared in ethanol or acetone.

Prior to enrichment, the environmental samples were exhausted of residual nitrogen by diluting them (50 g solid or 50 ml liquid) in nitrogen-free medium (0.5 litre) and incubation (7 days, static, ambient temperature).

Enrichment medium (100 ml) was inoculated (1 % v/v) with supernatant from the preparative step and incubated (7 days, static, ambient temperature). Supernatant from this incubation was then used for subsequent dilutions and incubations. This step was repeated until a negative control (nitrogen-free medium not supplemented with any form of nitrogen) no longer showed growth (this indicated that background nitrogen was totally exhausted). At this stage, several serial dilutions of supernatant from the enrichment cultures were plated out on LB-agar plates and plates that contained solidified enrichment medium. Colonies were picked and used to inoculate enrichment medium to confirm the ability of the obtained axenic strains to utilize nitro-organic compound as source of nitrogen for growth. The ability of the strains to liberate nitrite from various nitro-organic compounds was analysed using the whole cell assay (Section 2.4.1). Strains were stored in glycerol stock (Section 2.2.1).

2.2.5 Screening by enrichment

Anaerobic selective medium was (Appendix B) transferred (10 ml) to a narrow glass tube (15mm Diameter) or a volumetric flask was filled up to the fill-mark. A glass bead or magnetic stirring rod was placed at the bottom of the tube or flask and the medium was covered with mineral oil (Sigma, Molecular Biology grade). The tube or flask, including medium and oil was autoclaved (20 min, 120 ^{0}C). The S^{2-} in the medium removed all the oxygen as indicated by the rezasurin indicator (S^{2-} 0.4 g/l, rezasurin 2.5 mg/l, see Appendix B). After inoculation and addition of the appropriate nitrogen source, the tube was placed on a magnetic stirrer or shaking table so that the medium was swirled gently. The oil provided a good barrier to exclude oxygen but gases formed during fermentation could escape. An additional advantage of using oil was that samples could be taken without disturbing the anaerobic conditions.

Great care was taken to minimise the amount of nitrogen that was introduced into the medium by inoculation. The inoculate (an *E.coli* culture) was washed (3x) as described below. The cells were collected by centrifugation and resuspended in Anaerobic selective medium.

2.3 Molecular Biology

DNA manipulations were carried out in accordance with Sambrook (1989) unless specified otherwise.

Molecular Biology Part I The introduction of foreign DNA into *E. coli*

2.3.1 Preparation of electro competent cells

Electro competent cells were prepared according to a protocol adapted from Cooley (1991). During optimisation of the protocol it was concluded that it was important to minimise the time that elapsed between the start and the end of the protocol. Gentle handling of the cells was found to be less important. Lack of conductivity of the cell suspension was found to be a good indicator for the competence of the cells.
SOB (500 ml) medium was transferred to conical flasks (2.5 litre capacity) that allowed good aeration of the medium (wide neck). The medium was inoculated with an over night culture (10ml) and incubated (37 °C shaking at 200 rpm). After 3 h the attenuance of the culture was measured at 30-minute intervals. The culture was placed on ice (30 min) when the appropriate attenuance (D_{600} = 0.5) was reached. The cells were washed five times by centrifugation (6000 x g) followed by resuspension in electroporation buffer (1 x 2 litre, 2 x 50ml, 2 x 4ml) and finally resuspended in electroporation buffer (2 ml). The suspension was divided among several Eppendorf tubes in small aliquots (140 µl), snap-frozen in liquid nitrogen and stored at –80 °C.

2.3.2 Transformation by electroporation

When handling competent cells with automatic pipettes, truncated tips were used. Electro competent cells were thawed on ice. For every transformation, plasmid or ligation mix (1-2 µl) was combined with competent cells (40 µl) in a sterile Eppendorf tube. Care was taken to achieve good mixing. The mixture was transferred immediately to an electroporation cuvette (2 mm, Biorad, HemelHempstead, UK).

The cuvette was placed in an electroporation apparatus (Micropulser, Biorad) and pulsed using the recommended settings (1.8 kV, 6.1 mS at zero conductivity). Cells were immediately diluted in SOC medium (1 ml, ambient temperature), transferred to a universal tube (5 ml) and incubated (37 °C, 30-60 min, shaking). The appropriate volume was plated out on selective LB-agar medium.

2.3.3 Transformation using chemically competent *E. coli*

Chemically competent TOP10 cells were purchased (Invitrogen) and transformed according to the manufacturer's manual. Plasmid or ligation mixture (1 µl) was added to a vial of competent cells that had been thawed on ice (volume 40-100 µl). Cells were incubated on ice (30-60 min), heat shocked (42 °C, 30 s) and diluted in SOC medium (250 µl). After incubation (37 °C, 30-60 min, shaking) the appropriate volume of cells was plated out on selective LB-agar medium.

Molecular Biology Part II Plasmid extraction and DNA purification

2.3.4 Plasmid extraction and rescue

E. coli containing the plasmid of interest was cultured (Section 2.2.1), cells were harvested by centrifugation (6000 x g) and resuspended in P1 buffer (100 µl). (Colonies obtained from by "Plasmid Rescue" (Section 2.4.4) were also subjected to the remainder of this protocol). Cells were lysed by addition of P2 buffer (100 µl). The solution was mixed gently by inversion until it appeared to be homogeneous. Cell debris (proteins, genomic DNA etc.) was precipitated by addition of chloroform (100 µl) and P3 buffer(100 µl). Chloroform and cell debris were separated by centrifugation (14000 x g). The aqueous supernatant was transferred to a fresh Eppendorf tube and DNA was precipitated using isopropanol.
As an alternative to the classical mini-prep method described above, DNA-binding spin-columns (Amersham, Buckinghamshire, UK) were used to purify and concentrate DNA, following the manufacturer's instructions. DNA samples that were cut from an agarose gel were always recovered using spin-columns.

2.3.5 Butanol extraction

Dry butanol (1- or 2- butanol) extracts water from a DNA preparation. This allows the volume of the sample to be reduced until it is suitable for isopropanol precipitation. The sample (2.4 - 10 ml) was transferred to a centrifuge tube (15 ml) that was then filled with butanol. The preparation was mixed thoroughly and a centrifugation step (6000 x g) separated the phases. The butanol-phase was removed and the procedure was repeated until the volume of the aqueous sample was reduced sufficiently (1.2 ml). The sample was distributed over two Eppendorf tubes and, after a short centrifugation step (6000 x g), the last traces of butanol were removed with a Pasteur pipette. The sample was now ready for isopropanol precipitation.

Molecular Biology Part III Manipulation of DNA

2.3.6 *In vitro* amplification of DNA by PCR

In vitro amplification of DNA molecules was performed using the Polymerase Chain Reaction (PCR). The composition of the PCR reaction mixture and the PCR temperature-cycling programme are listed in Table 2.1 A and B.

Table 2.1. PCR Reaction mixture & conditions

The conditions listed here were used throughout the thesis (with exception of the mutagenic PCR protocol). Template: Topo-*nerA* plasmid, primers Topo Fw and Topo Rv (Table 2.2) unless indicated otherwise.

(A) PCR reaction	
Forward primer	100 ng
Reverse primer	100 ng
dNTP's	200 µM
$MgCl_2$	1.5 mM
Reaction buffer	1x
Taq polymerase	2 U
Template DNA	<10 ng
Ultra pure water	To 50 µl

(B) PCR conditions		
Hot start	96 °C	1 min.
30x		
Denaturing	96 °C	1 min
Annealing	52 °C	1 min
Elongation	72 °C	2 min
Final Elongation		10 min

The characteristics of PCR primers are listed in Table 2.2.

The polymerase (source: *Thermus aquaticus* strain YT1), $MgCl_2$ and buffer were supplied by Promega. Multiple volumes of the entire reaction volume were prepared in a sterile Eppendorf tube. *Taq* polymerase was added last. After mixing by inverting, the PCR mix was divided in 50 µl portions in thin-walled PCR tubes (100 µl, BDH, Poole UK). The PCR programme was performed in a Primus (MWG biotech, Milton Keynes, UK) thermal cycler. In most cases PCR primers Topo FW and Topo Rv were used (Table 2.2).

PCR screen

PCR was used to establish whether a sample of microbial culture or purified DNA contained the Topo vector, whether the Topo vector contained an insert and the length of the insert. The PCR mixture was prepared as usual but template DNA, in the form of cell culture (1 µl) or purified DNA (< 10 ng) was added last. PCR primers that anneal upstream and downstream of the insert in the Topo vector were used (Topo Fw and Topo Rv, see Table 2.2). If no insert was present a short (< 0.5 kb) DNA fragment was obtained. By using a PCR primer that anneals with the start of *nerA* in combination with a PCR primer that anneals downstream of the insert in the Topo vector, the correct orientation of the *nerA* insert could be verified (*nerA* start1 and Topo RV, see Table 2.2).

Mutagenic PCR

PCR was used to generate mutated copies of the *nerA* gene. The Topo-*nerA* plasmid (Figure 2.1 B) was used as a template. Regions flanking the insert were co-amplified in the PCR reaction. These flanking regions contained the *Nco I* and *Hind III* restriction sites that allowed re-introduction into the Topo ∅ vector. Primers Topo Fw and Topo Rv were used (see Table 2.2). The protocol was based upon regular PCR protocol but in order to increase the mutagenic rate the $MgCl_2$ (7 mM) and Taq polymerase (5 Units) concentrations were increased and the PCR reaction mixture was supplemented with manganese (0.1 mM $MnCl_2$). Elongation time was increased (10 min). The experiments underpinning this optimisation are presented in Section 4.1.

2.3.7 Restriction digestion

Restriction enzymes were used according to the manufacturer's recommendation. If DNA samples were suspected to be resistant to digestion (for example due to inhibitory substances in the sample) the sample was subjected to a range of concentrations of restriction enzyme. If DNA was digested for the use in cloning steps a large scale reaction was used (total volume 500 µl, DNA 5-20 µg, restriction enzymes 20 U each, BSA 10 µg, appropriate buffer according to manufacturer's recommendation). Addition of BSA during digestion was found to be crucial for the success of the subsequent ligation reaction, even if the manufacturer did not recommend the use of BSA with that particular restriction enzyme.

2.3.8 Ligation

New plasmids were constructed by ligating DNA molecules with cohesive ends that were obtained by restriction digestion. Each reaction (10 µl) contained ligation buffer (10x, 1 µl, Appendix B), ligase (10 Units / µl, 1 µl), linear plasmid (100 ng) and insert (30-200 ng). Negative control-reactions contained only either plasmid or insert. The reaction mixture was incubated (16 °C, 18 h) and the ligase was heat-inactivated (65 °C, 10 min). The plasmid construct was introduced into *E. coli* (Section 2.3.2 and 2.3.3). The ligation mixture could be analysed on agarose gel. If the ligation reaction had been successful ladders of multimers could be observed.
PCR products were also cloned directly into the Topo vector according to the manufacturer's recommendations (Section 2.1.2).

Molecular Biology Part IV Analysis of DNA

2.3.9 Agarose gel electrophoresis

DNA samples were separated on the basis of size by gel electrophoresis. A molecular weight marker, 1 kb DNA Ladder (New England Biolabs) was used to estimate the size and quantity of the DNA fragments in each sample.
DNA fragments were separated using 1% (w/v) electrophoresis grade agarose (Gibco BRL) gels prepared in 1 x TBE buffer. Ethidium bromide was added (0.5 μg /ml) to the gel and the running buffer. Electrophoresis was carried out in a horizontal electrophoresis gel tank (Bio-Rad) at a constant voltage (10 to 20 V/cm) until DNA fragments were separated sufficiently. Ethidium bromide-bound DNA molecules were viewed using an UV trans-illuminator. TAE buffer was used if DNA samples were to be recovered from gel (Sambrook, 1989).

2.3.10 Quantification

Prior to nucleotide sequence determination the concentration of DNA in each sample was measured by the attenuance at 260 nm. Before every digestion or ligation step, the total amount of DNA in a sample was measured using a fluorescent dye according to the manufacturer's recommendations (pico green, Molecular Probes, Leiden, the Netherlands). DNA samples of increasing dilution were prepared in TE buffer.
During optimisation of this protocol it was observed that dilutions made directly in dye solution give irreproducible readings. The fluorescence (excitation 480 nm, emission 520 nm) was measured on a plate reader (Bio-Tek FL600, Fisher scientific). The readings of the standard were used to draw a calibration graph. The readings of the unknowns that were within the linear region of the calibration curve were used to calculate the concentration of DNA in the sample.

2.3.11 Automated DNA sequencing

The sequence of the bases in a DNA sample was determined using the D-Rhodamine method and an Applied Biosystems (Warrington, UK) Prism 377 sequencer (G. Lewis, personal communication). Samples were plasmids containing a mutated *nerA* insert. The resulting chromatograms were translated into sequence information. Single mutations were found by alignment of at least three sequencing runs. If there was a disagreement between runs, the original chromatograms were used to establish which run was more reliable. If there were less than two runs with good quality an addition sequencing reaction was performed. The primers that were used for the sequencing are listed in Table 2.2.

Table 2.2 PCR primers

Table of PCR primers that were used when sequencing the *nerA* insert in the Topo vector. The direction of the sequencing run is indicated (+ coding strand, – non coding strand). The position of the 3' end of the PCR primer relative to the start of the *nerA* insert is also indicated.

Name	Sequence	Direction	Position
Topo Fw *	GAGGTATATATTAATGTATCG	+	-100
nerA outw 37 start	ATCGCCGGCCTGTGCCGGTTC	-	15
nerA start 1	ATGACCAGTCTTTTCGAACCG	+	21
J2I rvs	AGCGATTGCGGTGAGA	-	70
J2J fwd	TCTCACCCGCAATCGCT	+	85
J2S rvs	TCGGCCCACGTGCCAGATTTG	-	290
J2M fwd	CAAATCTGGCACGTGGGCCGA	+	310
J2R fwd	GATGATTACGGTGGCTCCAT	+	611
J2Q rvs	ATTGTTGGCGATCCACAG	-	900
J2P fwd	CTGTGGATCGCCAACAAT	+	918
J2T rvs	CCATAGAAGGTCGGCTGGTTC	-	1050
nerA stop	CTATTGGGCGAGGGCCGGATAG	-	1095
nerA outw 1080 stop	CGAAGGTTACACCGACTATCC	+	1101
Topo Rv *	CTGATTTAATCTGTATCAGGC	-	1466

* The Topo Fw primer anneals upstream of the insert and the Topo Rv anneals downstream of the insert.

2.4 Activity assays

2.4.1 Whole-cell assay

The nitrate-ester reductase activity of a sample of a microbial culture was routinely evaluated by NO_2^- production in a set of micro titre plates (96 well, flat bottom, clear polystyrene). A number of substrate and sample controls were included in the assay as described below.

Substrates

An appropriate amount of substrate solution was prepared freshly in LB medium, according to Table 2.3. Additionally control substrates (labelled "Negative", "Viability1 and 2" and "Consumption") were prepared in LB medium.

Table 2.3. Composition of substrate stocks and controls

	Concentration (mM)	Control substrate	Composition
EGDN	10	Negative	LB medium only
GTN	2	Viability 1	Tetrazolium** (1.3 mM)
PETN	8	Viability 2	NO_3^-(10 mM)
TNT	5	Consumption	NO_2^- (100 μM)
RDX	10		

** Tetrazolium (violet) : 2,5-diphenyl-3-α-naphthyltetrazolium chloride

The negative control substrate was included to detect the background concentration of NO_2^- in the samples. The viability control substrates were included to compare the metabolic activity among samples. Tetrazolium in control substrate "Viability 1" was routinely used to measure microbial activity (Gabrielson, 2002).
Control substrate "Viability 2" measured the nitrate-reductase activity of a sample as an additional indication of microbial activity. The consumption control was included to correct for interference of the NO_2^- detection by the sample, particularly caused by the consumption of NO_2^- by the microbial culture during the incubation of the assay. The major role of the control substrates was to identify samples that were unsuitable for analysis because of low viability or excess nitrite consumption.

Samples

Microbial cultures were grown as described in Section 2.2.1. Negative (wild type *E. coli*) and positive (*E. coli* expressing wild type *nerA*) biotic controls were included. Additional abiotic control samples were prepared in LB (Negative control: Medium only, Positive control: NO_2^- (100 μM)). The negative and positive control samples provided a calibration for the colorimetric detection of NO_2^- in the context of each individual substrate. The controls corrected for background attenuance, background NO_2^- concentration and interference with the colorimetric reaction or detection by the substrate.

Incubation and colorimetric detection of NO_2^-

Each sample (100 μl) was combined with each substrate (100 μl). The micro titre plates were incubated (ambient temperature) until the tetrazolium turned red, but not for more than 60 minutes. The nitrite concentration in all wells, except those containing tetrazolium or TNT was determined as described below.

A standard curve of NO_2^- (0 – 50 μM) was prepared and included in the analysis. Reagent A (50 μl, Appendix B) was added to all wells. This also stopped the enzymatic reactions. The blank attenuance of the plate was measured at 540 and 650nm (see below). After further incubation (minimal 1 min, but no longer than 15 min) reagent B (50 μl) was added. The reaction of NO_2^- with the azo-dyes in Reagent A and B gave raise to a violet colour (Litchfield, 1968) that is detected colorimetrically (Thermo max micro-well reader, Molecular Devices). Attenuance at 540 nm was measured and corrected for the (non-specific) attenuance at 650 nm. The blank attenuance that was read earlier was subtracted. The concentration of NO_2^- in the test samples was calculated using the standard curve. When required, samples were diluted and re-analysed.
Reduced tetrazolium (red) was also detected by the wavelengths used here.
Reduced TNT (orange) was detected qualitatively, by eye.

2.4.2 Cell free assay

The ability of an enzyme preparation or crude cell extract to reduce GTN was evaluated by measuring NADH oxidation in a micro-titre plate (96 well, flat bottom, clear polystyrene). An enzyme preparation (100 µl, containing between 1 and 50 mUnits of enzyme activity) was combined with a substrate solution (100 µl, 2 mM GTN, 1.2 mM NADH, 50 mM potassium phosphate buffer pH 7. Negative controls lacked GTN or NADH). If high background activity towards NADH was suspected (for example due to oxygenases present in the crude cell extract), the sample was supplemented with an oxygen-depleting agent (Na_2S, final concentration in assay 10 mM).

NADH concentration was monitored continuously by UV attenuance (340 nm, every 6 seconds over a period of 30 seconds to 15 minutes, ambient temperature in Thermo max micro-well reader, Molecular Devices). GTN could be replaced by any other substrate that was soluble in water and that did not interfere with the UV attenuance measurement. Colorimetric detection of liberated nitrite was not possible due to interference of NADH with the azo dyes in Reagent A and B.

The starting concentration of NADH (0.6 mM) gave an attenuance (D_{340}) of around 1 as was expected { [NADH] x ε_{NADH} x pathlength = 0.6 x 10^{-3} x 6200 x 0.3cm = 1.1 }. The decrease of UV attenuance in time (Δ D / min) was converted to Units of enzyme activity per well (or 100µl sample) by using Equation 2.1.

$$Enzyme\ Activity = \frac{reaction\ volume \times -\Delta D/\min}{\varepsilon_{NADH} \times pathlength} \times 10^6 = \frac{2x10^{-4} l \times -\Delta D/\min}{6200 M^{-1}cm^{-1} \times 0.3cm} \times 10^6$$

$$= -0.109 \times \Delta D/\min$$

Equation 2.1 Enzyme activity

Enzyme activity is a function of reaction volume, decrease of D / min (negative sign indicates that consumption of NADH results in positive activity), extinction coefficient of NADH and light path length = height of micro well. Definition: One unit of enzyme activity reduces 1 µmol of NADH per minute

2.4.3 High throughput characterisation

Mutants of *nerA* were expressed in *E.coli*. Cell free extract was obtained, desalted and submitted to gel filtration. The fractions that eluted from the gel filtration column were assayed for activity towards various nitro-organic compounds (six substrate analyses). The detailed protocol is described below.

Cell free extract

E.coli cultures expressing mutants of *nerA* were harvested (50ml, 3 to 4 h after incubation, D_{600} = 1) by centrifugation (6000 x g). The cell pellet was resuspended in TK buffer (5ml, Appendix B), supplemented with EDTA (final concentration 5mM) and lysozyme (5 µl x 1% w/v stock solution). After incubation (30 min on ice) DNAse (5 µl x 1% w/v stock solution), $MgCl_2$ (final concentration 10mM) and deoxycholate (3α, 12α, dihydroxy-5β-cholan-24-oic acid (mono Sodium salt), 5 µl x 1% w/v stock solution) was added. Further incubation (10 min on ice) allowed the disruption of the cell membrane, digestion of DNA and the escape of periplasmic enzymes. Cell wall and other debris were removed by centrifugation (6000 x g) followed by filtration (0.2 µM). The cell-free extract was desalted (PD 10 column, Pharmacia) according to the manufacturers recommendations and submitted to gel filtration

Gel filtration

The gel filtration column consisted of a glass column (Pharmacia XK16/70, length 700 mm, internal diameter 16 mm) filled with gel filtration medium (Superdex 200 prepgrade, Pharmacia, 130 ml). The column was connected to the FPLC set-up and operated at maximum recommended flow rate (50 cm / h, 0.7 ml/min). When the medium was settled (63 cm, column volume is 128 ml) the dead volume above the medium was eliminated by lowering the plunger to the top of the packed bed. The gel filtration buffer was of moderate salt strength to avoid dissociation of tertiary structures (100 mM NaCl, 50 mM potassium phosphate buffer pH7).

The desalted sample (2.5ml) was placed in the sample loop. The FPLC pump was started and after initial equilibration (3 ml) the sample was injected (2 ml). The pump was stopped after all fractions had washed off the column (140 ml). However the next sample could be loaded after 90 ml if the previous sample did not contain an excess of small solutes (Figure 2.2).

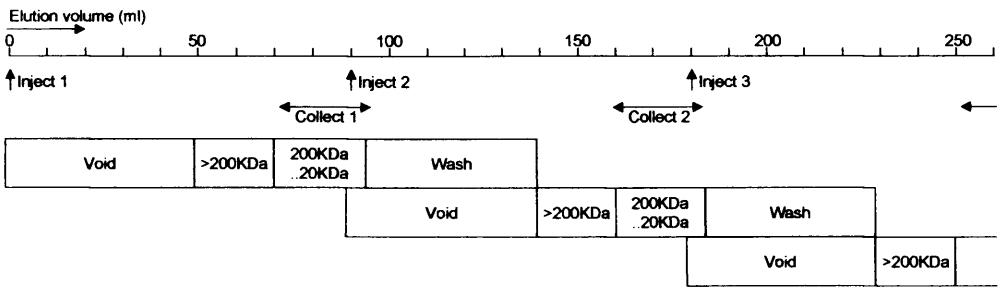


Figure 2.2 Continuous loading of Gel filtration column

To save time, the second sample could be loaded while the fractions of the first sample were collected. The washing of the small eluates from sample 1 coincided with the void volume of sample 2.

Fractions (1 ml) were collected between 71 ml and 95 ml eluent (corresponding approximately to the range of 20 to 200 kDa). The fractions were assayed for reductase activity as described below. The gel filtration column was calibrated for two sets of standards as described below.

The elution volume of each standard was determined from the UV trace. The elution volume of Blue dextran corresponded to the void / exclusion volume of the column. The total volume of the column was calculated from diameter and height of the column (128 ml).

The retention factor (Rf) of the standards was calculated according to Equation 2.2

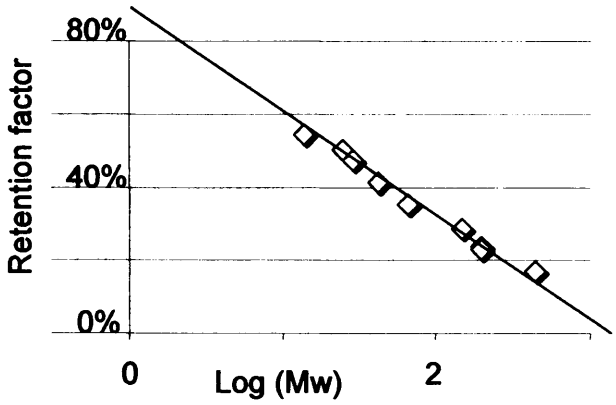
$$Rf(\%) = \frac{Elution\ volume\ sample - Void\ volume}{Total\ column\ volume - Void\ volume}$$

Equation 2.2 Retention factor

The retention factor was platted against the Log (Size) of various standards to obtain a calibration curve. With help of the calibration curve the molecular weight of an unknown substrate could be estimated (Figure 2.3).

Calibration substrates with molecular weights ranging from 14 to 440 kDa were effectively separated.

Substrate	Size (kDa)	Rf Run1	Rf Run 2
Blue dextran	> 1000	0%	0%
Apo ferredoxin	443	17%	
β Amylase	200	24%	23%
Alcohol dehydrogenase	150	29%	
Albumin	67		35%
Ovalbumin	43	41%	
Carbonic anhydrase	29		47%
Chymotrypsin	25	50%	
Ribonuclease	14		54%



$$\text{Size (kDa)} = 10^{(3.22 - 3.77 \times Rf)}$$

Figure 2.3 Calibration curve gel filtration

The gel filtration column was calibrated for two sets of standards (Run 1 and Run 2).

Six substrate analyses

The fractions that eluted from the gel filtration column were assayed for activity towards six nitro-organic compounds. Substrate stock solutions were prepared according to Table2.4

Table2.4. Substrates for reductase assay

Substrate stock solutions were prepared in KP50 (Appendix B) and contained NADH (final concentration 0.6 mM). Stocks were used within 2 hours. The stock solutions were 2x concentrated to compensate for the dilution by the sample.

Compound	Final Concentration in Assay (mM)
Blank	
GTN	1
GTN	0.15
TNT	0.5
EGDN	2.5
PETN	1
1,3 Dinitrobenzene	2.5
1-Chloro-2,4-dinitrobenzene	2.5

The assay was started by combining the sample (100 µl) with substrate stock solution. Consumption of NADH was followed and enzyme activity calculated as usual (Section 2.4.2). The results of a high-throughput characterisation were displayed in an activity diagram: the reductase activities in a certain fraction were plotted (on the Y-axis) against the size (X-axis) that corresponds to that fraction (see for example Figure 6.5).

2.4.4 Library screen

Spray gun

An aerosol spray gun (Fisher Scientific UK, supplied with a 15 ml glass bottle, Figure 2.4) was used to apply substrates or developing agents in reproducible doses to agar plates.

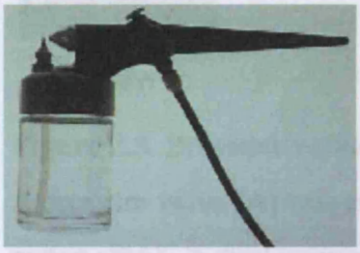


Figure 2.4. (A) Aerosol spray gun (B) Alternative reservoir

The gun was resistant to all solvents that were routinely used to dissolve explosives. When it was necessary to spray an exact volume, the glass bottle was replaced by a screw cap Eppendorf tube (2 ml, Figure 2.4B) The lid of the Eppendorf tube was pierced and the tube of the spray nozzle was passed through it.

The manufacturer of the spray gun suggested the use of a can of aerosol propellant to power the spray gun. However, the can cooled down during use (adiabatic cooling) and the delivered pressure decreased. This was unacceptable, as the performance of the gun was dependent on pressure. Instead the gun was powered by the outlet of a bench-top air pump, moderated (2 Bar) by a pressure valve (Figure 2.5).

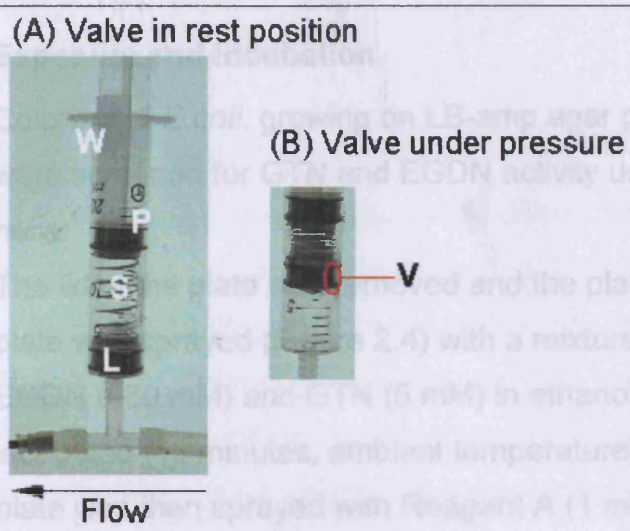


Figure 2.5. Pressure valve

A pressure valve (A) was constructed from a plastic syringe barrel (20 ml), fitted with a spare piston seal L, a steel spring S (500 N/m), and a syringe piston P which was fixed into position with a wedge W. The applied air pressure pushed seal L up the barrel, compressing the spring as shown in B. The pressurised air could only escape through a vent V at the 7 ml mark. The performance of the valve was found to be dependent on the mobility of the seal, which is dependent on a smooth inner surface of the barrel. Therefore great care was taken when the vent was cut in the barrel.

Exposure and incubation

Colonies of *E.coli*, growing on LB-amp agar plates and expressing mutants of *nerA* were screened for GTN and EGDN activity using the screening method described here.

The lid of the plate was removed and the plate was placed in a fume-cabinet. The plate was sprayed (Figure 2.4) with a mixture of nitrate esters (1 ml, containing EGDN (660 mM) and GTN (5 mM) in ethanol unless indicated otherwise) and incubated (15 minutes, ambient temperature) to allow for the release of nitrite. The plate was then sprayed with Reagent A (1 ml), incubated (1 min) and sprayed with Reagent B (1 ml), to detect nitrite colorimetrically. The plate was incubated until a violet colour developed. Alternatively the plate was sprayed with TNT (1 ml, 100 mM in acetone) or tetrazolium (1.3 mM in water see Section 2.4.1). No further spraying was required as reduced TNT was orange and reduced tetrazolium was violet. The experiments underpinning this optimisation are presented in Section 5.2.1

Plasmid Rescue

The spray reagents A and B cause complete loss of cell viability in the colony, so it was not possible to subculture colonies of interest. In order to identify mutations in the *nerA* inserts, it was necessary to rescue the plasmids carrying the mutant genes. Colonies of *E. coli* showing interesting features in the screen were picked from plates as soon as was feasible. Care was taken to pick the entire colony and to harvest as much biomass from the plate as possible. The colony was resuspended in buffer P1 and plasmid was extracted from the suspension as usual (Section 2.3.4). The plasmid was re-introduced into *E. coli* and a culture expressing the mutant *nerA* was obtained.

2.5 Protein purification

2.5.1 Choice of chromatographic system

Both chromatographic systems described in this Section were a type of liquid chromatography where pumps controlled the solvent velocity. The following parameters played a role in the choice of chromatography system for the purification of protein. If it was important that the flow rate was high (>5 ml/min, large scale purification) or well-defined (gel filtration) the FPLC system was preferred. If temperature played an important role or if the dead volume of the system needed to be small (small scale optimisation), the Bio Logic system was preferred. Although both systems can load samples of high volume, the peristaltic pump of the Bio Logic system was more suitable to load large volumes than the displacement pump of the FPLC system, because the latter had higher dead volume and required more extensive washing.

Fast Protein Liquid Chromatography (FPLC)

The FPLC apparatus (Pharmacia) consisted (in order of flow direction) of two syringe pumps (P500), gradient mixer (GM-1), injection valve with displacement loop (<5 ml, constructed from two syringes), column of choice, UV recorder (254 nm attenuance, D < 0.5), and a fraction collector (2 ml) including by-pass valve (waste / collect). Sample loading and elution programmes were controlled by a programmable device (LCC500).

When the FPLC column and pipe work was not in use it was stored with buffer (20% v/v ethanol in KP50 buffer) that prevents microbial growth. The fraction collector was a converted multi-well titration robot (Titertek, Kirkham, UK). A PC that was slave to the LCC500 controlled the device.

Bio Logic system

The Bio Logic system (Bio Logic LP) consisted (in order of flow direction) of 3 storage vessels (A,B,C, for buffers), an inlet valve that allowed selection of buffers, mixer, peristaltic pump (maximum flow rate 5 ml/min), by-pass valve that allowed the flow to be diverted around the column, column of choice, UV and conductivity detector, fraction collector including by-pass valve (waste / collect). The system was controlled by a programmable device.

The peristaltic pump was calibrated at zero pressure. The flow rate was assumed to be constant (as programmed) although the relationship between flow and pressure was unknown. The inlet valve could be programmed to switch frequently between buffer A and B and thus, in combination with the mixer, a gradient could be delivered. The flow was diverted around the column while the tubing of the system was washed. UV (254 nm) and conductivity of column effluents were registered continuously by a PC (data logging software "LP dataview" (Biorad)).

2.5.2 Cell free extract

E.coli Top10 carrying the Topo *nerA* plasmid and expressing *nerA* were cultured in a large volume (7 l, see Section 2.2.2), collected by centrifugation (6000 x g) and resuspended (KP50 120 ml).

A French press (internal diameter 25mm, American Instrument Company, Travenol laboratories Inc.) fitted with an input and an output valve was pre cooled to 8 °C and assembled. The French press was washed twice by passing cold buffer (KP50) through it.

The cell suspension was supplemented with DNase (100 µl of a 1% w/v stock). The suspension was drawn into the French press and extruded at a pressure of at least 110 kPa. The eluent was collected in an ice-cold vessel. Cell wall / membrane material and other debris was removed by centrifugation (6000 x g). The supernatant was further cleaned and sterilised by micro-pore filtration (5 µm pore-size, as the pressed sample was sometimes too concentrated to be filtered through 0.2 µm pore-size filters), transferred to disposable tubes, assayed for NerA activity and desalted.

A major disadvantage of using the French press to rupture cells is the fact that some of the cell membrane is torn into small particles that carry oxygenases. The oxygenases interfere with NADH dependent enzyme assays. The membrane particles also tend to clog the sterilisation filters. Enzymatic disruption of cells (used in Section 2.4.3) punctures rather than ruptures the cell membrane. The membrane fractions remain large enough to be removed by centrifugation.
The French press method was used only in combination with large scale culturing of *E.coli* to obtain heterologously expressed protein.

2.5.3 Desalting

A column (Hiprep 26/10 Pharmacia) was connected to a peristaltic pump. Flow rate was at maximum (10 ml/min). The column was equilibrated (250 ml) with KP50 buffer. Sample was loaded (15 ml) and after loading the eluent (17 - 20 ml) was collected. Between samples the column was equilibrated (150ml) with KP50 buffer. For long term storage the column was washed (100 ml, 0.1 M NaOH, reverse flow) and stored (250 ml, 20% v/v ethanol in KP50) according to the manufacturer's recommendations. The desalted cell extract was collected and submitted to anion exchange chromatography.

2.5.4 Anion exchange

Preparation of Anion exchange column

Diethylaminoethyl cellulose (16 grams of dry powder, DE22 {No longer available on the market} Whatman, Kent, UK) was "pre-cycled" (alternative acid and alkaline washes) according to the manufacturer's recommendation. All of the wet resin (100 ml) that was obtained was loaded into a column with a cooling jacket (Pharmacia XK50). The column was connected to an FPLC system and operated at controlled temperature (2-10 °C) and constant flow rate (10 ml/min). The plunger of the column was lowered as the medium settled. Optimisation studies indicated that this column had a binding capacity of 1.5 g protein.

Routine operation of anion exchange column

The column was pre-equilibrated with KP50 buffer (200 ml), loaded with sample as required and the column was rinsed with KP buffer until the UV attenuance (254nm) of the eluent returned to baseline level.
The sample was eluted in a salt gradient (100 ml, 0 to 0.3 M NaCl in KP buffer, followed by 100 ml 0.3 M NaCl in KP buffer) and fractions (10 ml) were collected as soon as the UV attenuance of the eluent rose above baseline level. The column was cleaned (200 ml 0.1 M NaOH, 200 ml ethanol / KP50 1:4 v/v), disconnected from the FPLC system and stored at room temperature.

2.5.5 Affinity chromatography

Blue Sepharose (6 Fast Flow, Pharmacia) contains groups that mimic NAD^+ and that can bind enzymes with affinity for NAD^+. These enzymes usually have an even higher affinity for NAD^+ or NADH and can be eluted using these cofactors. High salt concentrations can also weaken the binding of protein to column.

Preparation of affinity column

A slurry of Blue Sepharose (8 ml) was transferred to a column (internal diameter 12 mm). The column was connected to the Bio Logic system and operated at constant flow rate (0.5 ml/min). When the medium was settled (50 mm) the dead volume above the medium was occupied by lowering the plunger to the top of the packed bed. The column was cleaned with alkali (10 ml, 0.1 M NaOH) and high salt (10 ml, 1 M NaCl in KP50), equilibrated (20 ml KP50) and stored (20 ml ethanol / KP50 1:4 v/v).

Routine operation of the affinity column

The column was equilibrated (20 ml KP50) before use. The column was loaded with sample (1 - 200 ml) and rinsed (KP50) until the UV trace returned to base line level. The column was eluted with a gradient (0 - 2 mM, 50 ml) and a chase (2 mM, 20 ml) of NADH. Fractions (1 ml) were collected during elution. During optimisation of the protocol, effluent of loading and washing stages was also collected.

3 STATISTICAL ASPECTS OF THE EVOLUTIONARY CYCLE

The cycle of generating change and evaluating performance lies at the very heart of the exploration of the protein landscape (Section 1.2.1). The first two steps in the cycle (mutate and screen) have statistical parameters that are often neglected but that affect the success of the cycle. It is the aim of this Chapter to identify these parameters.

The XY plot that is going to be explored in the protein landscape must be defined first. This plot is in fact a collection of structures that have equal distance to the origin of that evolutionary step. As long as little is known about the structure / function relationship in NerA, all of these structures have equal potential with regards to improved catalytic properties and all of them should be generated and evaluated. If it is not feasible to evaluate all the structures, the plot must be limited. Section 3.2 investigates what kind of XY plot can be covered with mutagenic PCR. The collection of structures that represent an XY plot of choice can be predicted in most cases (for example the amino acid sequences of all structures that are one amino acid away from native NerA can be listed). The methods (mutagenic PCR, gene shuffling) that generate change can only approach the predicted list because these methods have a random character. In addition to this randomness the methods can be biased. In Section 3.3 the effects of randomness and bias on the composition of the offspring are investigated.

3.1.1 Offspring and screen size

Much of the offspring is useless background (redundant) because they do not contribute new points to the XY plot of choice. However in order to find the structures of interest, the entire offspring, including background, must be screened. On average, a change in structure will lead to degeneration of function, so the average level of plot lies below the origin. A good example of this is described in an article by Aita (1996); Most of the random mutations in the Lac promoter result in poorer expression of a reporter gene. Only a minority of mutations actually result in an improved performance. If a goal is set to explore a certain amount of XY space, it is paramount to explore *all* of it, so that those few structures with improved function can be found (Figure 3.1).

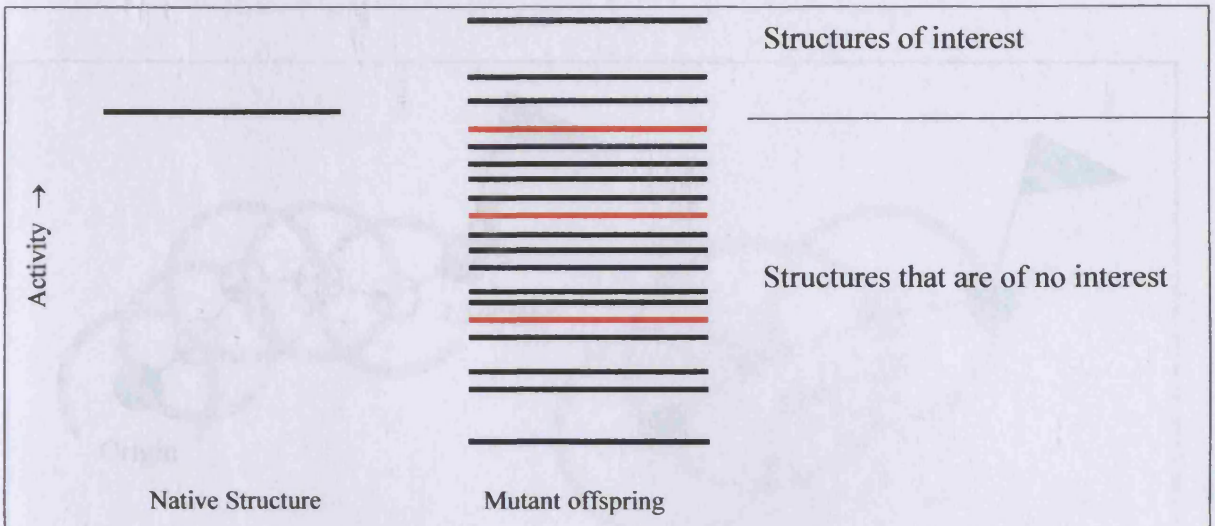


Figure 3.1 Degeneracy in Offspring

If a screen covers only a minority (red) of the offspring, the chance of finding a structure with enhanced activity becomes very small.

The evolutionary cycle is therefore optimal if

1. The XY plot that is explored is of maximal allowed size.
2. The offspring contain all[1] structures in that XY plot.
3. All of the offspring are passed through a screen.

Section 3.4 investigates the relationship between screen size and offspring. The size of an evolutionary step (XY plot, step 1) that can be explored is restricted by the size of the screen (step 2). Therefore it is necessary to optimise screen size to find a balance between resources available for screening against the number of evolutionary cycles to reach a new landmark in the protein landscape (Figure 3.2).

[1] In this context "All" means a number that approaches 100%.

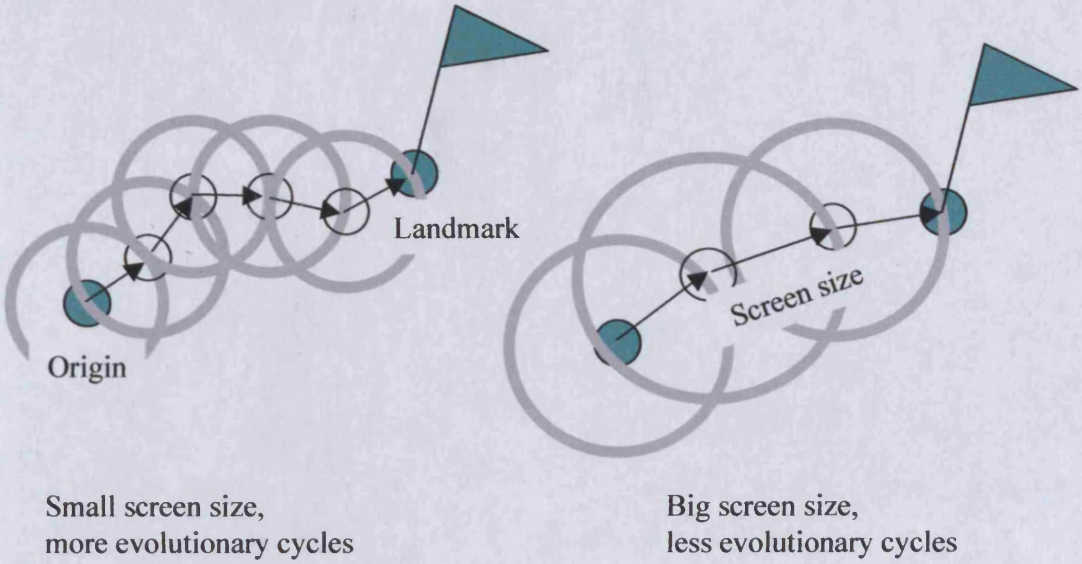


Figure 3.2 Evolutionary steps

The screen sets the size of the evolutionary step. If the step size is smaller, more evolutionary cycles are required to walk from the origin to a landmark.

3.2 Generating change with mutagenic PCR

3.2.1 Basic aspects of mutagenic PCR

In a normal double stranded DNA molecule a pyridine base (T or C) is paired with a purine base (A or G) (Figure 3.3). When Taq-polymerase incorporates a new nucleotide in a growing DNA strand it will match the appropriate purine to a pyrimidine in the native strand. In the event of a mismatch it is more likely that the polymerase incorporates the wrong purine opposite a pyrimidine than that it will incorporate a pyrimidine opposite a pyrimidine. Therefore mutations from A to G and from C to T (and *vice versa*) are more likely to occur (Vartanian, 1996). These mutations are called transitions. The mutations where a purine is changed to a pyrimidine or *vice versa* are called transversions.

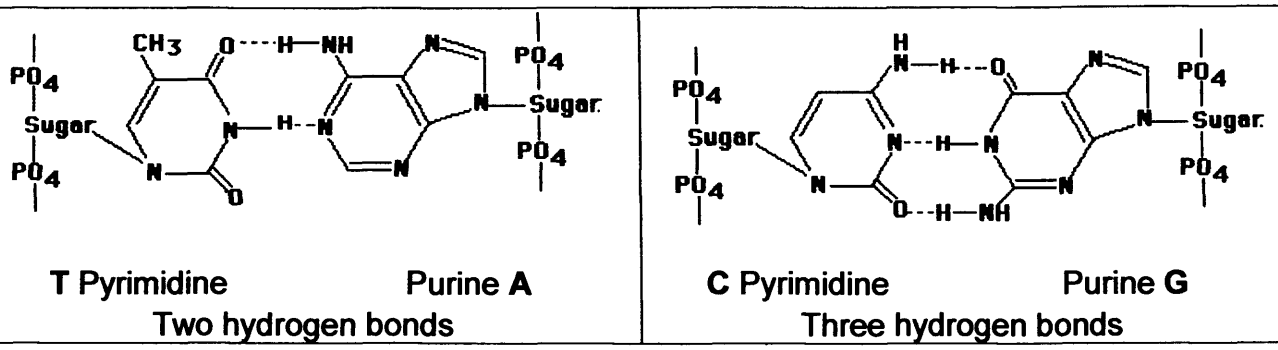


Figure 3.3 Base pairing in a double DNA strand

A pyrimidine pairs with a purine.

There are 12 mutagenic events (4 nucleotides can mutate into any of the 3 remaining nucleotides). Every event has its anti-coding counterpart. For example, T to G is equivalent to A to C in the opposite strand. The 12 events can be paired with their equivalent strand event to form 6 pairs. Two pairs are transitions, the remaining four are transversions. Transversions can be further classified according to number of hydrogen bonds in the nucleotides involved (see Table 3.1).

Table 3.1 Classes of mutagenic events

Classes in the same row cannot be distinguished when PCR based methods are used. The name of the classes (1..6) were chosen arbitrarily and were used only for cross reference in this thesis.

Class	Coding strand Event	Non-coding strand Event	"Mutation" in Hydrogen bonds
			Transitions
1	A to G	T to C	2 to 3
2	C to T	G to A	3 to 2
			Transversions
3	A to T	T to A	2 remains 2
4	C to G	G to C	3 remains 3
5	A to C	T to G	2 to 3
6	G to T	C to A	3 to 2

The numbers that were chosen to identify a class are arbitrary and were used only for cross-reference within this thesis.

If the mutagenic protocol is unbiased, every row in Table 3.1 is equally represented. The occurrence of every row would be 1/6, so a pool of six mutants could represent all possible mutations. However if the mutagenic protocol is biased, the rows are no longer equally represented. The class with the least occurrence will determine the size of the pool that represents all mutants (this matter will be further discussed in Section 3.3, see Figure 3.8).

Mutagenic PCR can be biased towards any of the classes mentioned in Table 3.1 or combinations thereof. See examples below.

- Preference for nucleotide at which mutation occurs
  (in this example bias towards classes 1,3 and 5).

Native	ACTG**A**CTGACTG**A**CTGACTGACTGATCG**A**TCG**A**TCGATCGA
Mutant	ACTG**C**CTGACTG**T**CTGACTGACTGATCG**G**TCG**C**TCGATCGA

- Preference for what nucleotide is used in place of the native one
  (in this example bias towards classes 2,3 and 6). This can be enhanced using biased nucleotide concentrations in the PCR mixture (Vartanian, 1996).

Native	ACT**G**ACTGACTGA**C**TGACTGACTGATCGA**T**C**G**ATCGATCGA
Mutant	ACT**A**ACTGACTGA**A**TGACTGACTGATCGA**A**C**A**ATCGATCGA

- There can be mutational hotspots in the sequence.

Native	ACTGACTGACTGA**C**TGA**CT**G**A**CT**G**ATCGATCGATCGATCGA
Mutant	ACTGACTGACTGA**G**TGA**AA**G**C**CT**C**ATCGATCGATCGATCGA

All these biases (PCR bias) influence the statistical parameters describing the profiles of the offspring of random mutagenesis.

3.2.2 Limitations of mutagenic PCR

In this Section *nerA* is used as an example. It is 1116bp long and encodes 372 codons. Consider the primary structure of 372 amino acids. A certain amino acid in this structure could be replaced by any of the 19 other amino acids, with help of site-directed mutagenesis. This way there are 372 x 19 = 7068 structures that are one amino acid away from native NerA in the XY plane. It is however not feasible to perform so many site-directed mutagenic manipulations.

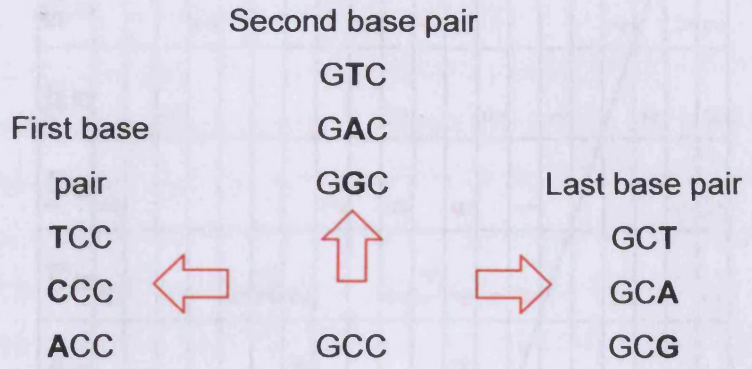
Random mutagenic PCR can manipulate the sequence of DNA more easily and is "only a translation away" from manipulating the primary amino acid sequence. Consider the gene of 1116bp. A certain nucleotide could be replaced by any of the 3 other nucleotides. In this way, there are 1116 x 3 = 3348 sequences that are one nucleotide away from native *nerA* in the XY plane. It becomes clear that mutagenic PCR, changing only one base per strand, cannot generate all 7068 structures that are one amino acid away from native NerA. This effect is referred to as "translational bias" in this thesis and it is further investigated in this sub-Chapter (3.2).

3.2.3 Translational bias

In the case of random mutagenesis not all exchanges are possible because rather than replacing an entire codon, only a single nucleotide is replaced (this is explained in Figure 3.4). In the case of *rational* (as opposed to random) mutagenesis, an entire codon is replaced with help of site-directed mutagenesis. Any of the 20 amino acids can be replaced by any of the 19 other amino acids. A total of 20 x 19 = 380 exchanges are allowed. These exchanges are represented in Table 3.2. Columns indicate native amino acids and rows indicate new amino acids. Table 3.2 is a 20x20 matrix. A diagonal line cancels 20 boxes that refer to a replacement of an amino acid by itself. The boxes that refer to exchanges that are allowed by a single mutagenic event are occupied by one or more numbers. The numbers refer to the mutagenic class (see legend or Table 3.1) that made that exchange possible. A table similar to Table 3.2 was first published by Mikes (1977), to study homology in proteins. The example in Figure 3.4 B shows how Table 3.2 was obtained. The software that was used to generate Table 3.2 can be found in Appendix C.

Figure 3.4 Single mutation of codon

A. A codon can be mutated at three different positions, and every position can be mutated in three different ways. This offspring off 9 mutant codons obviously does not represent all imaginable 19 other amino acids.



B. The codon GCC encodes for alanine. The codon has an offspring of 9 single mutated codons. Red arrows indicate the non-silent mutagenic events. The offspring is limited to 6 new amino acids (Ala to Val, Asp, Gly, Thr, Pro or Ser). These non silent events were achieved with mutagenic classes 2,6,4,2,4 and 6 respectively. This analysis was performed on all codons, resulting in Table 3.2.

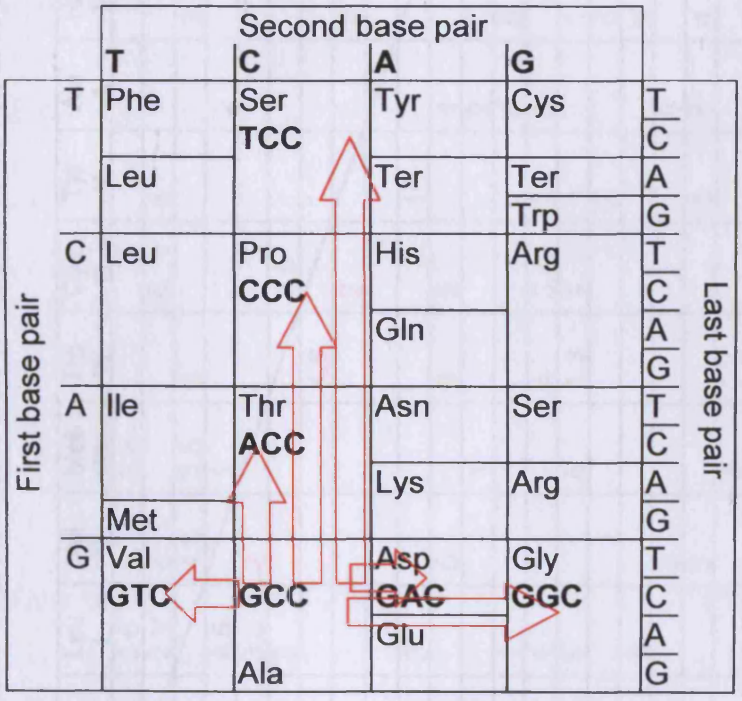


Table 3.2 Amino acid replacements possible as a result of one-point mutations

This table represents a 20x20 matrix of all amino acid exchanges that are imaginable. 20 boxes have been cancelled by a diagonal line because they refer to an amino acid being exchanged by itself. Exchanges that are possible as a result of a one point mutation are indicated by a number in the appropriate box. Numbers indicate what mutagenic class (see legend below) leads to that replacement.

Ile **I**	Phe **F**	Leu **L**	Val **V**	Met **M**	Trp **W**	Cys **C**	Tyr **Y**	Ala **A**	His **H**	Gly **G**	Thr **T**	Pro **P**	Ser **S**	Arg **R**	Asn **N**	Gln **Q**	Glu **E**	Asp **D**	Lys **K**	.	
	3	3,6	2	2,4							2		6	6	3				3	Ile	**I**
3		2,V	6			6	3						2							Phe	**F**
3,5	1,V		4,6	3,5	6				3			2	2	6		3				Leu	**L**
1	5	4,5		1				2		6							3	3		Val	**V**
1,4		3,6	2								2			6					3	Met	**M**
		5				4,5				6			4	2,3						Trp	**W**
	5				4,6		1			6			3,4	2						Cys	**C**
	3					2			2				6		3			6		Tyr	**Y**
			1							4	1	4	5				5	5		Ala	**A**
		3					1					6		2	5	V		4		His	**H**
			5		5	5		4					1	1,4			1	1		Gly	**G**
1				1				2				6	3,4	4	5				5	Thr	**T**
		1						4	5		5		1	4		5				Pro	**P**
5	1	1			4	3,4	5	6		2	3,4	2		V	1					Ser	**S**
5		5		5	1,3	1			1	2,4	4	4	V			1			1	Arg	**R**
3							3		6		6		2					2	V	Asn	**N**
		3							V			6		2			4		5	Gln	**Q**
			3					6		2						4		V	1	Glu	**E**
			3				5	6	4	2					1		V			Asp	**D**
3				3							6			2	V	6	2			Lys	**K**

Transitions		Class
Coding	Non Coding	
A to G	T to C	1
C to T	G to A	2

Transversions		Class
Coding	Non Coding	
A to T	T to A	3
C to G	G to C	4
A to C	T to G	5
G to T	C to A	6

The symbol "V" stands for any transversion (class 3,4,5,6).

3.2.4 Silent mutagenic events

Silent mutagenic events do not lead to an exchange of amino acid after translation. It is the goal of this section to determine exactly what fraction of mutagenic events are non-silent. The value of this fraction must be known in order to estimate the relation between mutagenic rate (average number of base pair changes / gene) and the average number of non-silent events in a gene.

In Table 3.2 silent mutagenic events have been cancelled by a diagonal line. Some of these boxes are actually occupied by events that belong to a certain mutagenic class. The program that was used to generate Table 3.2 was also used to take a closer look at silent events and events involving stop codons (Table 3.3).

The total number of mutagenic events is 64 codons x 9 = 576. The number of mutagenic events per class is 576 / 12 = 96. By subtracting the number of "Silent" and "Stop" events from the total number of events, the number of Non-silent events was found.

Table 3.3 Non silent events

The program that was used to generate Table 3.2 was also used to count the number of silent events and events involving a stop codon, per class (Columns "Silent" and "Stop"). The number of non-silent events was calculated by subtracting the "Silent" and "Stop" events from the number of mutagenic events per class (See text, results in Column "Non silent"). The fraction of Non Silent events was calculated by dividing "Non Silent" by the total number of events per class (96).

	Class	Silent	Stop	Non Silent	Fraction Non Silent
Transitions	1	31	7	58	60%
	2	31	7	58	60%
Transversions	3	18	14	64	67%
	4	16	4	76	79%
	5	19	9	68	71%
	6	19	9	68	71%
Total		134	50	392	68%

The total number of transitions is { 2 x 96 = 192 }. The total number of non-silent events that can be achieved with transitions is { 2 x 58 = 116 }. The fraction of transitions that are non-silent is therefore: { 116 / 192 = 60% }.

The fraction of transversions that are non silent is { (64 + 76 + 68 + 68) / (4 x 96) = 72% }. These fractions are used in Equation 7.4

3.2.5 Discussion of possible amino acid replacements

A few observations can be made in Table 3.2.

1. Only 150 out of 380 boxes in the 20x20 matrix are occupied. This supports the point that is being made in this section. Methods that rely on the incorporation of random single mutations can access only a limited number { 150 / 380 = 40% } of exchanges that can be imagined.
2. The majority { 122 / 150 = 81% } of the boxes that are occupied, are occupied by only one number. This implies that these amino acid exchanges can be achieved only by one particular mutagenic class. The implications of this are discussed by using mutagenic class 1 as an example. There are 22 boxes in Table 3.2 that are occupied by the number 1 only. These 22 boxes refer to amino acid exchanges that can only be achieved by a mutagenic event that belongs to class 1. If a mutagenic method is biased in such a manner that class 1 mutagenic events are very unlikely, the 22 amino acid exchanges that depend on this class also become very unlikely to occur. In this case the reach of random mutagenic PCR is further limited.
3. In the previous Section it was observed that there are 392 non silent events distributed among the six classes (Table 3.3). However, as was observed in point 1 on this page, only 150 boxes in Table 3.2 are occupied. This can be explained by the fact that there is great redundancy in the mutagenic events. For example the mutagenic event A to G can yield the exchange of Thr to Ala in four different ways. All four codons starting with AC all encode Thr. They can be mutated in all four codons starting with GC, all encoding Ala. To investigate this further the occurrence of the classes in Table 3.2 was counted (Table 3.4). The symbol "V" was of course counted as class 3,4,5 and 6.

Table 3.4 Occurrence of classes

The occurrence of every class in Table 3.2 was counted. For example the number "1" referring to a mutagenic event of class 1 was found 26 times in Table 3.2.

Class	Occurrence
1	26
2	26
3	40
4	36
5	37
6	37

Class 1 and 2, representing transitions, each occur 26 times in Table 3.2. Transversions occur on average { (40 + 36 + 37 + 37) / 4 } = 38 times (see Table 3.4). It must be concluded that there is greater redundancy in Classes 1 and 2, compared to Classes 3,4,5 and 6. Redundant events are repetitions and go at the expense of the information content of a certain class. The information content of transitions is lower than that of transversions due to the difference in redundancy. Earlier it was explained that mutagenic PCR is likely to be biased in favour of transitions (Figure 3.3). The fact that mutagenic PCR is biased in favour of transitions and that transitions have a smaller information content indicates that the genetic code is robust against changes.

When a mutagenic protocol is used to generate change it would be beneficial if this protocol was biased in favour of the class that has the highest information content: Class 3.

4. The cells in the first four columns and the top four rows of Table 3.2 are all occupied. This means that any of the four most hydrophobic amino acids (Ile, Phe, Leu and Val) can be exchanged by each other by a single mutagenic event. One could argue that the exchange of one hydrophobic residue by another hydrophobic residue has less impact on the overall enzyme performance than an exchange that would involve a great change in hydrophobicity. Indeed there is a hypothesis that the standard genetic code has evolved to be robust against single mutagenic events by minimising the effect of such an event (Freeland, 1998; Luo, 2002).

In Table 3.2 the amino acids have been placed in order of hydrophobicity to make the effect of robustness visible. Trinquier has taken the theory even a step further. Trinquier shuffled the order of the amino acids and tried to find in what order the robustness would become most visible. Trinquier then argued that this optimised order of amino acids has an evolutionary origin (Trinquier, 1998).
A possible source of robustness is inheritance from the "RNA world". The properties of a codon are reflected in the properties of the amino acids they encode (Di Giulio, 1989; Lehman, 2000). It has been shown that codons specifying amino acids that share the same biochemical synthetic pathway tend to have the same first nucleotide (Freeland, 1998). A comparison of these theories can be found in an article by Sella (2002). The implication of any of these theories is that single mutagenic events not only have limited access to amino acid exchanges but that the exchanges that are allowed are also relatively conservative in terms of physiological properties.

5. When Table 3.2 was first made, it was observed that classes 3,4,5 and 6 often co-occurred. Therefore the symbol "V" was introduced, referring to any transition. The explanation for this co-occurrence is very simple. The standard genetic code is usually represented in 16 groups of four codons that have the first two nucleotides in common. Eight of these 4-membered codon groups encode a single amino acid. In these codon groups mutation of the third nucleotide yields only silent events. Five other codon groups encode two amino acids that occupy two codons each (Phe and Leu, His and Gln, Asn and Lys, Asp and Glu, Ser and Arg). In these codon groups any transition leads to an exchange with the other amino acid in the codon group (Figure 3.5). Indeed in the columns in Table 3.2 that refer to the ten amino acids mentioned before, the symbol "V" is found. This indicates that these amino acids can be exchanged by the other amino acid in the codon group, with any transversion. It is must be observed that the "V" symbol is often found near the diagonal line. This means that two amino acids that are found in the same codon group are similar in terms of their hydrophobicity and that their exchange is relatively conservative. This further supports the point made in 3). The amino acid exchanges that are more likely to occur because they are accessible through four different mutagenic classes are generally more conservative.

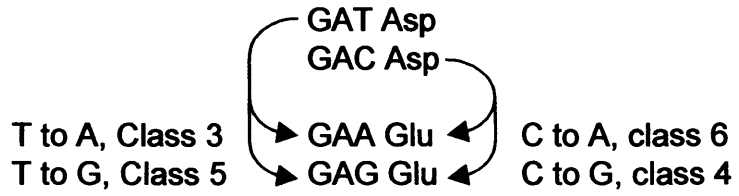
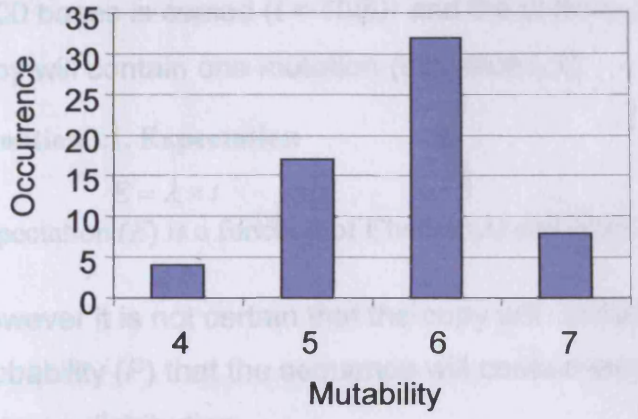


Figure 3.5 Transversions in codons that have first two nucleotides in common
The codons for Asp and Glu have the first two nucleotides in common. Any transition (class 1 or 2) will lead to a silent event. The exchange of Asp to Glu can be achieved with any transversion.

3.2.6 Codon mutability

In the program that was used to generate Table 3.2 every codon was mutated 9 times (Figure 3.4). In the example used in Figure 3.4 the codon GCC has an offspring of 9 mutant codons of which 6 are non-silent with regard to the original. Not all codons have an offspring of exactly 6 non-silent events as can be seen in Figure 3.6. Some codons have an offspring of only 4 non-silent events. An example of this is CGA. Mutation of the third nucleotide leads only to silent events, mutation of the first nucleotide leads to one stop codon (TGA), one silent event (AGA) and one non silent event (GGA) and mutation of the second codon leads to three non-silent events. Other codons have an offspring of 7 non-silent events. An example of this is CAU. Mutation of the first and second nucleotide each give raise to 3 non-silent events and mutation of the third nucleotide yields one additional non-silent event.



Average mutability =

349 (Sum) / 61 (Codons) =

5.7 Mutant offspring / codon

Figure 3.6 Mutability of codons

The number of non-silent mutant codons was determined for every codon not encoding a stop codon. The average mutability of a codon was calculated by dividing the total number of non silent events (349) by the number of codons (61).

Figure 3.6 indicates that an offspring of 6 non-silent events is the most common. All mutant codons together (349) divided by all codons that were manipulated (61) equals the average number of non-silent mutant codons per codon (5.7). This value is commonly used to estimate the average number of non-silent offspring of a codon (Voigt, 2001).

3.3 The offspring of mutagenic PCR

3.3.1 Poisson distribution

Random mutagenesis involves a large number of bases being copied (Trials t) and a very small chance (λ) of a mutagenic event to take place. Similarly screening methods involve a large number of individuals (Trials, t) to be screened but a very small chance (λ) of picking one unique individual. These two processes are best described using the Poisson distribution, which is therefore briefly introduced here. When a DNA strand is duplicated by Taq polymerase a great number of bases are processed. The mutagenic character of PCR means that every time a base is matched with its counterpart there is a very small chance of a mismatch. In statistical terminology the number of base pairs that are processed is t (Trials) and

the chance of a mismatch is λ. Intuitively one would expect (E) that if a sequence of 1000 bases is copied (t = 1000) and the chance of a mismatch is (λ = 1/1000), the copy will contain one mutation (Equation3.1).

Equation3.1. Expectation

$$E = \lambda \times t$$

Expectation (E) is a function of Chance (λ) and Trials (t).

However it is not certain that the copy will contain exactly one mutation. The probability (P) that the sequence will contain exactly x mutations is found using the Poisson distribution.

Equation 3.2. Poisson distribution

$$P(x) = e^{-\lambda} \times (\lambda t)^x / x!$$

Probability (P) of an event to take place exactly x times is a function of x, Chance λ and Trials t.

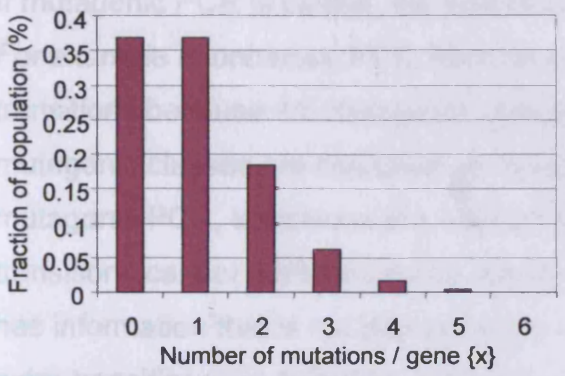
The statistics of screening also involve the Poisson distribution. The chance that a particular individual is picked is λ. The probability that it is not found (x=0) must be kept very small say <5% (P(0) < 5%). The number of colonies that must be screened (t) can be found using Equation 3.2 (this matter is further explored in Section 3.4).

3.3.2 Poisson distribution applied to mutagenic PCR

In Table 3.3, it was observed that between 60% and 79% of mutations of each class are non-silent. The goal is to generate 1 amino acid exchange per gene so the optimal mutagenic rate will be { 1 / (60% to 72%) = 1.4 - 1.6 } mutations per gene. There is no control over the position of the mutation, so there is no control over the actual number of mutations per gene. The Poisson distribution can be used to find how much of the population contains exactly one non-silent mutation.
Suppose the mutagenic rate is set at 1 non-silent mutation / gene { (λt) = 1 }. The probability that a gene contains exactly x mutations is

$$P(x) = e^{-1} \times 1^x / x! \qquad P(1) = 37\%$$

So the chance that the gene contains exactly 1 non silent mutation is only 37%.



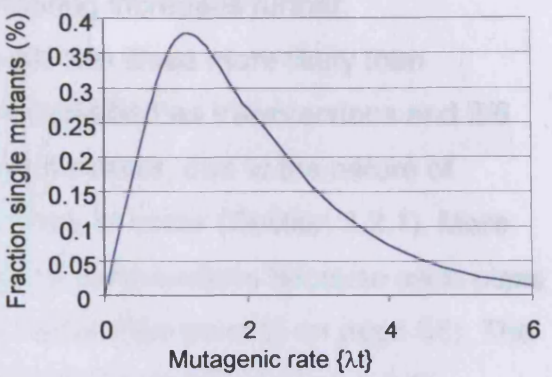


Figure 3.7.A. Composition of mutated pool **B. Fraction of single mutants as function of mutagenic rate**

A If the mutagenic rate is tuned to 1/gene (λt=1) only 37% of the pool will consist of mutants with exactly 1 mutation. The rest of the pool will consist of mutants with less (37%) or more (26%) mutations. **B** If the mutagenic rate is not well tuned, the fraction of single mutants drops further.

The offspring of mutagenic PCR contains a background of under mutated (native) and over mutated sequences (Figure 3.7A; {$P(x)$} versus {x}).
Usually there is poor control over the mutagenic rate. If the mutagenic rate is sub optimal, the fraction of useful sequences decreases and the background increases (Figure 3.7B; {$P(1)$} versus {λt}).

If mutagenic PCR is biased, the size of the offspring increases further. For example in unbiased PCR, transversions are two times more likely than transitions because 4/6 mutagenic classes are classified as transversions and 2/6 mutagenic classes are classified as transitions. However, due to the nature of mutagenic PCR, transitions are actually more likely to occur (Section 3.2.1). More transitions cannot compensate for the absence of transversions because each class has information that is not present in the other class (See point 3) on page 66). The extra transitions are therefore redundant and cause background (Figure 3.8).

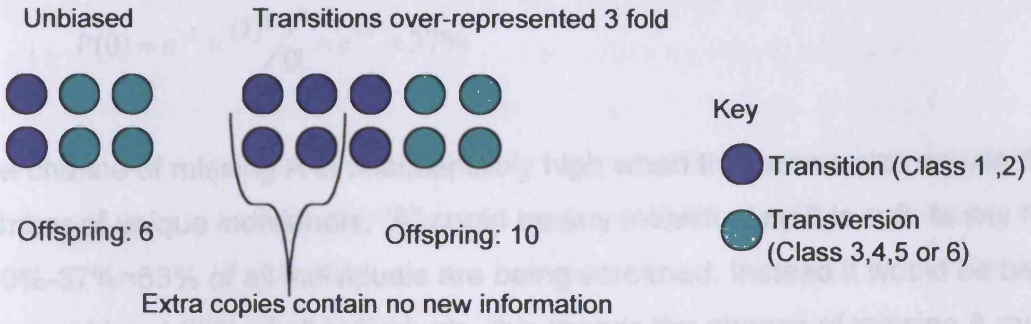


Figure 3.8 Increased background due to PCR bias

If transitions are 3-fold over represented , the redundant transitions cause background and an increase in the offspring. The two sets contain the same amount of information, but the biased set is 67% bigger in size (10/6 = 167 %).

3.4 Screen size

Screening can also be described with the Poisson distribution (Section 3.3.1). Suppose there is a pool of *N* unique individuals, containing one particular individual called A. The pool can be much bigger than *N*, but when randomly picking one sequence from the pool, the chance of picking sequence A is $1/N$ ($1/N = \lambda$). One screen can consist of many trials (t). Suppose there is a pool with 1000 unique sequences and 1000 colonies are screened: then $\lambda t = 1$. The probability that A is **not** present in this screen is $P(0)$:

$$P(0) = e^{-1} \times {(1)^0}/{0!} = e^{-1} = 37\%$$

The chance of missing A is unacceptably high when the screen size equals the number of unique individuals. "A" could be any individual so it is safe to say that only 100%-37%=63% of all individuals are being screened. Instead it would be better to screen at least 95% of all individuals, this means the chance of missing A must be < 5%.

$$P(0) = e^{-\lambda t} \times {(\lambda t)^0}/{0!} = e^{-\lambda t} = 5\%$$

{ λt = - Ln(0.05) = 3, with $\lambda = 1/N$ }. Screen size t must be 3 times *N*. Indeed as a rule of thumb **a gene pool must be "overscreened" 3 times to cover 95% of its content**

If you want to see all faces of the die, you'll have to throw more than six times.

Additionally there could be a contamination of the screen by individuals that contain no structure at all (for example self ligates).

3.5 Conclusion

The XY plot in reach of random mutagenic PCR

When using mutagenic PCR to generate change, a codon can be mutated in 9 different ways and, on average, these 9 mutants encode 5.7 new amino acids. For NerA (372 amino acids) mutagenic PCR can generate { 5.7 x 372 codons = 2120 } new structures that are one amino acid away from the native protein.
These structures represent only a fraction { 2120 / 7068 = 30% } of the total number of structures that are one amino acid away from native NerA. It is expected that these structures are relatively similar to the native structure because of the robustness of the standard genetic code to single mutations. It must also be remembered that random mutagenic PCR is sensitive to bias because lack of one certain class of mutagenic events can not be compensated for by abundance of another class.

Screen size

The offspring of the mutagenic PCR reaction will contain background of unwanted structures. If the mutagenic rate is optimal for the exchange of 1amino acid / gene the fraction of single mutants would be 37%. The remaining 63% is background of native or over-mutated sequences. For every (1 / 37% =) 2.7 members of the offspring one is a single mutant.
Under ideal circumstances (No PCR bias, optimal mutagenic rate) the size of the offspring will be { 2.7 x 2120= 5700 }. If the circumstances are not optimal, for example due to biased PCR, the composition of the offspring will be sub optimal. The screen size must then be increased to account for the increase in background. A screen must have the size of { 3 x 5700 = 17000 } to evaluate 95% of the 2120 new structures generated by mutagenic PCR.
Screening for double mutants would require the screening of at least { 3 x 2.7 x $(2120)^2 = 36*10^6$ } colonies. Usually the number of structures that can be evaluated (screen size) is limited due to logistical reasons. Ultimately this will limit the plot of XY plane that can be explored in a single evolutionary step. If the plot exceeds the screen size, the evolutionary cycle is sub optimal (Figure 3.1). Screening for double mutants is therefore not feasible.

4 RANDOM MUTAGENESIS OF *NERA*

4.1 Random Mutagenesis

The results obtained in this study are presented in the same order as that is used when performing an evolutionary cycle: mutagenic PCR and cloning of the PCR product (this Chapter), Functional screening of the mutants (Chapter 5) and finally Enzymatic characterisation (Chapter 6). However, the various protocols were developed simultaneously and not necessarily in a chronological order. For this reason there are several references in this Chapter to Sections that lie ahead. For example the protocol for cloning PCR product (Section 4.2) was developed before the protocol for mutagenesis (Section 4.1) was developed.

4.1.1 Mutagenic PCR

A number of ways have been described that decrease the fidelity of the PCR reaction. The most common procedure is to increase Mg^{2+} concentration or to complement the PCR reaction mixture with low concentrations of Mn^{2+} (Cadwell, 1992). If the PCR reaction was biased because preference was given to the incorporation of a certain nucleotide, that bias could be controlled by reducing the concentration of that nucleotide (Vartanian, 1996). Conditions that make PCR mutagenic are often sub-optimal for DNA replication. Therefore elongation time was extended (10min) and in some cases more *Taq* polymerase was included in the reaction mixture.

Increased concentration of magnesium

The first experiment (1) was based on the regular protocol described in Section 2.3.6, except that increases were made in the Mg^{2+} concentration (from 1.5 to 2.5 mM) and the elongation time (from 2 to 10 min). The PCR products were cloned and the resulting expression library was screened as described in Chapter 2. Colonies expressing functional *nerA* were red whereas colonies that did not express functional NerA remained white (Section 5.2.2).

Five white colonies containing mutants of *nerA* that no longer encoded a protein with activity towards EGDN or GTN were obtained. These mutants were analysed by sequencing and all base pair exchanges were identified. Some mutations were found in all strains. Possible explanations were that the deposited sequence contained errors or that the *nerA* gene on the plasmid that was the template for the mutagenic PCR reaction already contained a few mutations. The effect of mutagenic PCR on the *nerA* sequence was analysed using the sequence data of the mutants. The base pair changes were identified and counted. The results are shown in Table 4.1

Table 4.1. Base pair changes caused by mutagenic PCR (1)

Rows indicate the mutation class (See Table 3.1), the last column indicates number of occurrences in that class. Data obtained from 5 mutants. PCR conditions: Mg2+ concentration 2.5 mM, elongation time 10 min, template: the Topo *nerA* plasmid, primers: Topo Fw and Topo Rv. For other conditions see Table 2.1.

Class	Event	Number of mutations
Transitions		
1	A to G / T to C	1
2	C to T / G to A	12
Transversions		
3	A to T / T to A	2
4	C to G / G to C	0
5	A to C / T to G	0
6	G to T / C to A	1

The most common mutagenic rate in the population-fraction that was analysed was around 2 to 3 per gene. The average mutagenic rate was 16 per 5 genes ($\lambda t = 3.2$). Divided by the gene length (1.1kb) it amounts to 0.29% per base pair. One insertion and two deletion events were found per five genes and all these three events occurred at a guanidine. It is not known if it is common that insertion or deletion events are directed towards guanidine.

The majority of base changes were class 2 (For an overview of mutagenic classes see Table 3.1). This means that the mutations were biased towards transition, as was expected and reported before (Cadwell, 1992) (see also Section 3.2.1). There is only one occurrence of incorporation of C or G. There seemed to be a preference to incorporate A or T (base pair with 2H bonds). It was concluded that the mutagenesis was biased towards transitions and towards incorporation of an A or T. The opposite, where the mutagenesis were biased towards A to G and T to C transitions (Class 1), has been reported before (Cadwell, 1992). The extent of bias of the PCR protocol was unacceptable and an alternative protocol was developed.

Effect of manganese

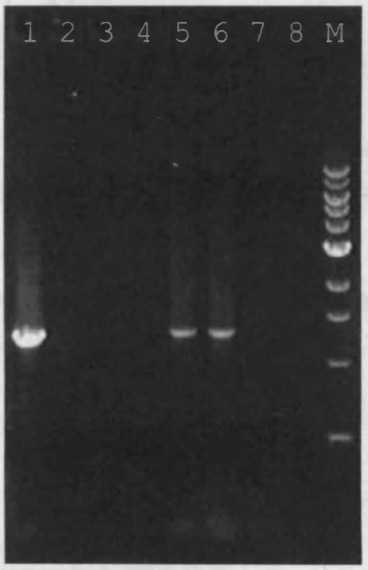
In a second experiment the influence of Mn^{2+} was examined. Initial experiments showed that Mn^{2+} had a strong inhibitory effect on PCR. However this effect was compensated by increased concentrations of Mg^{2+} (Figure 4.1A). The amount of *Taq* polymerase was also optimised (Figure 4.1B). Further iterations were performed to optimise the combination of Mn^{2+}, increased concentration of Mg^{2+} and *Taq* polymerase. The protocol that was the result of this optimisation can be found in Section 2.3.6 "Mutagenic PCR" (Mn^{2+} 0.1 mM; Mg^{2+} 7 mM; *Taq* polymerase 10 Units).

Figure 4.1. Optimisation of yield of mutagenic PCR protocol

(A) The concentrations of magnesium and manganese were optimised. PCR conditions: Mg^{2+} and Mn^{2+} concentration as indicated in Table (A) below. Elongation time 10 min, template: the Topo *nerA* plasmid, primers: Topo Fw and Topo Rv. Taq polymerase: 2U / 50 µl. For other conditions see Table 2.1. (B) Decreasing amounts of *Taq* polymerase were used. PCR conditions: Mg^{2+} concentration 2.5 mM, Mn^{2+} 0 mM. Taq polymerase as indicated in Table (B) below. Elongation time 10 min, template: the Topo *nerA* plasmid, primers: Topo Fw and Topo Rv. For other conditions see Table 2.1.

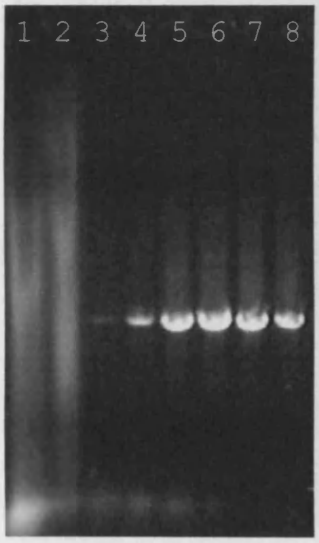
A. Mn^{2+} and Mg^{2+}

Lane	Mn^{2+} (mM)	Mg^{2+} (mM)	Yield
1	0	2.5	++
2	0.2	2.5	-
3	0.8	2.5	-
4	2	2.5	-
5	0.2	3.5	+
6	0.4	3.5	+
7	0.8	3.5	-
8	2	3.5	-



B. Taq polymerase

Taq Units	Yield
100	-
50	-
25	-
12	+
6	+++
3	++
1.5	++
0.8	++



4.1.2 Mutagenesis of *nerA*

The mutagenesis protocol as described in Section 2.3.6 "Mutagagenic PCR" was used to generate a new library of single mutants of *nerA* (Experiment (2)). Seventeen library members with NerA activity were sequenced and the characteristics of the new PCR protocol were analysed as discussed earlier in this section. The results are displayed in Table 4.2. Of the seventeen library members five were identical to native NerA, and two library members occurred twice in the tested population. These doubles were not taken into consideration when analysing the characteristics of the mutagenic protocol. The finding of doubles is however significant and will be discussed in Section 7.1.6. The characteristics of the seventeen mutants are described in greater detail in Section 6.4.2 and Appendix A.

Table 4.2 Base pair changes caused by mutagenic PCR (2)

Rows indicate the mutation class, the last column indicates number of occurrences in that class. Data obtained from 15 mutants (doubles were entered into the analysis only once). PCR conditions as described in Section 2.3.6 "Mutagenic PCR": Mg^{2+} concentration 7 mM, Mn^{2+} 0.1 mM. Taq polymerase 5U / 50µl. Elongation time 10 min, template: the Topo *nerA* plasmid, primers: Topo Fw and Topo Rv.

Class	Event	Number of mutations
Transitions		Total 15
1	A to G / T to C	3
2	C to T / G to A	12
Transversions		Total 9
3	A to T / T to A	5
4	C to G / G to C	2
5	A to C / T to G	1
6	G to T / C to A	2
Total		25

The mutagenic rate (λt) was optimal: around { 25 / 15 = 1.7 } mutations per gene. As observed before, there was still a bias towards transitions and towards incorporation of A or T. However all classes of mutagenic events were represented using this protocol.

The statistical population was too small to pass any judgement on frequencies of individual mutagenic classes but it was apparent that transitions were over represented. These results are compared to the results obtained by other authors in Section 7.1.

4.2 Cloning the PCR product

4.2.1 The "Topo – one step" cloning protocol

There was a system available that allowed PCR product to be directly introduced into an expression vector: the "Topo one-step" cloning system (Invitrogen, 1999). The *nerA* gene was amplified using a construct that was available in the lab (pLex-*nerA* , S.M. Marshall, personal communication.). The PCR product was introduced directly into the Topo vector according to the manufacturer's recommendations. The construct was introduced into chemically competent cells and the entire batch of cells was spread on a single selective LB-agar plate. The transformation reaction yielded an average of 20 colonies per plate.

Colonies were revived and subjected to the whole cell assay, to detect their ability to release nitrite from nitrate esters, according to the protocol in Section 2.4.1. Of the 46 cultures tested, 15 were found to have NerA activity. A number of cultures were also subjected to the PCR screen and the results of the PCR screen were in total agreement with the results of the whole cell assay. Cultures with no activity in the whole-cell assay showed up either as containing no insert or containing the insert in the wrong orientation. This one-step cloning protocol was not efficient enough to be used to generate comprehensive libraries of mutant *nerA*. The number of colony-forming units was low and only 30% of these colonies contained a plasmid with *nerA* in the correct orientation. An alternative cloning and expression system, based on the Topo vector but using a classical cloning technique was therefore developed.

4.2.2 Classical cloning protocol

Because the Topo cloning protocol was rejected, a classical cloning protocol using Topo ∅ as a receiving expression vector and the Topo vector containing *nerA* (Topo *nerA*) as a template for mutagenic PCR was optimised. Topo ∅ was obtained as described in Figure 2.1. Topo *nerA* was prepared as follows.

From one of the cultures that was obtained in the "Topo – one step" protocol (previous Section) that had nitro-reductase activity and that carried the Topo vector with the *nerA* insert in the right orientation, plasmid was extracted. The plasmid was named "Topo *nerA*-pLex".

A short sequence that originated from the original plasmid (pLex) was cut out using the *Eco RI* restriction site. The resulting plasmid was named "Topo *nerA*" (see Figure 2.1B) and used as a template for the generation of all mutagenic libraries described in this thesis.

4.2.3 Restriction Digestion and Ligation

Experiments indicated that the digestion of PCR product was greatly enhanced by the presence of BSA. It was also observed that the subsequent ligation reaction was very inefficient unless the DNA substrates had been digested in the presence of BSA. It was concluded that BSA must be included in the restriction digestion reaction, even if the manufacturer of the enzymes does not recommend this. These findings have been incorporated in the protocol that can be found in Chapter 2.

Mutated *nerA* (including flanking regions that facilitated re-entry into Topo ∅, see Figure 2.1) was obtained by mutagenic PCR as described in Chapter 2. The PCR product was purified by phenol extraction and 10 μg was subjected to restriction digestion as described in Section 2.3.7 (using Nco I and Hind III). Plasmid (Topo ∅) was prepared in a large scale (30 μg) and was subjected to an additional phenol extraction followed by column purification. The DNA was eluted in Tris buffer (50 μl, yield 10 μg) and subjected to restriction digestion (using Nco I and Hind III). The digestion reactions were incubated (37 °C, 1 h) and a small sample was analysed on agarose gel to confirm that the digestion was complete. The digestion reaction mixtures were supplemented with DNA loading buffer, loaded on an agarose gel (TAE buffer system) and the DNA fragments were separated according to size by electrophoresis (10V / cm, 100min). The appropriate bands were cut from the agarose gel and DNA was recovered using spin columns (Section 2.3.4).

The digested PCR products were ligated into the digested vector as described in Section 2.3.8. The digests were combined at various molecular ratios (Columns 1,2 in Table 4.3). The total amount of DNA was around 200 ng per reaction. The ligation reaction (1 μl) was introduced into *E. coli*, by electroporation. The host cells were dispersed on selective plates in various dilutions and analysed after incubation (37°C, 16h). The colonies on each plate were counted.

The results are displayed in Table 4.3. The preparation of insert alone did not give rise to any colonies so it was not contaminated with plasmid that might originate from the PCR reaction. The empty plasmid (lane 1) was found to capable of self-ligation to a certain extent. This gave rise to some background colonies that did not contain the *nerA*-insert. However this background (20 colonies per plate) was negligible in ligation reactions where the ligation of insert into plasmid had been very successful (reaction 5, 10^4 colonies per plate). The molecular ratio (1:1, reaction 5) gave the highest yield in terms of colonies / plate.

Table 4.3 Ligation of digested PCR product into digested Topo ∅.

Digested and purified plasmid (Topo ∅) and insert (*nerA* + flanking regions) were combined in various molecular ratios and ligated. The ligation product was introduced into *E.coli* and the number of colonies bearing plasmid was counted.

Reaction	Ratio (Insert : Plasmid)	Colonies / plate
1	Only plasmid	20
2	1:15	40
3	1:7	10^2
4	1:3	10^2
5	1:1	10^4
6	3:1	10^2
7	7:1	20
8	Only insert	-

Products of the ligation reaction 5 (Table 4.3) were introduced into *E.coli* once more and the appropriate dilution of transformed cells was plated on 200 LB-Amp plates. This resulted in an expression library of approximately 30000 colonies. This library was screened for activity towards EGDN and GTN as described in the next Chapter.

5 FUNCTIONAL SCREENING OF *NERA* MUTANTS

5.1 Enrichment

An attempt was made to develop an enrichment medium suitable for enrichment of *E. coli* heterologously expressing *nerA*. Because nitrite is an intermediate in the conversion of GTN to GDN, conditions that would allow *E. coli* to grow on nitrite were tested. The nitrogen free medium (Appendix B) that was commonly used for enrichment of environmental isolates was not suitable because *E. coli* requires anaerobic conditions in order to utilise nitrite as a source of nitrogen for growth. The enrichment method was optimised using nitrogen free medium as a starting point and resulting in a new medium called "anaerobic selection medium" (Appendix B). Even under anaerobic conditions *E. coli* did not readily utilise nitrite as a source of nitrogen for growth. The medium contained a trace amount of ammonia and an additional nitrogen source of interest (nitrite or nitro-ester). Without a co-substrate like ammonia or an amino acid *E. coli* would not utilise a substrate like nitrite or GTN. A method to maintain anaerobic conditions in this anaerobic selective medium was also developed, as is described in Section 2.2.5.

These conditions supported growth of *E. coli* on nitrite under anaerobic conditions. Swirling of the medium had a very positive effect on growth. If the medium was supplemented with nitrite (0.2 mM, in addition to the ammonia already present) the observed growth of an *E. coli* culture was considerably (3x) higher than on blank medium.

However, two observations, led to the conclusion that the enrichment strategy is not suitable for selection of *E. coli* host heterologously expressing *nerA*:

1. After addition (1 h) of GTN to incubations with cells heterologously expressing *nerA*, nitrite could be detected in the medium. The nitrite released by NerA was available for all cells in the incubation under these circumstances. The ability to release nitrite from GTN was therefore not a competitive advantage.
2. Negative controls (over-night cultures of *E. coli* carrying only empty plasmid, Topo ∅) had the ability to release nitrite from GTN (See Section 6.3.2). Expression of NerA gave therefore no competitive advantage.

5.2 Screening on solid medium

5.2.1 Development of screening method

Because selection by enrichment was not feasible, an alternative screening method was developed, whereby colonies were screened on a solidified medium. This section described various attempts to develop such a screening method.

1. Resistance to nitro-organic compounds.

Various experiments with LB-agar plates that contained gradients of nitro-organic substrate proved that NerA does not provide the host with a higher tolerance to any of these substrates.

2. Solidified medium that will detect nitrite colourimetrically.

LB-agar medium was supplemented with the azo dyes that form the active ingredients of the NO_2^- detection assay (sulphanilic acid, 6 mM; naphthylethylenediamine, 4 mM; Section 2.4.1 and Appendix B). The medium supported growth so it was concluded that *E. coli* tolerated these reagents. However, the detection of NO_2^- by the azo dyes was dependent on low pH. Although the LB-agar medium was slightly acidified in the immediate surroundings of a growing colony, the pH remained too high for NO_2^- detection. Titration of the pH of the medium to lower values inhibited growth. Other experiments indicated that the diffusion of NO_2^- through the solid medium was too high. After the incubation that was required to obtain colonies on a plate (16h) it would be impossible to distinguish which colony had produced the NO_2^-. The high diffusion of NO_2^- through the medium was also a problem in the enrichment strategy.

Because of the diffusion effect it was decided that full-grown colonies had to be exposed to the nitro-organic compound for a defined amount of time. Therefore substrate and subsequent reagents were applied to the surface of the LB-agar plate with help of a spray gun (Method described in Section 2.4.4). This idea was obtained from an article describing a method to detect nitro-organic compounds on Thin Liquid Chromatography plates (Yinon, 1993). Validation experiments showed that colonies expressing *nerA* could easily be distinguished from colonies that did not express *nerA*, because of a purple / red colour that was formed in the colony itself or in a halo around it (see also Figure 6.1).

5.2.2 Screening the expression library

White colonies

The expression library of mutants of *nerA*, described in the previous Chapter was screened as described in Section 2.4.4.

Every plate (200 in total) contained around 150 colonies. Less than half (40%) of the colonies were red but the majority of the colonies (60%) were white. The absence of red colour indicated that there was no NerA activity in these colonies. A PCR screen on a number of white colonies confirmed that the Topo *nerA* plasmid was present in these colonies. Loss of activity could not have been caused by promoter mutations since only the coding sequence of *nerA* had been mutated (the mutated *nerA* coding region was cloned back into a unmutated vector). It was assumed that the white phenotype was the result of a knockout mutation in the *nerA* gene.

Earlier it was observed that the mutagenic rate (λt) was around 1.7 and the Poisson distribution was used to predict how this would affect the distribution of mutations in the library (calculations are given later in Section 7.1.4). It was expected that the more mutations a certain mutant had (genotype), the more likely it was that the enzyme would have lost its function (phenotype). The expected distribution of phenotypes was compared to the results found in the screen (Figure 5.1).

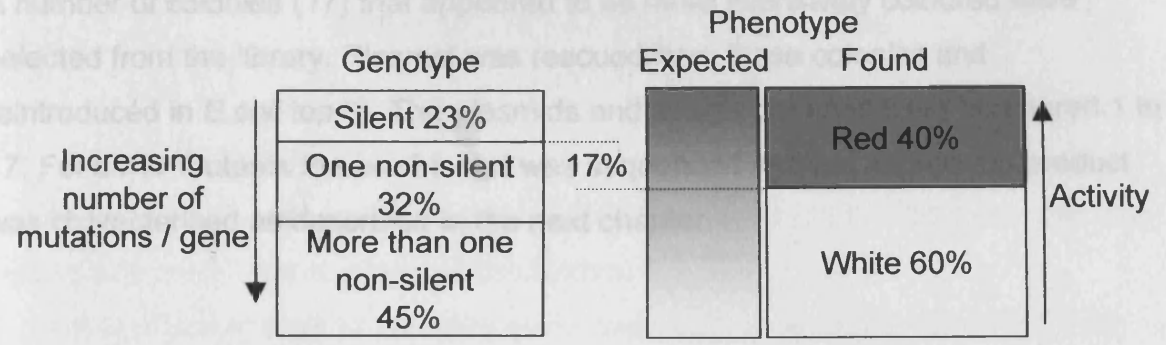


Figure 5.1 Population analyses

The composition of the library in terms of number of non-silent mutations per gene (using Poisson distribution and $\lambda t = 1.7$) is displayed on the left. It was expected that the enzyme is more likely to loose its function as the number of mutations in the gene increase, displayed as "Expected Phenotype" in the middle. The results that were found in the screen are displayed on the right. If all silent mutants are red, { 40% - 23% = 17% } red colonies remain, see text.

It was expected that the mutants with only silent mutations would be still functional, this would account for 23% of all colonies in the library (calculations in Section 7.1.4). The number of red colonies found was 40%. This leaves { 40% - 23% = 17% } red colonies. These 17% red colonies found their origin in the population that carries one non-silent mutation or more { 32% + 45% = 77% }. This implied that only { 17% / 77% = 22% } of the mutants are still functional and that the remaining 78% were knocked-out. It is unusual that a gene is so sensitive to mutations. Guo (1997) for example found that only 31% of non silent mutations led to loss of function in the protein that was studied. Guo also observed that between 30% and 50% of the knock-out events were caused by frame shift mutations. The explanation for the relatively high frequency of knockouts is therefore likely to be found in the high frequency of insertion and deletion events caused by mutagenic PCR. In early experiments (Section 4.1.1 "Increased concentration of magnesium") these frame shift events were found in 3 out of 5 knock-out library members.

Red colonies

A number of colonies (17) that appeared to be more intensively coloured were selected from the library. Plasmid was rescued from these colonies and reintroduced in *E.coli* top10. The plasmids and strains obtained were numbered 1 to 17. For all 17 mutants the *nerA* insert was sequenced and the expression product was characterised as described in the next chapter.

6 ENZYMATIC CHARACTERISATION OF ORGANIC NITRATE REDUCTASES

6.1 Environmental isolates

Enrichment is the most commonly used method to obtain microbial strains that can biodegrade nitro-organic compounds (Section 1.1.3).

A number of soil and water samples were taken from a freshwater pond that was used for several years as a soak-away for waste effluent from a gelignite factory. A number of axenic strains, capable of utilising either EGDN, GTN, PETN or TNT as a sole source of nitrogen for growth were obtained from these samples using the enrichment strategy described in Chapter 2. These strains, together with a number of randomly selected, unrelated environmental isolates that had not been exposed previously to nitro-organic compounds, were grown on LB. The resulting cell cultures were submitted to the whole cell assay for nitrite release from nitrate esters Section 2.4.1).

The strains that were obtained by enrichment on a nitro-ester could release nitrite from all of EGDN, GTN, PETN and TNT, even if they were isolated using only one of these substrates. The capacity to bio-convert nitrate esters and TNT was also found in *E .coli* (Section 6.3) and in the majority of the unrelated environmental (uncharacterised) isolates. It was concluded that the ability to bio-convert nitrate esters and TNT might be widespread in nature.

Eight isolates were replica-spotted on five LB plates, thus obtaining five identical libraries of eight spots each. The libraries were incubated over-night and each library was sprayed with a different substrate (EGDN, GTN, TNT, Tetrazolium or no substrate (blank)). The libraries were further processed as usual ("Library screen ", Section 2.4.4). Results are shown in Figure 6.1.

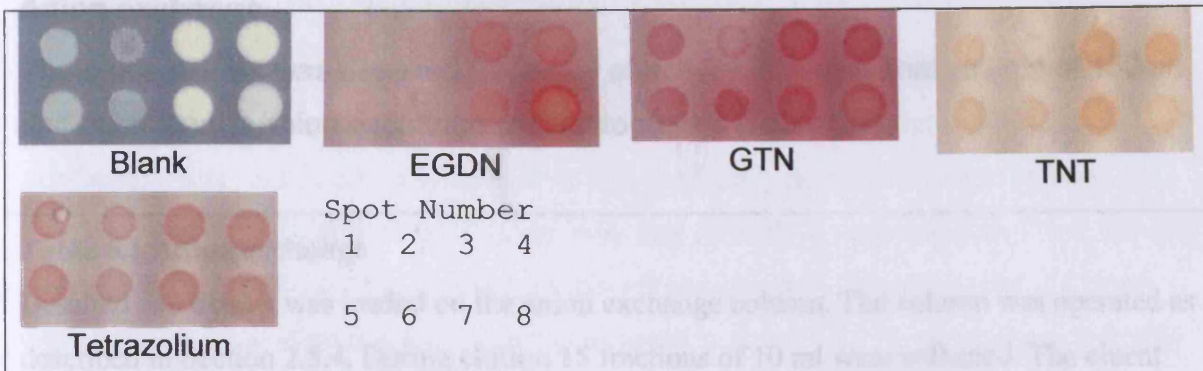


Figure 6.1. Validation of library screen

Eight isolates were replica-spotted, each on five LB-agar plates. After over-night incubation to allow grwoth, the plates were submitted to the library screen using the substrates indicated below each panel.

The eight cultures in Figure 6.1 had distinct activity signatures. For example spot 8 has high activity towards EGDN, moderate activity towards GTN but no activity towards TNT whereas spot 3 has moderate activity towards EGDN but good activity towards GTN and TNT. These environmental isolates were not further analysed, but served to demonstrate the usefulness of the method.

6.2 *Purification and Characterisation of* NerA

The protocol for the purification of NerA (Snape, 1997) was re-optimised. The optimised protocol can be found in Chapter 2. The purified wild-type NerA was used in validation studies of the high throughput characterisation assay (Section 6.4.1).

6.2.1 Purification of NerA

NerA protein was purified as described in Chapter 2 and the results are given here. A large scale culture of *E.coli* expressing *nerA* was harvested (10.5h after inoculation, yielding 6l cell culture attenuance at 600nm =1.5). The harvested cells were lysed by mechanical cell lysis, resulting in 120ml crude cell extract containing 3 g of protein and 24.4 U of GTN reductase activity.

Anion exchange

The crude extract was desalted (recovery of activity 85% in a final volume of 150ml) and submitted to anion exchange chromatography (Table 6.1).

Table 6.1 Anion exchange

Desalted cell extract was loaded on the anion exchange column. The column was operated as described in Section 2.5.4. During elution 15 fractions of 10 ml were collected. The eluent that eluted from the column during loading (Unbound, 150 ml) and washing (Wash, 250 ml) was also collected and analysed. ** Sometimes the oxygenase activity was so great that the activity towards GTN could not be measured.

Fraction	Total Volume (ml)	Total Protein (mg)	Total Activity (U)	Specific activity (mU / mg)
Load	150	2439	20.9	9
Unbound	150	206	**	N.A.
Wash	250	176	3.7	21
Fractions 1..15	150	638	12.9	20

The recovery of the anion exchange step was 62% { Activity recovered / activity in = 12.9 / 20.9 = 62% } and the purification factor was 2.2 { Specific activity recovered / specific activity in = 20 / 9 = 2.2 }. Remarkable was that the specific activity in the "Wash" (Table 6.1) was comparable to the specific activity in the combined fractions 1 to 15, implying that the "Wash" also represented purified NerA. However the "Wash" contained 2.5 U of activity towards TNT whereas the combined fractions 1 to 15 contained only 1.2 U of activity towards TNT. The activity towards TNT in the "Wash" must be allocated to reductases that originate from the host (Section 6.3.4) whereas the activity towards TNT in fractions 1 to 15 could be due to NerA activity. Although the recovery and the purification factor of the anion exchange step was poor, it effectively removed oxygenase and *E. coli* reductase activity from NerA activity. These two activities would otherwise compete with NerA for binding places on the affinity column.

Affinity column

During earlier optimisation studies it was concluded that a moderate concentration of salt (0.3M NaCl) could elute NerA from the affinity column. In the previous purification step NerA was eluted with a salt gradient and some of the "elution" fractions (number 12 and above) contained salt (0.3 M). It was expected that the salt would interfere with binding of NerA to the affinity column. Therefore the fractions with number 12 and above were set aside and fractions 5 to 11 were loaded on the affinity column. These fractions represented approximately 50% of the yield of the previous step and the remaining fractions (1 to 4 and 12 to 15), were saved for another experiment. Fractions 5 to 11 were loaded in reverse order so that the fraction with the highest salt concentration (fraction 11, 0.28M NaCl) was loaded first. In case the concentration of salt in that fraction was too high to allow NerA to bind, only the NerA in that fraction would remain unbound. If fraction 11 was loaded last, the salt in this fraction might elute all NerA from previous fractions. Furthermore it was expected that in a clean column high affinity binding places that could bind NerA even under moderate salt conditions would still be available. If fraction 11 was loaded last, these high affinity places would have already been occupied. The reverse loading procedure was preferred over desalting as the latter has an efficiency of only 85% (recovery of activity).

Results of the affinity purification step are shown in Figure 6.2.

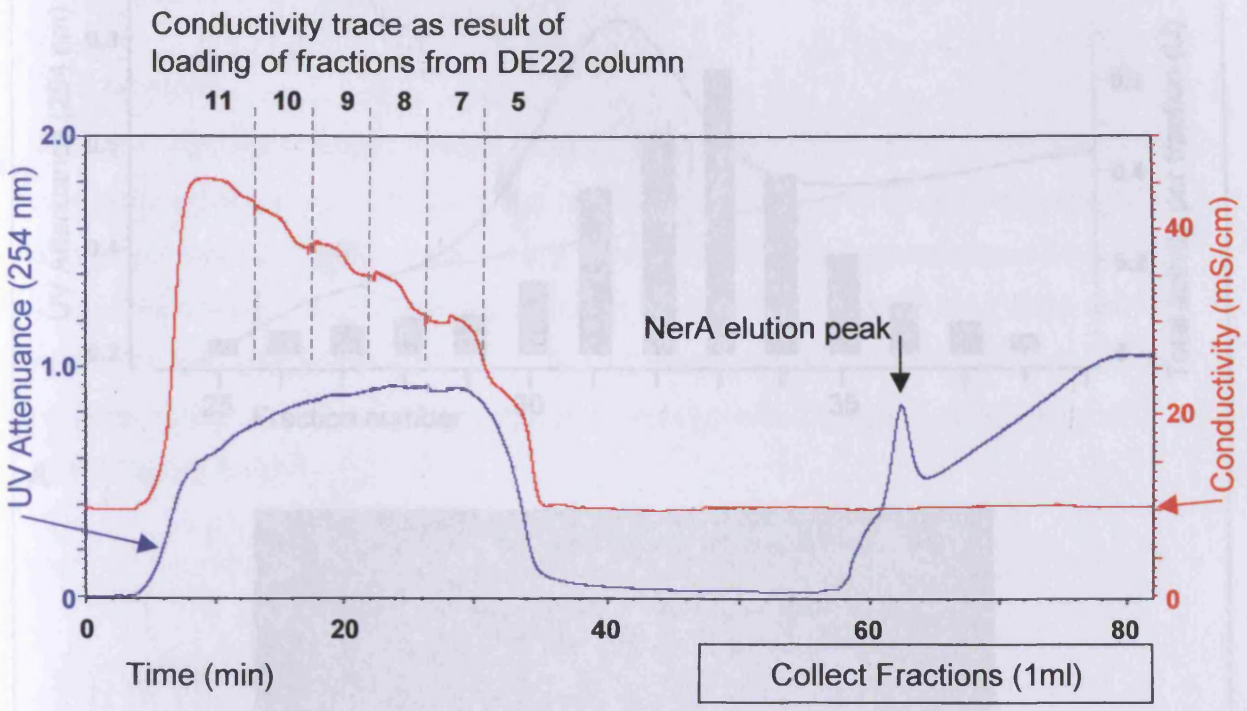
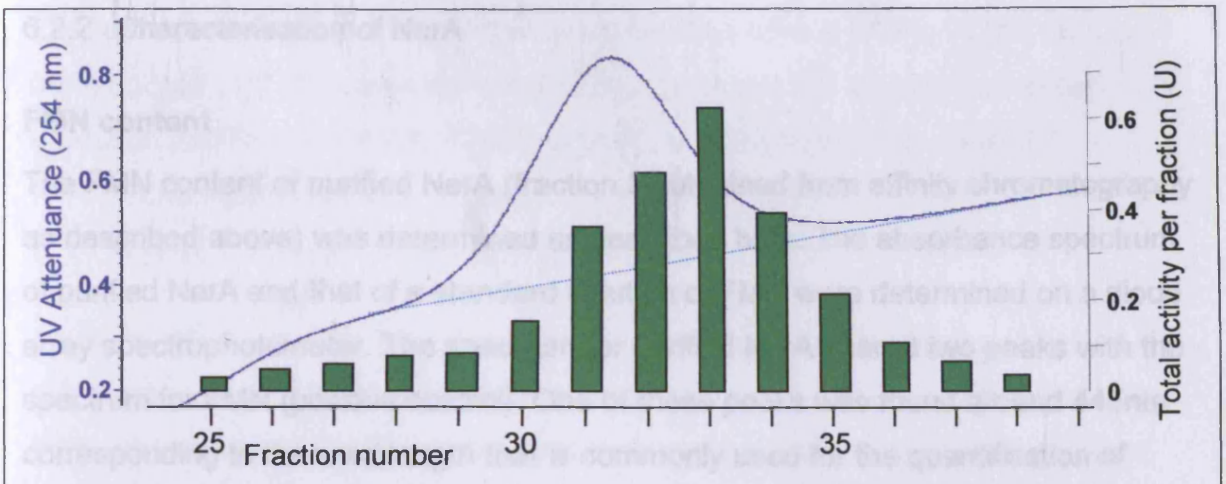


Figure 6.2 Affinity chromatography of NerA

The affinity column was operated as decribed in Section 2.5.5. Fractions 5 to 11, obtained from anion exchange chromatography were loaded in reverse order on an affinity column. The conductivity trace indicated the salt concentration in the loaded fractions. The column was washed and NerA was eluted in a NADH gradient.

The salt concentration in the loaded fractions was detected in the eluent during loading. The salt concentration decreased stepwise as the next fraction was loaded and therefore the conductivity trace had a step-like profile. The elevated UV trace detected the protein that did not bind to the affinity column and that eluted from the column during loading. The fractions around the elution peak were analysed in greater detail (Figure 6.3).



A

B

Figure 6.3 Elution peak of NerA

(A) Detail of taken from Figure 6.2. The inclining UV trace (dotted line) corresponds to the NADH gradient. (0 to 2mM). The peak on top of the inclining line is proposed to correspond to NerA being eluted around 0.88 mM NADH. Fractions were collected (1ml) and analysed for GTN-reductase activity (green bars). There was a time lag between the UV peak and the activity peak corresponding to the distance (2ml) between the UV detector and the fraction collector.

(B) SDS PAGE of fractions 27 to 36. The blue shade under the bands is a shadow.

NerA eluted from the affinity column in the NADH gradient (between 0.7mM and 0.98mM, estimated from UV trace Figure 6.3). The fractions that contained the bulk of the enzyme activity (fraction 30 to 36) contained 2.3U of enzyme activity, 19mg of protein in a volume of 7ml. A single band was visible on SDS PAGE.

6.2.2 Characterisation of NerA

FMN content

The FMN content of purified NerA (fraction 32 obtained from affinity chromatography as described above) was determined as described here. The absorbance spectrum of purified NerA and that of a standard solution of FMN were determined on a diode array spectrophotometer. The spectrum for purified NerA shared two peaks with the spectrum for FMN (positive control). One of these peaks was found around 445nm corresponding to the wavelength that is commonly used for the quantification of FMN (Figure 6.4).

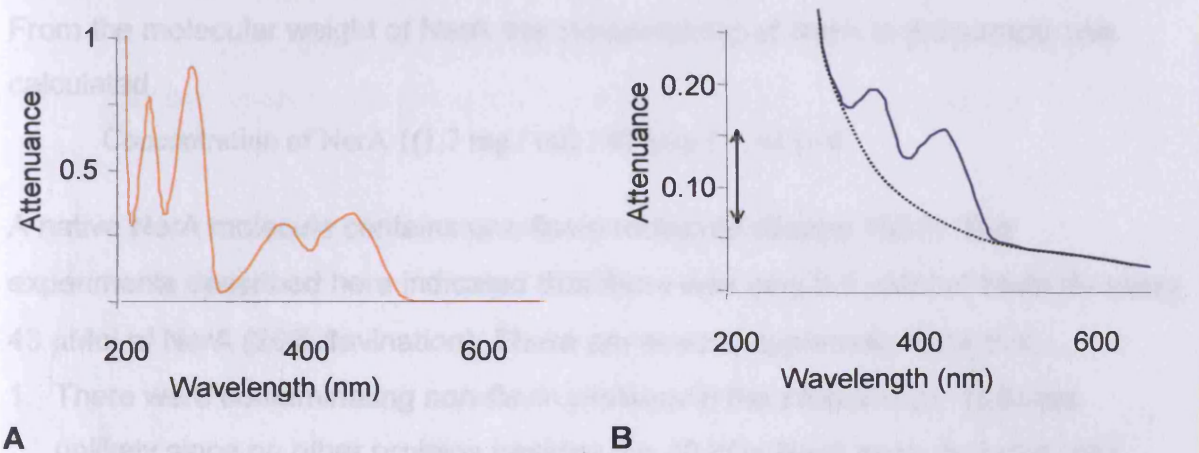


Figure 6.4 FMN content of purified NerA

UV/visible spectrum of (A) FMN (11µg / ml), (B) NerA (1.7 mg protein / ml). Peak heights in the NerA spectrum were measured relative to an extrapolated baseline (attenuance of protein) indicated by a dotted line. The concentration of NerA protein was measured using UV attenuance (280nm).

The attenuance at 445nm measured in the positive control (FMN, 11μg / ml, D_{450} = 0.35) agreed with the expected value (0.36), based on the specific molecular extinction coefficient of FMN (12500 M^{-1} cm^{-1} at 450 nm, Brown, 1998).

Concentration of FMN: { 11 mg / l } / { 370 g / mol } = 29 μM.

Expected attenuance {12500 $(M\ cm)^{-1}$ } x 29 μM = 0.36.

From the attenuance at 450nm measured for purified NerA (240mg protein / ml, D_{450} = 0.1) the concentration of FMN in that sample was estimated (assuming that the molecular extinction coefficient for FMN in the enzyme is the same as the values reported for Oye (Brown, 1998).

Concentration of FMN: {0.1 / 11700 $(M\ cm)^{-1}$ } = 8.5 μM.

From the molecular weight of NerA the concentration of NerA in the sample was calculated.

Concentration of NerA {(1.7 mg / ml) / 40 kDa } = 43 μM.

A native NerA molecule contains one flavin molecule (Snape 1997). The experiments described here indicated that there was only 8.5 μMol of flavin for every 43 μMol of NerA (20% flavination). There are several explanations for this:

1. There were contaminating non-flavin proteins in the preparation. This was unlikely since no other proteins besides the 40 kDa NerA were detected with SDS PAGE electrophoresis.
2. A part of the NerA in the preparation was de-flavinated (present as Apo enzyme). This was disproved, as addition of FMN did not increase reductase activity. (Reports indicate that inactive apo-protein readily takes up FMN and activity is restored (S. M. Marshall, personal communication).
3. The preparation contained deactivated NerA that has lost the ability to bind flavin and thus the ability to act as a reductase. It would be of interest to analyse the preparation with native gel electrophoresis so that the non-flavin binding enzyme can be distinguished from native enzyme.
4. Estimate of protein or FMN concentration was inaccurate. FMN concentration was measured at 445 nm. At this wavelength there was a high background attenuance of protein. This background could have interfered with the FMN measurements.

5. The extinction coefficient of FMN in NerA is lower than the reported value of FMN in Oye. The difference in extinction coefficient between free FMN and FMN bound in Oye is less than 10%. It is therefore unlikely that the difference in attenuance is significant. It would however be of interest to denature a known amount of purified NerA and detect the FMN that is liberated.

NerA analysed by High Throughput Characterisation

Fractions 33 and 34 (Figure 6.3) obtained by affinity chromatography were combined and submitted to the high throughput characterisation protocol (gel filtration followed by six substrate analyses as described in Section 2.4.3).
The first experiment was performed with a very high concentration of NerA (2ml of purified enzyme containing 8.3mg protein and 1 Unit of activity, Figure 6.5A). A second set of experiments was performed with a moderate concentration of NerA (25 mU in 2ml) whereby the influence of addition of extra FMN (1 mg) was studied (Figure 6.5B).

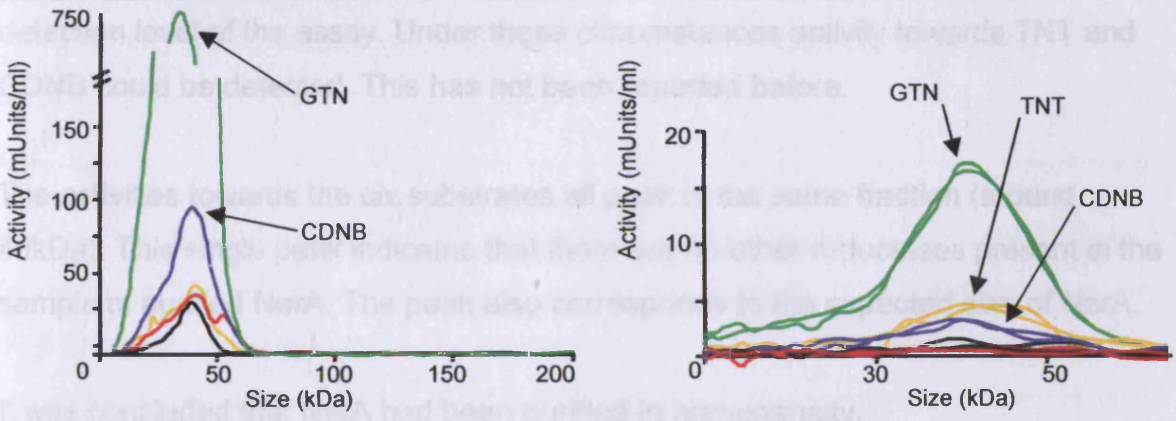


Figure 6.5 Characterisation of NerA

High throughput characterisation of (A) highly concentrated purified NerA (1U in 2ml), (B) purified NerA (25mU in 2ml). In an additional experiment purified NerA was supplemented with FMN (1 mg) before analysis. The results are superimposed on the diagram. Colours refer to the various substrates: GTN green, TNT yellow, DNB red, CDNB blue and Blank black.

In experiment (A) the reductase activity of NerA in the absence of substrate could be detected (autonomous activity in the blank = 25 mUnits / ml). It remains unclear by what mechanism NerA can oxidise NADH without having a substrate to reduce. It could be possible that oxygen was a substrate so that NerA acted as an oxygenase. Of the substrates tested with purified NerA, GTN was the preferred substrate. Activity towards CDNB could also be detected. Addition of DNB or TNT did not increase the NADH oxidation rate under the circumstances tested.
Addition of FMN to the diluted sample of purified NerA did not affect the NerA activity of that sample, indicating that there was no apo protein present in the sample (Figure 6.5B).

Comparing Figure 6.5A and B and noting the difference in scaling of X and Y axis, the activity towards GTN was 40 fold lower as a result of loading 40 times less enzyme. This agreed with the assumption that NerA activity towards GTN would be proportional to the concentration of enzyme. At lower enzyme concentration the autonomous activity (no added substrate) was of the same magnitude as the detection level of the assay. Under these circumstances activity towards TNT and CDNB could be detected. This has not been reported before.

The activities towards the six substrates all peak in the same fraction (around 40kDa). This single peak indicates that there are no other reductases present in the sample of purified NerA. The peak also corresponds to the expected size of NerA.

It was concluded that NerA had been purified to homogeneity.

6.3 E.coli Reductases

There have been reports on enzymes native to *E.coli* that have activity towards nitro-organic compounds. *E.coli* NfsA and NfsB, for example have activity towards nitro-toluene and nitro-phenol (Zenno 1996 a; Zenno 1996 b). NfsB also has activity towards TNT and a sample of NfsB, isolated from *E.coli* K12 was donated by Andrew Lovering (Lovering 2001). Whilst purifying NerA from cell-free extract it also became apparent that *E.coli* has several enzymes that have activity towards a number of nitro-organic compounds including EGDN, GTN and TNT.

E.coli cell extract (obtained from *E.coli* Top10, a derivative of the K12 strain) as well as the purified NfsB were submitted to the high throughput characterisation protocol.

6.3.1 NfsB

Purified NfsB (100 μg in 2ml, Lovering, 2001, Section 2.1.3) was analysed by the high throughput characterisation protocol. Results are shown in Figure 6.6.

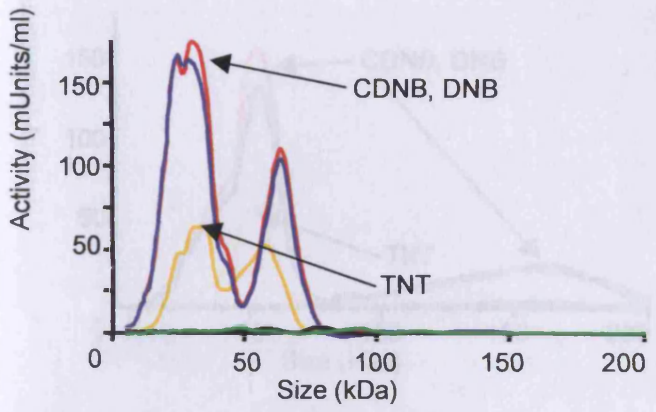


Figure 6.6 High throughput characterisation of NfsB

Purified NfsB (100μg in 2ml) was analysed by gel filtration followed by six substrate analyses.

NfsB has a molecular weight of 24 kDa and forms homodimers. The sizes that were found by gelfiltration were 30 kDa and 60 kDa respectively. These two values could be explained by the existence of a monomer and a dimer, but they are greater than the expected 24 kDa and 48 kDa. It is also surprising that the monomer and dimer were detected separately. If the dimerisation / monomerisation process was taking place in the column, it was expected that the enzyme would elute as a combined peak.

NfsB has a distinct substrate specificity. Its preferred substrates are CDNB and DNB, it also has some activity towards TNT. Activity towards EGDN, GTN or PETN was not detected.

6.3.2 *E.coli* cell-free extract

An *E.coli* Top10 culture harbouring the Topo ∅ plasmid was submitted to the high throughput characterisation protocol. This experiment served as a negative control for future experiments with *E.coli* cultures expressing *nerA*.

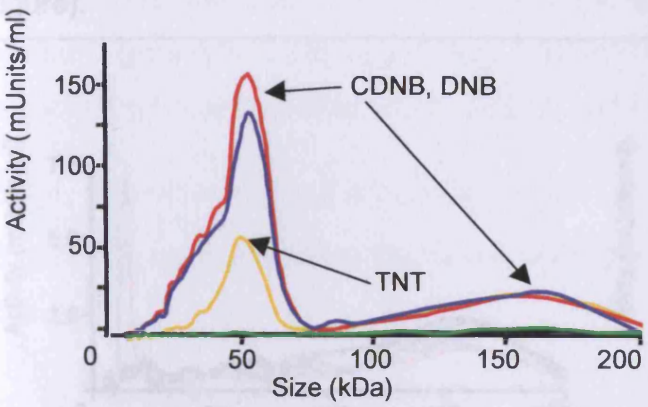


Figure 6.7 Negative control in high-throughput characterisation

An *E.coli* Top10 culture harbouring the Topo ∅ plasmid was submitted to the high throughput characterisation protocol.

A 50 kDa peak with a substrate specificity very similar to that of NfsB was found. The size corresponds very well with the expected size of the NfsB dimer (48 kDa). It would be of interest to repeat this experiment with an E.coli strain that is *nfsB* negative, to confirm that the 50 kDa peak found here is indeed caused by NfsB. There is an additional peak around 160 kDa. It can be distinguished from the 50 kDa peak not only by size but also by substrate specificity. The 160 kDa peak has no activity towards TNT but some activity towards GTN was detected. Because the substrate specificity of the 160 kDa peak is so different from that of NfsB or the 50 kDa peak, it was assumed that the 160 kDa peak represented an additional *E.coli* nitro-reductase. Hereafter it is referred to as 160 kDa *E.coli* nitro-reductase. The quaternary structure of 160 kDa *E.coli* nitro-reductase remained unclear. If the enzyme later proves to be a multimer, the name "160 kDa" might be inappropriate.

6.3.3 *E.coli* EGDN reductase

The high-throughput characterisation protocol was usually used to analyse cultures that have been incubated for 3 to 4 h. In an additional experiment the culture that served as a negative control as described above was incubated for an additional 24 h and analysed. The results of this experiment were compared to those obtained from the negative control (Figure 6.8; note that every substrate has a separate figure).

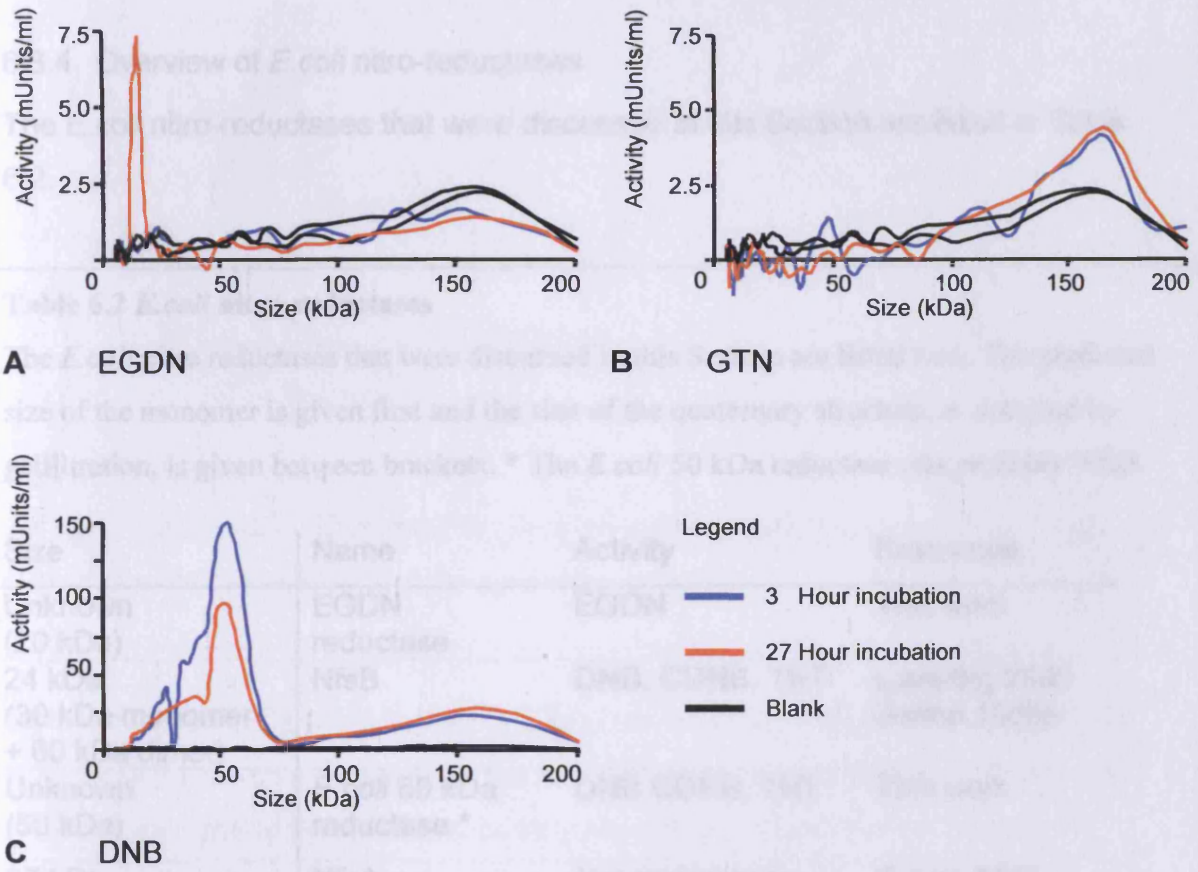


Figure 6.8 Effect of growth phase on *E.coli* reductases

An *E.coli* Top10 culture harbouring the Topo ∅ plasmid was submitted to the high throughput characterisation protocol (indicated in blue). The same strain was incubated for 27 h rather than the usual 3 h and also analysed (indicated in orange). The results obtained for the substrates CDNB and TNT are comparable to those obtained for DNB (C).

The 50 kDa and 160 kDa peaks are still found after additional incubation. There is less activity associated with the 50 kDa peak whereas the 160 kDa peak remained unaffected. This is additional proof for the assumption that the 50 kDa and 160 kDa peaks refer to two independent nitro reductases. In Figure 6.8B the activity of the 160 kDa peak towards GTN becomes visible. The most striking result however is the finding of a high activity of very small (± 20 kDa), late-induced EGDN reductase that has no detectable activity towards any of the other nitro organic compounds tested (Figure 6.8A). The quaternary structure as well as the identity of the *E.coli* enzyme that is responsible for the EGDN reductase activity, remains unclear.

6.3.4 Overview of *E.coli* nitro-reductases

The E.coli nitro-reductases that were discussed in this Section are listed in Table 6.2.

Table 6.2 *E.coli* nitro-reductases

The *E.coli* nitro reductases that were discussed in this Section are listed here. The predicted size of the monomer is given first and the size of the quaternary structure, as detected by gelfiltration, is given between brackets. * The *E.coli* 50 kDa reductase was probably NfsB.

Size	Name	Activity	Reference
Unknown (20 kDa)	EGDN reductase	EGDN	This work
24 kDa (30 kDa monomer + 60 kDa dimer)	NfsB	DNB, CDNB, TNT	Lovering 2001 Zenno 1996b
Unknown (50 kDa)	*E.coli* 50 kDa reductase *	DNB CDNB, TNT	This work
27 kDa (Not analysed)	NfsA	Not tested here	Zenno 1996a
Unknown (160 kDa)	*E.coli* 160 kDa reductase	DNB, CDNB, GTN	This work

6.4 *Mutants of* NerA

6.4.1 Validation High throughput characterisation

An *E.coli* Top10 culture harbouring the Topo *nerA* plasmid (wild type *nerA* insert) was submitted to the high-throughput characterisation protocol. This experiment served as a positive control for future experiments with *E.coli* cultures expressing mutant *nerA*. Results are shown in Figure 6.9.

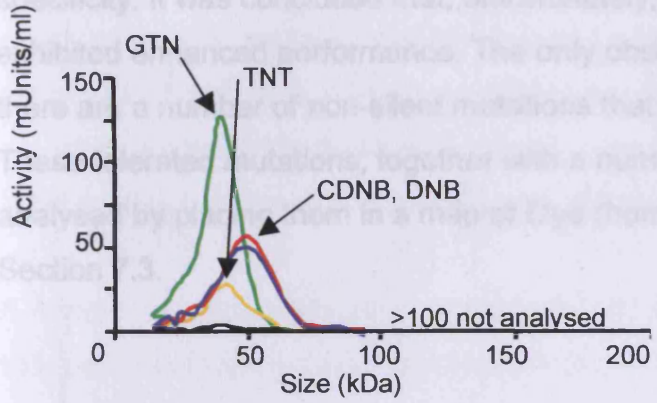


Figure 6.9 NerA in *E.coli*

An *E.coli* Top10 culture harbouring the Topo *nerA* plasmid (wild type *nerA* insert) was submitted to the high-throughput characterisation protocol.

As can be seen in Figure 6.9, the 50 kDa *E.coli* reductase could be distinguished from NerA based on difference in size and difference in substrate specificity. NerA could be recognised by its unique activity towards GTN and the 50 kDa *E.coli* reductase could be recognised by its unique activity towards DNB. Both NerA and the 50 kDa reductase have activity towards TNT and this resulted in a combined peak between the NerA peak and the 50 kDa peak. This limited the sensitivity to TNT activity by NerA but detection of enhanced activity towards TNT was not the main focus of this thesis. The previous screening step (screen of expression library) focused on enhanched activity towards EGDN and GTN. The high-throughput characterisation protocol was suitable to detect mutants of NerA with enhanced activity towards EGDN or GTN because there was no overlap between NerA and *E.coli* reductases that had activity towards EGDN or GTN. The 160 kDa reductase

has some activity towards GTN but it was effectively purified away from NerA by gel filtration. The EGDN reductase was late expressed and was never found in the incubations that were used to analyse mutants of NerA.

6.4.2 Mutants of NerA analysed by high throughput characterisation

The strains that were obtained from the screening of the expression library (numbers 1 to 17) were submitted to the high throughput screening method as usual. None of the library members showed significant change in substrate specificity. It was concluded that, unfortunately, none of the apparent "winners" exhibited enhanced performance. The only observation that could be made was that there are a number of non-silent mutations that did not negatively affect activity. These tolerated mutations, together with a number of knockout mutations were analysed by placing them in a map of Oye (homologue of NerA) as described in Section 7.3.

7 DISCUSSION

7.1 Statistical parameters governing mutagenesis and screening

In this Sub-Chapter the findings of Chapter 3 will be applied to the results found in this thesis (Chapter 4). There will also be a comparison with work by other authors.

7.1.1 Mutagenic classes

In Chapter 3 all mutagenic events were divided over 6 mutagenic classes. It was observed that the majority of the amino acid exchanges that can be achieved using single mutagenic methods can be caused by one mutagenic class only. In order to reach the full potential of mutagenic PCR it is important that all mutagenic classes are represented in the mutagenic offspring (see Section 3.2.5). In this section the representation of the mutagenic classes will be discussed.

A number of articles describing mutagenic PCR were analysed (Table 7.1) so that the results in these articles could be compared to the results obtained in this thesis. Comparison of the data in Table 7.1 was made possible by standardising the various results. The left half of the table indicates some of the components of the PCR mixture, the right half of the table displays some of the statistical parameters of the mutagenic offspring.

The over-representation of a class (column 5) was calculated according to Equation 7.1

Equation 7.1 Over representation of most abundant class

The over representation of the most abundant class was calculated by dividing the occurrence of that class by the average occurrence of the other classes. The average occurrence of the other classes was calculated by taking the sum, Σ of all other classes and dividing it by the number of other classes (5).

$$Overrepresentation(class) = \frac{Occurrence\ class}{\sum occurrences\ other\ classes / 5}$$

The 2HB Bias (column 6), indicating the preference towards incorporating an A or T (2 Hydrogen bonds in base pair bond) was calculated according to Equation 7.2.

Equation 7.2 Two Hydrogen Bond Bias

(A) The 2HB Bias (Two Hydrogen Bond Bias) was calculated by dividing the number of times an A or T was incorporated by the number of times a C or G was incorporated. Class 2,3 and 6 all involve the incorporation of an A or T so the second formula can also be used.

(B) The 2H bond bias is also often indicated by this notation

(A) $$2H\ Bond\ Bias = \frac{Incorporation\ of\ A\ or\ T}{Incorporation \cdot f\ C\ or\ G} = \frac{\sum class\ 2,3,6}{\sum class\ 1,4,5}$$

(B) $$2H\ Bond\ Bias = \frac{\{A+T\}}{\{C+G\}}$$

It must be noted that some of the authors reported the amino acid exchanges that they found without reporting the underlying mutagenic events, or silent mutagenic events that might have occurred. In these cases the mutagenic events that were reported were "reverse translated" as is explained in the next section.

Table 7.1 Some parameters of mutagenic PCR

The conditions in the PCR mixture used by various authors is given in column 2-4. The last two columns display the result of the analyses of the offspring of the PCR reaction. Concentrations of Mg^{2+} and Mn^{2+} are listed in the columns 2 and 3. The concentrations of nucleotides are listed in column 4. Unless indicated otherwise the four nucleotides were added in the same concentration (abbreviations: A = dATP etc.). Column 5 describes the most common class of mutagenic event and its over-representation compared to other classes (Equation 7.1, in some cases estimated, see next section). The last column displays the 2H Bond Bias (Equation 7.2).

Reference	Mg^{2+} (mM)	Mn^{2+} (mM)	Nucleotides (μM)	Most common class	2H Bond Bias
This work Section 4.1.1	7	0.1	200	Class 2 5 fold	3.2
Cadwell (1992)	7	0.5	200 A, G 1000 C, T	Class 3 < 2 fold	1
Guo (1997)	6	0.64	200	Class 2 (RT) 2.3 fold	1.7
Lin-Goerke (1997)	2	0.25	20	Class 1 7.5 fold	0.43
Lönn (2002)	1.5	0.5	200 A, G 1000 C, T	Class 1 (RT) > 3.3 fold	n.a.
Riegert (2001)	4.7	0.05	3500 C 100 A, G, T	Class 1 (RT) ± 5 fold	0.2 – 0.5
Shafikhani (1997)	7	0 – 0.5	200 A, G 1000 C, T	Class 3* 3 fold*	1.6*
Svetlov % (1998) II	0	0.5	125	Class 1 8.8 fold	0.4
III	1.5			Class 2 4 fold	1
Vartanian (1996)	2.5 – 5	0.5	5-100 A 1-100 C 50 – 1000 G 1000 T	Class 2 "vast majority"	0.3 – 60
Xu (1999)	2	0.04	200 ++	Class 1 2.5 fold	1

* Average results over three experiments using 0, 0.15 or 0.5 mM of Mn^{2+}.

% Svetlov tested a PCR mixture with or without Mg^{2+} (experiments III and II respectively).
++ Xu used a second mutagenic step involving dITP.
(RT) Statistical parameters obtained by "Reversed Translation of amino acid exchanges", see text.

7.1.2 Reverse translation of amino acid exchanges

Three authors (Lönn, 2002; Riegert, 2001 and Guo, 1997) in Table 8.1 failed to report the mutagenic events that caused the amino acid exchanges that they analysed. Table 3.2 was used to look up what mutagenic events might have caused the amino acid exchanges. The mutagenic events were tabulated and three statistical parameters were calculated with help of this table:

1. The most abundant mutagenic class.
2. The relative over-representation of that class (Equation 7.1).
3. The 2HB Bias (Equation 7.2).

Lönn reports three amino acid exchanges. Table 3.2 was used to look up what mutagenic classes could have caused these exchanges (Table 7.2. This look up procedure is referred to as a "reverse translation of an amino acid exchange" in this thesis).

Table 7.2 Amino acid exchanges reported by Lönn

The amino acid exchanges reported by Lönn are placed in the first column. Table 3.2 was used to look up what mutagenic classes could have caused these exchanges (column 2).

Exchange	Class
F 163 L	1 or any transversion (V)
E 372 G	1
V 379 A	1

Although the statistical population is very small (N=3) it is obvious (Table 7.2) that in the mutagenic offspring generated by Lönn, class 1 was over represented. The article mentioned that 80% of the mutagenic events were transitions. Class 1 and class 2 are transitions, so that {class 1 + class 2} = 80%. Since class 2 events did not appear in Table 7.2 it was assumed that class 1 was more abundant that class 2 { class 1 > class 2 }. Taking these two conditions into account class 1 must be between 40 and 80% (and the other classes between 60 and 20% respectively). Using Equation 7.1 the over representation of class 1 is between { 40% / (60% / 5) = 3.3 and 80% / (20% / 5) = 20 } = at least 3.3 fold (but no more that 20 fold).

Riegert reports six amino acid exchanges (Table 7.3).

Table 7.3 Amino acid exchanges reported by Riegert

For explanation see legend Table 7.2.

Exchange	Class
K 22 E	1
M 37 T	1
F 66 L	1 or any transversion (V)
N 96 T	5
K 106 R	1
D 155 Y	6

It is again obvious that class 1 is over represented (Table 7.3). Making the moderate assumption that the exchange F 66 L was caused by an unknown transversion, rather than by a class 1 event, at least three out of the six mutagenic events involved class 1. The over representation of class 1 is estimated to be { 3 / ((6 – 3) / 5) } = 5 fold

Only the exchange D 155 Y and perhaps exchange F 66 L, involved incorporation of an A or T. The incorporation ratio was therefore estimated at {(1 to 2) / (4 to 5)} = 0.2 to 0.5

Guo reports around 250 non-silent events resulting in 200 amino acid exchanges (some exchanges were found more than once). Guo examined a population of approximately 100 clones. A table similar to Table 7.2 and Table 7.3 was made for the first 171 non-silent events (not all 250 events were looked up because 171 seemed to be a representative population. Results not displayed here). The occurrence of all classes was counted (Table 7.4).

Table 7.4 The occurrence of mutagenic classes in the offspring described by Guo

The first 171 non-silent events (those events displayed in the first column of FIG4 in the article by Guo (1997)) were looked up in table 3.2 as described earlier in this section. The occurrence of each class was counted (column 2). The sum of the occurrences was slightly greater than 171 because some exchanges can be caused by more than one class.

Class	Occurrence
1	34
2	58
3	36
4	20
5	16
6	9
Any Transition (V)	11
Sum	184

Class 2 is the most abundant in the offspring described by Guo (Table 7.4). Class 2 was over represented { 58 / ((184 – 58) / 5) } = 2.3 fold.

7.1.3 Comparison of results

PCR was used to introduce random mutations in the *nerA* gene. Initial attempts depended on an increased Mg^{2+} (2.5 mM) concentration. In this initial attempt there was a strong bias towards transitions in the offspring. Class 2 mutagenic events were over represented 15-fold. A bias towards transitions was expected (see Section 3.2.1). Because these results were unsatisfactory the PCR protocol was further developed. The final mutagenic PCR protocol depended on the presence of Mn^{2+} (0.1mM) and a high concentration of Mg^{2+} (7mM). In this offspring all mutagenic classes were represented although class 2 events were still over-represented 5-fold (Section 4.1.1 and Table 7.1, line 1).

As can be seen in Table 7.1 in most cases the PCR protocol generated an offspring with an abundance of transitions (class 1 and class 2). Only Cadwell and Shafikhani seem to have found a method that is not biased in favour of transitions. In these cases class 3 was over represented.

In the method reported by Cadwell the over representation of class 3 was less than 2. This implies that the overall bias is low. This is also reflected in the 2HB Bias. The 2HB Bias was 1, indicating that there was no preference for incorporation of any particular nucleotide.

The PCR protocols that were used by Cadwell and Shafikhani have a number of conditions in common: high Mg^{2+} (7mM), presence of Mn^{2+} (0.5 mM) and biased concentrations of nucleotides (dATP and dGTP 200 mM, dCTP and dTTP 1000 mM). It would seem obvious that the PCR protocol used in this thesis would benefit from increased concentration of Mn^{2+} and biased nucleotide concentrations. However under the conditions tested in this thesis, Mn^{2+} was found to have a strong inhibitory effect on the PCR reaction. The protocols of Lönn, Svetlov and Vartanian yielded an offspring that has a strong bias in favour of transitions in spite of the increased Mn^{2+} concentration. It must be noted however that these authors used a Mg^{2+} concentration that was much lower than the 7 mM used by Cadwell and Shafikhani.

As can be seen in Table 7.1 in most cases the PCR protocol generates an offspring with a 2HB Bias. It remains unclear what exactly influences the 2HB Bias. Vartanian used extreme nucleotide ratios in the PCR mixture and reported that high concentrations of dTTP in combination with low concentrations of dCTP can increase the preference for incorporation of A or T. Work by Riegert indicated that the opposite is also true. Riegert used high concentrations of dCTP and found preference for incorporation of C or G. It must therefore be possible to optimise the PCR protocol used in this thesis by adjusting nucleotide concentrations. This remains a major challenge and it must be noted that the excellent results reported by Cadwell were obtained after several iterative optimisation steps. A protocol that works well in the hands of Cadwell might not perform so well in the hands of others. Shafikhani used Cadwell's method and found a biased offspring in spite of using the same conditions.

7.1.4 Mutagenic rate

In Chapter 3 it was predicted that the number of mutations per gene would follow the Poisson distribution. It was concluded that single mutants are accompanied by a background of native and over-mutated sequences. The fraction of the population that carries exactly one non-silent mutation is a function of the mutagenic rate. Various authors use different definitions for "mutagenic rate". To allow comparison of the various results Equation 7.3 was used to calculate mutagenic rate per base pair (λ) or per gene (λt)

Equation 7.3 Mutagenic rate

Mutagenic rate was expressed as average number of mutations per base pair (λ) or as average number of mutations per gene (λt). The symbols λ and t find their origin in the Poisson distribution (Section 3.3.1).

$$\lambda = \frac{\sum Mutations}{\sum Base\ \ pair\ \ sequenced}$$

$$\lambda t = \lambda \times gene\ \ length(Base\ \ pair)$$

In Table 3.3 it was explained that the various mutagenic classes have different fractions of non-mutagenic events. When analysing the relationship between mutagenic rate at nucleotide level and the mutagenic rate at amino acid level, this should be taken into account (Equation 7.4A).
However in many cases the λt of any particular class has a large margin of error, or the participation of the class in the overall λt is unknown. In these cases the difference in the fractions in Equation 7.4A becomes irrelevant and Equation 7.4B is more appropriate.

Equation 7.4 Mutagenic rate at amino acid level

A. The average number of non-silent events per gene ($\lambda t_{\text{Non Silent}}$) can be calculated by weighting the various mutagenic classes. B. In some cases the λt of the various mutagenic classes is unknown so an average weighting factor is used.

A $\lambda t_{\text{Non Silent}} = 0.60 \times \lambda t_{class1} + 0.60 \times \lambda t_{class2} + 0.67 \times \lambda t_{class3} + 0.79 \times \lambda t_{class4} + 0.71 \times \lambda t_{class5} + 0.71 \times \lambda t_{class6}$

B $\lambda t_{\text{Non Silent}} = 0.68^* \times \lambda t$

* The average of the six weighting factors.

Some authors define the mutagenic rate as the fraction of clones in a library that contain one or more non-silent mutations (Lin Goerke, 1997 Equation 7.5A). This definition fails to distinguish between clones that contain exactly one mutation and those that contain several mutations. It also fails to appreciate the difference between mutagenic rate at nucleotide level and the mutagenic rate at amino acid level. Another disadvantage is that silent mutations are not taken into account. Silent mutations carry information about the properties of the mutagenic offspring. This information is lost when silent mutations are not taken into account. Rather than using Equation 7.5A it would be more appropriate to use the Poisson distribution to find the number of clones that contain one or more mutations (Equation 7.5B).

Equation 7.5 Fraction of library carrying non silent mutations

A. Some authors define the mutagenic rate (λt) as the fraction of the clones in a library that carry one or more non-silent mutations. This formula does not agree with the definition of λt in the Poisson distribution. B. The Poisson distribution can be used to find the fraction of the library that carries a non silent mutations.

A $$\lambda t = Clones\ with\ non\ silent\ mutation / Library\ Size$$

Although definition (A) is often used by other authors it is not used in this thesis because the definition of λt in this thesis does not agree with this formula

B $$Clones\ with\ non\ silent\ mutation / Library\ Size = 1 - e^{-0.68 \times \lambda t}$$

The weighting factor 0.68 in the exponent of equation Equation 7.5B. corresponds to the weighting factor used in Equation 7.4B.

The work of various authors was analysed and the mutagenic rate per base pair (λ) or per gene (λt) was calculated using Equation 7.3. Not all authors that are listed in Table 7.1 gave sufficient data to calculate the mutagenic rate.

Table 7.5 Mutagenic rate

The work of various authors was analysed and the mutagenic rate per base pair (λ) and per gene (λt) was calculated using Equation 7.3. The mutagenic rate at amino acid level ($\lambda t_{Non\ Silent}$) was calculated using Equation 7.4. Not all authors use the same definition of mutagenic rate so depending on the source of information the calculations started with λ, λt or $\lambda t_{Non\ Silent}$ (See "Calculation of values" below).

Reference	λ ($^0/_{00}$)	λt	$\lambda t_{Non\ Silent}$
This work	1.4	1.6	1
Cadwell (1992)	6.6	2.3	0.48
Guo (1997)	12 – 28	3.1 – 7.2	2.1 – 4.9
Lin-Goerke (1997)	4.8	0.98	0.67
Shafikhani (1997)	6.6 – 29	6.7 – 29	4.6 - 20
Svetlov [%] (1998) II	11	1.2	0.82
III	5.4	0.6	0.41
Vartanian (1996)	0.72 – 100	0.17 - 23	0.1 – 16
Xu (1999)	1.3	1.2	0.82

Calculation of values:

This work see Section 4.1.1.

Cadwell reported a mutagenic rate of λ = 0.66% (16591 nucleotides were sequenced). This value was placed in "λ" column. The length of the gene was 352bp so λt was { 0.66% x 352 = 2.3 }. This value was multiplied by 0.68 to find $\lambda t_{Non\ Silent}$.

Guo reported between 2.1 and 4.9 non-silent mutations per gene (258 bp). These values were placed in the $\lambda t_{Non\ Silent}$ column and were divided by 0.68 to find λt. The values for λt were divided by the length of the gene (258 bp) to find λ.

Lin-Goerke found 278 base pair substitutions in 283 colonies that contained an insert. The mutagenic rate per gene (λt) was calculated { 278 / 283 = 0.98 } and placed in colum "λt". λt was divided by the gene length (300bp) to find λ. λt was multiplied by 0.68 to find $\lambda t_{Non\ Silent}$.

Shafikhani compared three different protocols. The results were: 30 mutations per 4518 base pair, 40 mutations per 3514 base pair and 286 mutations per 9800 base pair. The mutagenic rate per base pair (λ) was calculated to be 6.6 $^0/_{00}$, 11 $^0/_{00}$ and 29 $^0/_{00}$. These values were placed in the "λ" column. The other two values were calculated as usual (gene length = 1017bp).

Svetlov compared three different protocols of which two were mutagenic. The results obtained from 20 clones were presented as number of wild-type, single mutants double mutants etc. From these values the total number of mutations was calculated to be 11 and 24. These values were divided by the number of clones (20) to find the mutagenic rate per gene (λt): 0.55 and 1.2 respectively. These values were placed in the "λt" column. The other values were calculated as usual (gene length 102 bp).
Vartanian reported the mutagenic rate per base pair (λ) to vary between 0.72 $^0/_{00}$ and 100 $^0/_{00}$. The other values were calculated as usual (gene length 231 bp).
Xu found 12 mutations in 10 colonies. The mutagenic rate per gene (λt) was calculated to be 1.2. The other values were calculated as usual (gene length 900 bp).

End of Table 7.5

Guo, Shafikhani, Svetlov and Vartanian report very high mutagenic rates per nucleotide ($\lambda > 10$ $^0/_{00}$). An explanation for this was sought in the mutagenic protocol. Shafikhani re-amplified the PCR product up to six times, resulting in up to 78 PCR cycles, increasing the opportunities for mutation to occur. Vartanian reported mutagenic rates of up to $\lambda = 100$ $^0/_{00}$ using extreme nucleotide ratios and concentrations. It remains unclear why Guo and Svetlov found high mutagenic rates as their PCR protocols were very similar to the protocols used by authors like Cadwell and Lin-Goerke respectively.

The mutagenic rate of the protocol described in this thesis was close to optimal (1.7, Section 3.3.1). This results in a library of mutants that contain only silent mutations (32%), exactly one non silent mutation (23%) or more than one non-silent mutation (remainder = 44%). The protocols by Cadwell, Lin-Goerke, Svetlov and Xu had a mutagenic rate (λt) that was below the optimal mutagenic rate. Svetlov worked with short (102bp) DNA fragments so in order to obtain a reasonable mutagenic rate at gene level (λt) it was necessary to use a high mutagenic rate per nucleotide ($\lambda = 11$ $^0/_{00}$).

Guo and Shafikhani used the mutagenic method to investigate the tolerance of a protein to amino acid exchanges. They were not interested in single non-silent mutations. Therefore they used a very high mutagenic rate.

Xu reported 12 mutations in 10 clones. One of the clones was wild-type and therefore Xu calculated the mutagenic rate to be { 12 per 9 *mutants* = 1.3 }. However it is incorrect to ignore clones that are wild-type as they are an integral part of a population that was generated by a random process.

7.1.5 Distribution of mutations

Cadwell showed that the fraction of wild type / single / double / triple mutants followed the Poisson distribution (Figure 2 in the article by Cadwell, 1997).

In Section 4.1.1 it was described that in the present work, 24 mutations were found in 15 clones; 5 clones were wild-type, 4 contained exactly one mutation, 2 contained exactly 2 mutations and 4 clones contained 3 mutations or more. These results have been placed in Figure 7.1. To normalise these results, the relative frequencies were calculated by dividing the observed number by the total number of mutations (24). The expected frequencies were calculated using the Poisson distribution and the λt that was calculated in Table 7.5. In Figure 7.1 the observed and the expected results are compared.

Mutations / gene	Observed	Observed Frequency	Expected Frequency $\lambda t = 1.6$
0	5	33%	20%
1	4	27%	32%
2	2	13%	26%
3	1	7%	14%
4	2	13%	6%
5	1	7%	2%

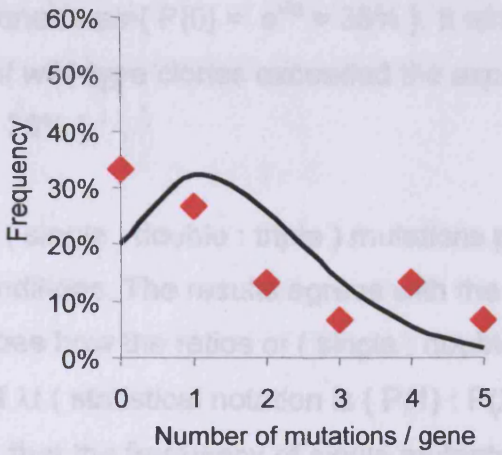


Figure 7.1 Population analyses of results Section 4.1.1

The number of mutations in each clone was analysed. The number of wild-type clones was counted as well as the number of clones that contained exactly one, two, three, four or five mutations (column "Observed"). These counts were converted to relative frequencies by dividing them by the total number of mutations (column "Observed Frequency"). The expected frequencies were calculated using the Poisson distribution. Both frequencies were plotted in a graph (right panel). The black line indicates expected frequency, the red diamonds indicate the observed frequencies.

The frequency of mutants that contain exactly 4 mutations is very high and therefore the observed frequencies do not fit very well with the expected frequencies. Besides this the fit is reasonable. It is beyond the scope of this thesis to prove or disprove that the number of mutations follows a Poisson distribution. (A Goodness of Fit test could be used (Cadwell, 1997)). Two important characteristics of the Poisson distribution are clearly visible:

1 The fact that the population contains a substantial number of wild-type clones.
2 The fact that the more mutations a clone has, the less likely it is to be found in the population.

Lin-Goerke analysed the offspring that was obtained and concluded that the distribution of mutations did not follow a Poisson distribution. It was not specified how this conclusion was reached. It was however obvious that the population contained an excess of wild type clones (50%).

Using the Poisson distribution and the λt that was calculated in Table 7.5 (λt=0.98), the expected fraction of wild-type clones was { $P(0) = e^{-\lambda t} = 38\%$ }. It remains unclear why the observed number of wild-type clones exceeded the expected number of wild-type clones { 50% > 38% }.

Svetlov discussed how the ratios of (single : double : triple) mutations per gene were influenced by altering PCR conditions. The results agrees with the Poisson distribution, which effectively describes how the ratios of (single : double : triple) mutations per gene are a function of λt (statistical notation is { P(1) : P(2) : P(3) }). An important consequence of this is that the frequency of single mutants { P(1) } cannot be manipulated without altering the frequencies of double and triple mutants, because all frequencies are a function of λt.

7.1.6 Doubles in a small sample

When the candidate-positives were analysed (present work) it became clear that there were two identical sequences in the population of 17 clones that were sequenced (Section 4.1.1). This is extremely unlikely to happen if the library contains thousands of unique individuals. Finding doubles in a small sample suggested that the number of unique individuals in the library was limited. An approach based on worst and best case scenario, was used to analyse the composition of the population.

- Best case scenario: The number of unique individuals in the library was high (N = H) and the fact that a sequence occurred twice in the batch (rather than once or not at all) must be explained by a freak accident that has a chance of occurring of only ρ%.
- Worst case scenario: The number of unique individuals in the library was very low (N = L) and the fact that a sequence occurred only twice (rather than three times or more) must be explained by a freak accident that has a chance of occurring of only ρ%.

The Poisson distribution was used to solve the best and worst case scenarios:
The chance λ of one individual being identical to another individual is 1 / N.

$$\lambda = 1 / N$$

The best estimate for the population size when finding 2 doubles in a batch of 17, was therefore 68. The scenarios were analysed for ρ = 10%. The population size for the best and worst case scenarios were 225 and 29 respectively. This could be interpreted as follows: If the population size was more than 225 or less than 29, it would become very unlikely (P(2) <10%) to find exactly two doubles in a batch of 17. It was concluded that the library contained only between 29 and 225 unique individuals. It was concluded that **the mutagenic PCR protocol had generated only a small fraction of the 2120 individuals** that lie one amino acid away from NerA.

It is not uncommon to find doubles in a small random sample.
Shafikhani (1997) reported that two doubles were found in a sample of 30 randomly picked clones. It is likely that the population generated by Shafikhani also contained a very limited number of unique individuals.

7.1.7 Screening

The work of some other authors was analysed for screen size and number of positives found (Table 7.6).

Table 7.6 Library size

Data were collected from various authors who had performed a functional screen on an expression library of random mutants. The size of library that was screened and the number of confirmed positives that were found, were displayed in this table.

Reference	Library	Positives
This work	30000	0
Lönn (2002)	Unknown	3
Riegert (2001)	5000	3
Khono (2000)	60000	1

It becomes clear that the results obtained by screening a random library were very meagre (Table 7.6). The fact that there were no positives found in this thesis is most probably due to the fact that only a limited number of mutants of the model enzyme had been generated (Section 7.1.6).

It remains unclear how fortunate Lönn was when finding the three positives, as no information was reported about the characteristics of the offspring or the size of the screen.

Riegert did not analyse the characteristics of the offspring, so nothing is known about λt. According to the calculations in Chapter 3, the screen size should be at least 17x the number of base pairs in a gene (if λt is optimal). The gene encoding the model enzyme that was used by Riegert (extradiol dioxygenase) has a length of around 900bp. This means the screen size should be at least { 17 x 900 = 15300 }. The screen size that was used (5000) was clearly too small. However in Chapter 3 it was assumed that all exchanges had to be examined in wild-type context. Only mutants that carried exactly one non-silent mutation were considered to be useful members of the offspring. This assumption does not hold under the conditions used by Riegert and therefore the screening strategy used by Riegert was re-examined.

The three positives found by Riegert contained two or three amino acid exchanges but only one of these exchanges was found to be responsible for the improved performance, the other exchanges in the positives were merely tolerated (Table 7.7).

Table 7.7 Amino acid exchanges achieved by Riegert

Riegert reported three mutants containing a total of 7 amino acid exchanges. All 7 exchanges were examined in wild-type context and only two of these exchanges were found to have an impact on the performance of the enzyme.

Impact	Exchange		Impact	Exchange	
-	M 37 T	Mutant 1	-	K 22 E	Mutant 3
+	F 66 L		+	D 155 Y	
-	N 96 T	Mutant 2			
-	K 106 R				

Mutant 1 and Mutant 2 shared the exchange (F 66 L) and this exchange was responsible for the enhanced enzyme performance (Table 7.7). This in fact means that Riegert found the exchange F 66 L two times in the library and that this exchange was not hindered by the company of one or two other exchanges.

If it is assumed that the majority of exchanges are tolerated, the exchanges that are beneficial can be found in non-wild type context. Several exchanges can be examined per clone and this greatly enhances the efficiency of the screening process. Suppose that the average exchange rate ($\lambda t_{Non\ Silent.}$) was around 2 per gene then the entire population that was screened (5000) contained { 5000 x 2 = 10000 } exchanges.
In an earlier analysis of Riegert's results (Table 7.3) it was concluded that around 50% of exchanges are caused by class 1 mutagenic events. Out of the 10.000 exchanges 5000 are thereforelikely to be caused by Class 1 mutagenic events. The remaining 5000 exchanges represent the five other mutagenic classes. If these assumptions are correct, the screen evaluated around 5000 amino acid exchanges. The length of the model enzyme used by Riegert was 900 bp or 300 codons. The number of structures that lie one amino acid away from this enzyme is { 300 x 5.7 = 1710 }.
The conclusion is that the 1710 structures were evaluated in a screen with the size 5000. The "rule of thumb" to over screen 3 fold was almost met (Section 3.4). With the proviso that the assumptions made in this section are correct, Riegert managed to evaluate all structures that lie one amino acid away from the model enzyme. This success relied on the fact that the majority of amino acid exchanges are tolerated and that beneficial exchanges can be found in spite of "pollution" by other mutations.

Khono found one mutant with enhanced properties. This mutant contained five base pair changes resulting in three amino acid exchanges. These exchanges were examine individually in wild-type context and one exchange proved to be responsible for the enhanced properties. The fact that the beneficial exchange was accompanied by two tolerated exchanges and two more silent mutations indicates that the mutagenic rate was not particularly low. It is therefore remarkable that the beneficial exchange was found only once in the over-sized library (60.000).

Possible explanations are:

- Pure coincidence.
- The library contained many clones without insert.
- Many exchanges were deleterious. The beneficial exchange did not manifest itself if it was accompanied by deleterious exchanges.

This matter can not be explored unless more is known about the characteristics of the offspring. It remains unclear how successful Khono was in exploring the sequence space around the model enzyme.

7.2 Structural and functional analysis of NerA *and its homologues*

NerA was classified as an α/β barrel oxido reductase, because of its sequence homology to other α/β barrel oxido reductases (Snape, 1997). The most extensively studied α/β barrel oxido reductase is Old Yellow Enzyme (Oye) and its 3D structure has been solved (Fox, 1994). In order to appreciate the structure and function of the Old Yellow family the tertiary structure of Oye will be discussed in Part I in this sub-chapter. Part I describes the arrangements of the secondary structures within the tertiary structure of Oye. The various structures were named for the purpose of cross reference in this thesis. Insight in the arrangement of secondary structures is later used to appreciate some features that become visible in a family alignment (Parts II and III). The purpose of this sub chapter is to gain insight into the 3D structure of Oye in order to understand the relevance of amino acid exchanges that were found in mutants of NerA (Section 7.3).

Part I. Three-dimensional structure of Old Yellow enzyme

This part describes the tertiary structure of Oye. Pictures like those depicted in Figure 7.3A and Figure 7.4 A, C and D have been generated with Swiss PDB viewer v3.7 (Guex, 1997) using the co-ordinates of Oye posted on the Swissprot database (Boeckmann, 2003).

7.2.1 The β Barrel

Oye is a globular protein with a β barrel "core" composed of 10 β sheets (Figure 7.3).

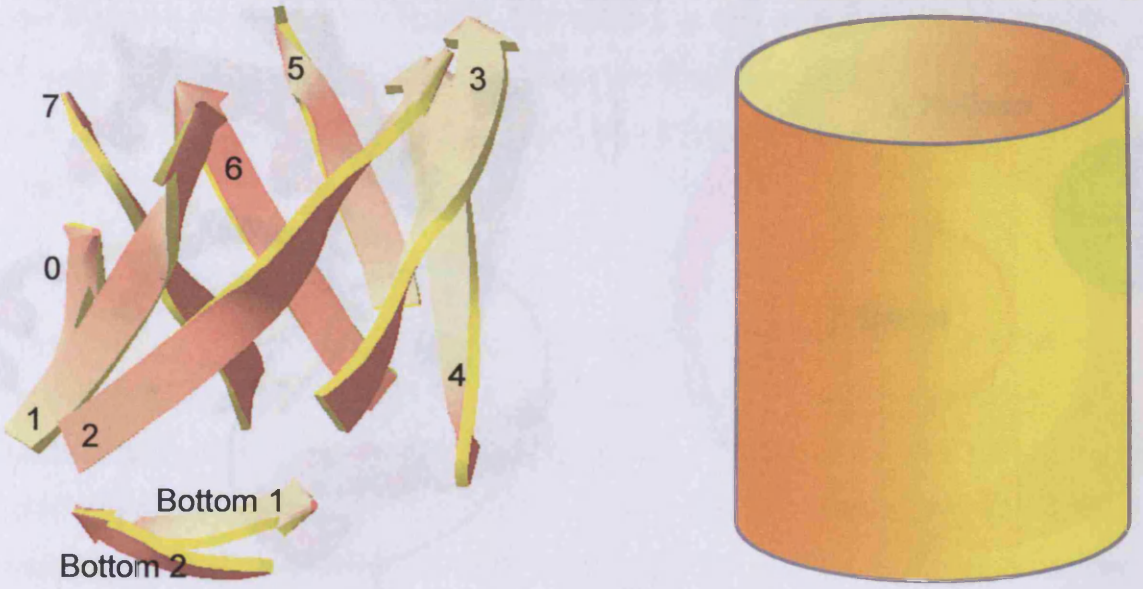


Figure 7.3 The β barrel structure

(A) 8 β sheets form a barrel. Most sheets are 6 amino acids long. These 8 sheets are numbered 0 to 7 according to their position within the encoding gene (See Part III) The bottom of the barrel is formed by two shorter β sheets (indicated as Bottom 1 and 2). (B) Cartoon representation of A. The yellow cylinder symbolises the β barrel in the pictures in this chapter.

The β barrel "core" is a common feature in all α/β barrel oxido reductases. The β barrel is surrounded by a wreath of α helixes that make the hydrophobic β barrel soluble (Figure 7.4A,B). The α helices are amphipathic and possess a hydrophobic inner side that interacts with the β barrel and a hydrophilic outside that interacts with the solvent (Figure 7.4 C,D). Any given α helix in the α wreath lies closest to the β sheet that bears the same number and the two are connected by short chain of amino acids. This short chain lies at the bottom of the β barrel. Therefore the coding regions for the α helix and the β sheet lie adjacent to each other on the *oye* gene. There is one exception to this rule. The coding region for β sheet number 0 lies near the N terminal end of the sequence but the code for α helix number 0 lies at the C terminal end (Figure 7.13).

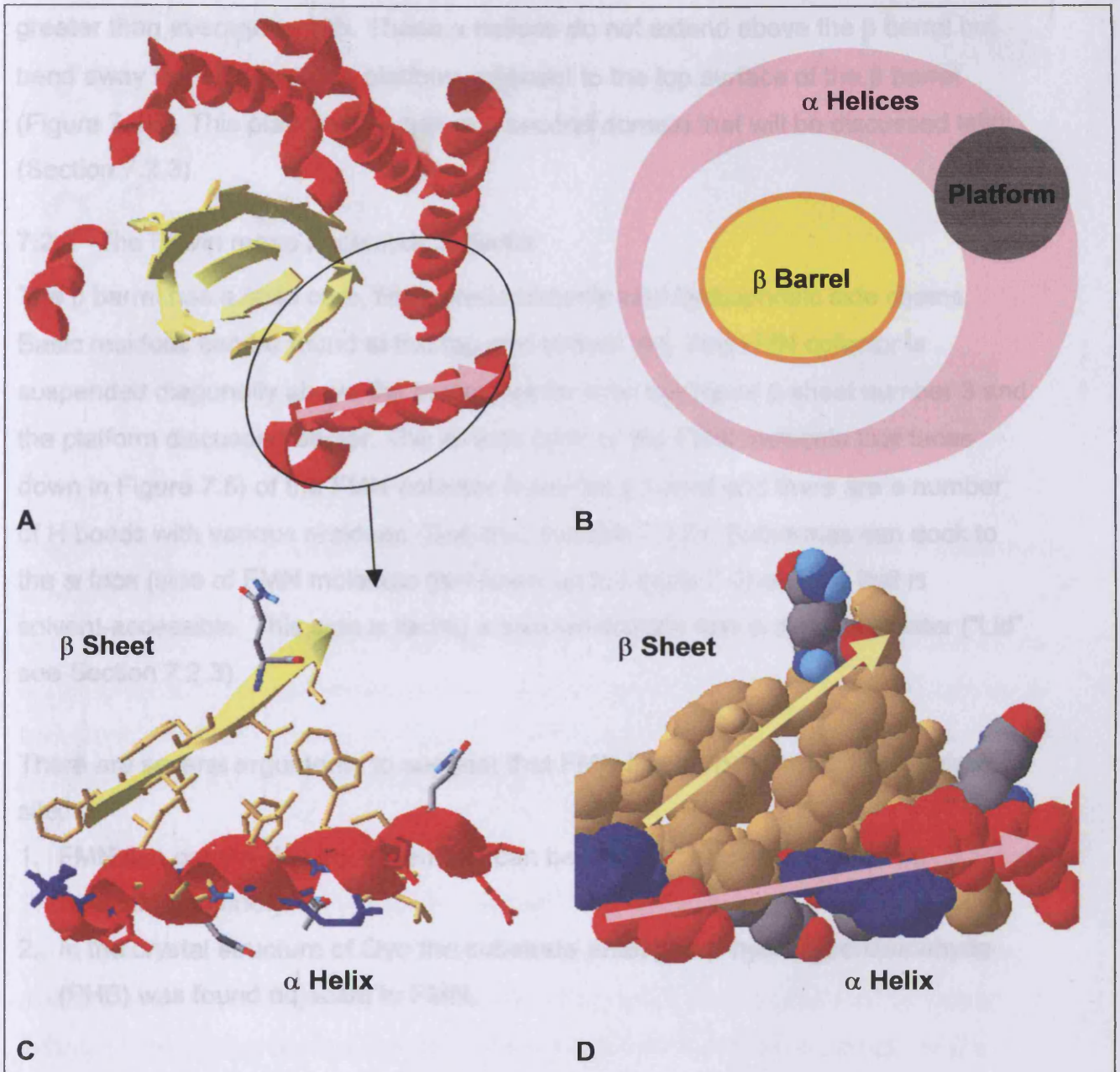


Figure 7.4 Central β Barrel surrounded by α Helixes

(A) Here the β barrel (Figure 7.3) is viewed from the top. The β barrel is surrounded by α helices. (B) The β barrel is surrounded by a wreath of α helices (Cartoon representation of A). A platform is formed by two α helices that bend away form the β barrel. (C) A schematic and (D) a space-filled representation of the interaction between an individual β sheet and an individual α helix (Detail taken from (A, oval section)). The α helix has a hydrophobic side (residues in yellow) that faces the β sheet, and a polar side (residues in red and blue) that interacts with the solvent.

The α wreath around the β barrel is asymmetric because two α helices are of greater than average length. These α helices do not extend above the β barrel but bend away from it, leaving a platform adjacent to the top surface of the β barrel (Figure 7.4B). This platform lies below a second domain that will be discussed later (Section 7.2.3).

7.2.2 The Flavin mono nucleotide cofactor

The β barrel has a solid core, filled predominantly with hydrophobic side chains. Basic residues can be found at the top and bottom rim. The FMN cofactor is suspended diagonally above the barrel, not far from the top of β sheet number 3 and the platform discussed earlier. The *re* face (side of the FMN molecule that faces down in Figure 7.5) of the FMN cofactor faces the β barrel and there are a number of H bonds with various residues (See also Section 7.2.6). Substrates can dock to the *si* face (side of FMN molecule that faces up in Figure 7.5) of FMN that is solvent-accessible. This side is facing a second domain that is discussed later ("Lid" see Section 7.2.3).

There are several arguments to suggest that FMN lies at the heart of the catalytic site.

1. FMN can carry redox equivalents (it can be reduced and re-oxidised in a reversible manner).
2. In the crystal structure of Oye the substrate analogue p-hydroxybenzaldehyde (PHB) was found adjacent to FMN.

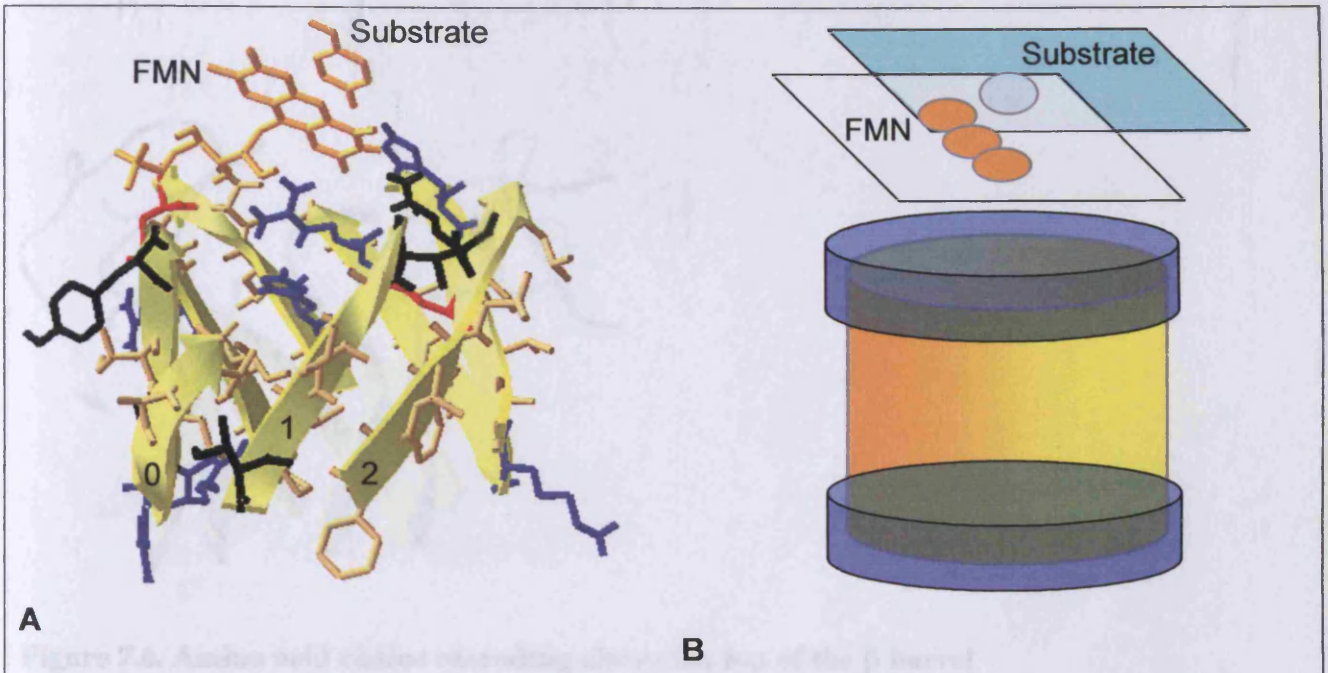


Figure 7.5. Location of amino acid residues, cofactor and substrate analogue around the β barrel in Oye

(A) The β barrel has a hydrophobic core (indicated in yellow). Basic residues (indicated in blue) can be found at the top and bottom rim. (B) Cartoon representation of A. The FMN cofactor is suspended in a slanting plane above the top of the barrel. The substrate analogue can be found in a plane behind and parallel to the FMN plane (shaded green).

The characteristic β barrel and α helix wreath have led to the name "α/β barrel oxido reductase". The β barrel forms the platform on which the enzymatic reduction takes place. Earlier in this section it was described how short pieces of amino acid chain that lie at the bottom of the β barrel connect the β barrel with the α wreath. At the top of the barrel the amino acid chains extend from the β sheets into the active site (Figure 7.6). These regions downstream of the β sheets are therefore expected to have a great influence on the catalytic properties of the enzyme (Fersht, 1999). Many conserved regions are found in these chains and some are involved in the binding of FMN (Figure 7.15).

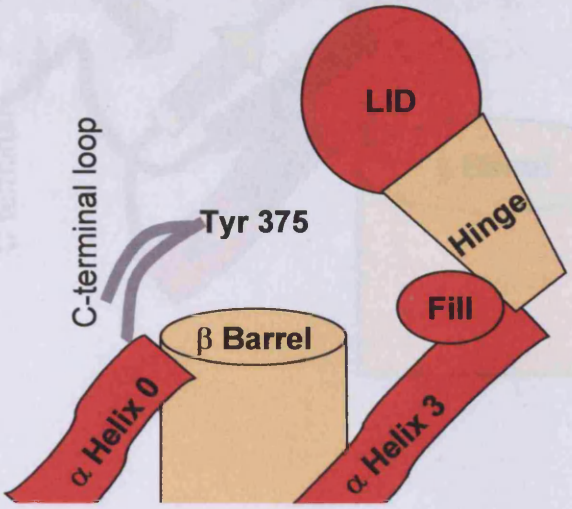
Figure 7.6. Amino acid chains extending above the top of the β barrel

Amino acid chains that lie downstream of the β sheets extent into the area above the β barrel.

7.2.3 A lid over the active site

The β barrel lies beneath FMN and the active site. There are also two domains that form a lid over the active site (Figure 7.7A). One domain is formed by a C terminal loop that reaches into the active site and places a Tyrosine residue (Tyr 375) close to the active site. The other domain lies on top of the platform that is formed by α helices 3 and 4 (indicated as "platform" in Figure 7.4B). This domain consists of two α helices (αTop1 and αTop2) that are connected to the bulk of the enzyme via a hinge that is made of two β sheets (βHinge1 and βHinge2, Figure 7.7C).

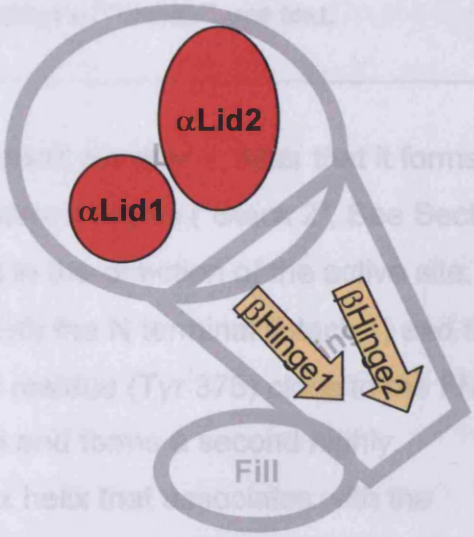
A



B

Figure 7.7 Lid over active site

A. Above the β barrel there are two domains that form a lid over the active site of Oye. B Cartoon representation of (A). On the left hand side there is a C terminal loop that places Tyr 375 very close to the active site. On the right hand side α helix 3 bends away from the β barrel and forms a platform for a lid that is connected to the bulk of the enzyme via a hinge. (Here the platform is viewed from the side, in Figure 7.4B it is viewed from the top).



C

C. The lid consists of two α helices. The hinge consists of two β sheets. There is an additional α helix that is found between the hinge and the β barrel. It is indicated by the name "fill".

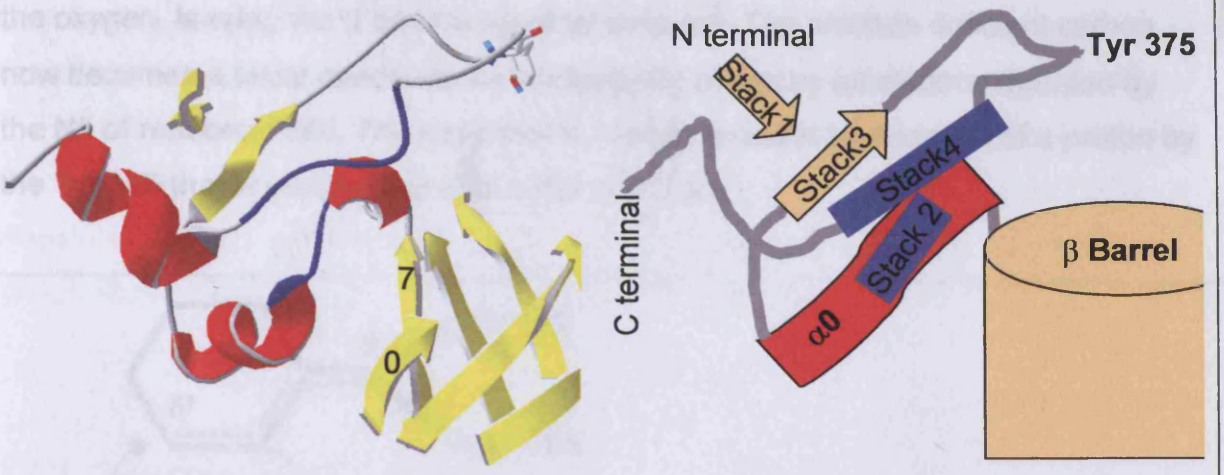


Figure 7.8 C-terminal loop

The C terminal loop reaches into the active site to place a Tyrosine residue close to the active site. The C terminal loop is stabilised by a number of "stacks", see text.

The C terminal loop leaves the β barrel after β sheet number 7. After that it forms α helix 0, this helix is interrupted by a highly conserved region ("Stack 2", See Section 7.2.7). After α helix 0 the C terminal loop bends in the direction of the active site. First it forms a β sheet (Stack 3) that interacts with the N terminal (Stack 1) and then it extends into the active site, placing a tyrosine residue (Tyr 375) close to the FMN cofactor. It now bends away from the active site and forms a second highly conserved region "Stack 4". Finally it forms an α helix that associates with the outside of the protein molecule. The interaction between Stack 2 and Stack 4, together with the interaction between the N terminal and β sheet "Stack 3" may play a role in stabilising the C terminal loop.

7.2.4 Catalytic reaction mechanism

Oye was crystallised in the presence of the substrate analog p-hydroxybenzaldehyde (PHB). Spain studied the binding of PHB in the active site and proposed a reaction mechanism for the reduction of 2 cyclohexenone (CHN), a known substrate of Oye (Spain, 2000). Spain's proposal was as follows: The substrate forms hydrogen bonds with His 191 and Asn 194 and is thereby positioned over the FMN cofactor (Figure 7.9). Tyr 375 is also involved in substrate

binding (not further discussed here). The hydrogen bonds displace the electrons to the oxygen, leaving the β carbon electron deficient. The electron deficient carbon now becomes a likely candidate for nucleophilic attack by an electron donated by the N5 of reduced FMN. The reduction is made complete by donation of a proton by the Tyr 196 that is suspended above the substrate.

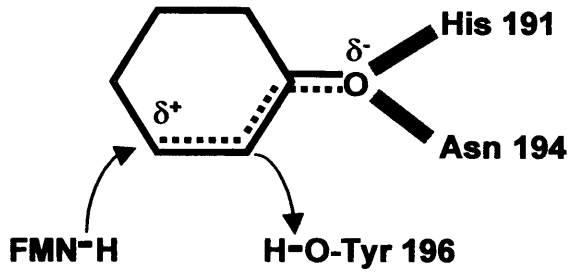


Figure 7.9 Reduction of 2-cyclohexenone by Oye

2 Cyclohexenone (CHN) forms hydrogen bonds with His 191 and Asn 194. The electron deficient β carbon in CHN is a likely candidate for nucleophilic attack by an electron, donated by reduced FMN. Taken from Spain, et al. (2000).

His191 and Asn194 play an important role in positioning the substrate / NAD(P)H above the FMN cofactor. Their role was studied in Oye by (Brown, 1998) using mutation H191N and a combination of mutation H191N and N194H. Although the H191N mutation does not greatly affect NADPH turnover in Oye, the effectiveness of CHN as an electron acceptor was greatly reduced. In the combined mutations H191N and N194H the rate of reduction by NADPH was 15-fold increased, whereas the rate of oxidation by CHN decreased dramatically. The double mutation is an example of how optimisation of one half-reaction (reduction by NADH) might go at the expense of the other half-reaction (oxidation by CHN). Brown therefore suggested that wild type Oye is not optimised for reduction by NADPH alone, but that it is a compromise that takes both reaction halves into account.

The optimisation of the reductive half-reaction by introducing the double mutation does not always go at the expense of the oxidative half reaction. The oxidation by oxygen, for example, was hardly affected. Presumably the double mutant makes a better oxygenase since the first half-reaction is improved and the second half-reaction remains unaffected.

Oye can reduce GTN and it was reported that the reductive half-reaction was the rate limiting step in the conversion of GTN by Oye (Meah, 2001). The reductive half-reaction can be optimised by introducing the H191N and N194H mutations and if this does not negatively affect the rate of oxidation by GTN, the double mutant would be expected to be a better GTN reductase. The H191N mutation alone has a negative effect on rate of oxidation by GTN but the double mutation has not been tested (Meah, 2001).

Part II. Family alignment

7.2.5 Sequence homologues of NerA

Snape performed a sequence analysis on the deduced amino acid sequence of NerA and concluded that NerA had high sequence similarity to Old yellow enzyme (Oye3, Snape, 1997). Snape also aligned NerA to a number of other Oye homologues. In order to update this alignment the amino acid sequence of NerA was submitted to the SWISSPROT database (Fast search, default parameters Pearson, 1988). Sequences that had a score higher than that of old yellow enzyme (Oye3) were aligned using ClustalW, default parameters. (Thompson, 1994).

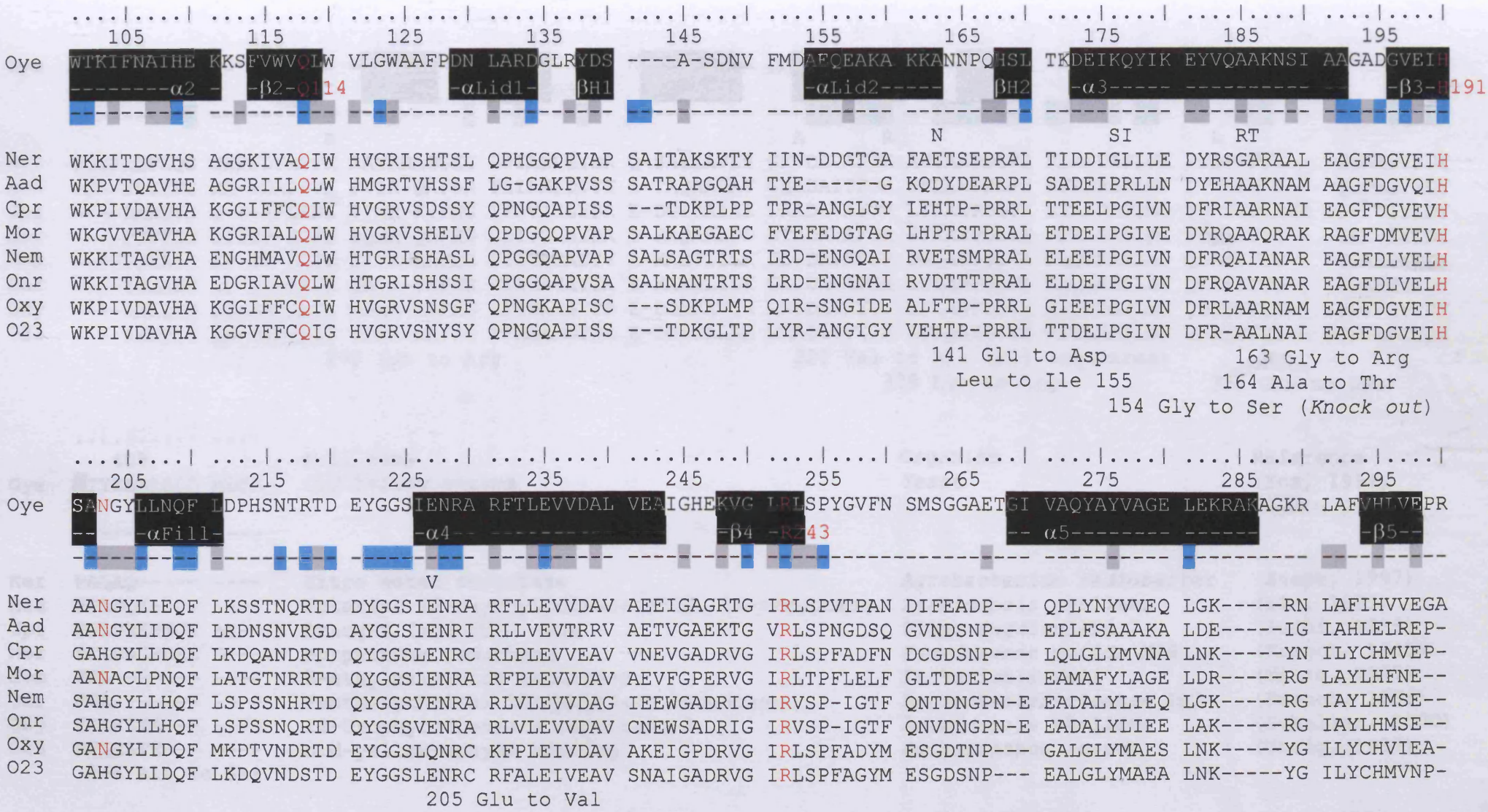
One particular sequence encoded the putative protein CprD. It was obtained from a cDNA library (*Vigna unguiculata,* Cow Pea). Unlike the other sequences it was not cloned in an expression vector. The other sequences encoded genes that had been cloned, expressed and analysed. These genes all encoded NAD(P)H dependent reductases. Three different copies of 12-oxyphytodienoate-10, 11-reductase were among the sequences that were found. Only the copy with the highest sequence similarity to NerA was used in the final alignment Figure 7.10.

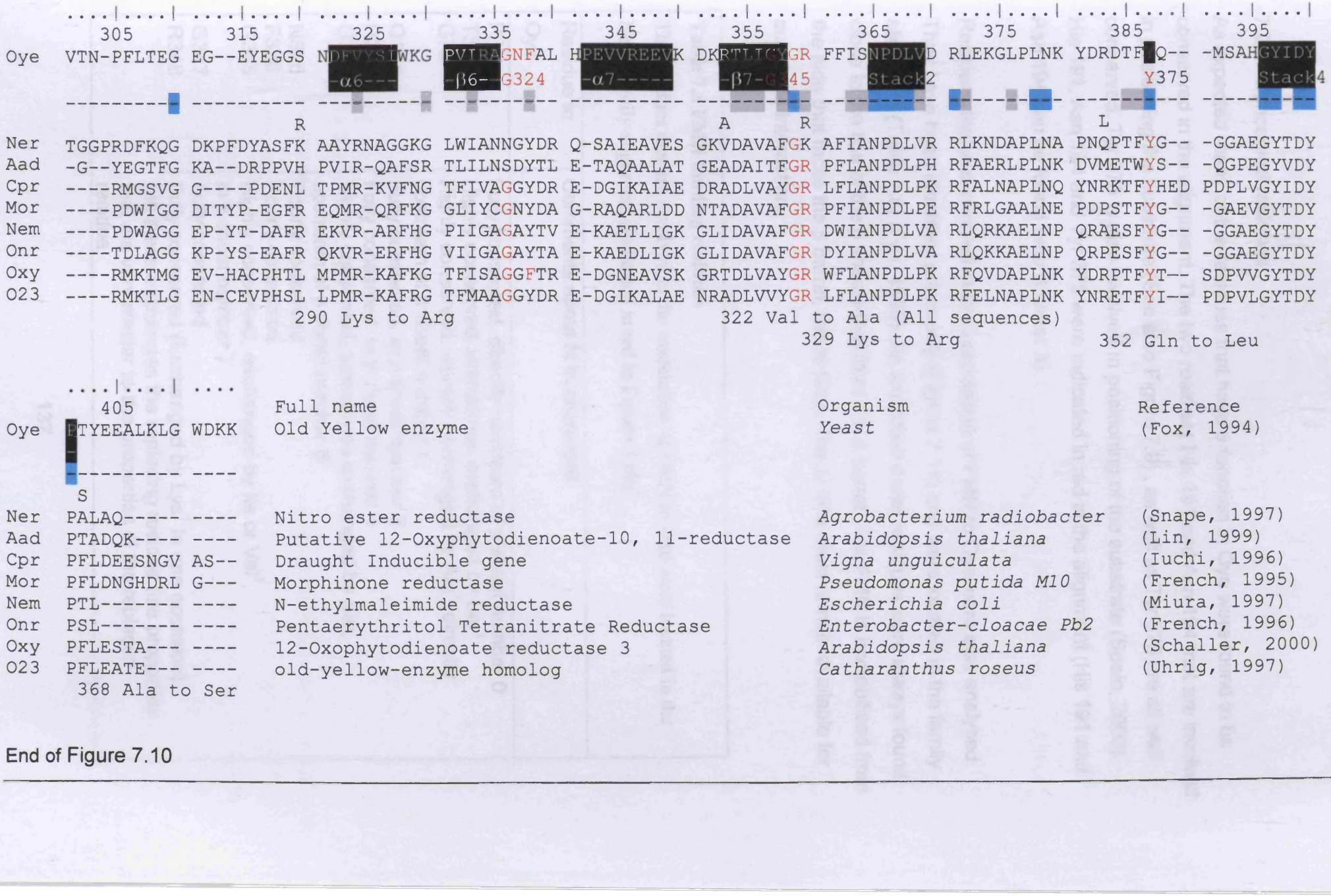
Onr, an organic-nitrate reductase from *Enterobacter cloacae* (Section 1.1.3) was among the sequences with highest similarity. It has activity towards GTN, just like NerA and Oye. It was not tested if any of the other sequences in the alignment encode proteins with activity towards nitro-organic compounds.

Figure 7.10 Alignment of homologues of NerA

Alignment of the amino acid sequences of Old yellow enzyme (Oye), nitro ester reductase (Ner), 12-oxyphytodienoate-10, 11-reductase (Aad), Draught Inducible gene (Cpr), morphinone reductase (Mor), N-ethylmaleimide reductase (Nem), pentaerythritol tetranitrate reductase (Onr), 12-oxophytodienoate reductase 3 (Oxy) and old-yellow-enzyme homologue O23. The origin of these sequences can be found on the last page of this alignment. The alignment was generated using ClustalW, default parameters. Some secondary structure landmarks were indicated by black boxes in the primary sequence of Oye. Residues that were involved in the binding of substrate or FMN were indicated in red. If these same residues were found in other sequences in the alignment, they were also indicated in red. The sequence line between Ner and Oye indicates conserved regions. If only one amino acid was found at a certain location, this location was marked in blue. If no more than two different amino acids were found, the location was marked in grey. Mutations that were created in NerA in the present work are indicated by single code, three letter code and location.

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               5         15         25         35         45         55         65         75         85         95
Oye       ----SFVKDF KPQALGDTNL FKPIKLGNNE LLHRAVIPPL TRMRALHPGN IPNRDWAVEY YTQRAQRPGT MIITEGAFIS PQAGGYDNAP GVWSEEQMVE
                                 BTM   BTM-  β0-   P35                          -α1- --          -β1---  G72                    -α2---
          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
          Mutations: Single letter code                          P                    T
Ner       ---------- ----MTS--L FEPAQAGDIA LANRIVMAPL TRNR--SPGA IPNN-LNATY YEQRAT-AG- LIVTEGTPIS QQGQGYADVP GLYKREAIEG
Aad       ---------- ----MPT--L FDPLTLGDLQ SPNRVLMAPL TRGR-ATREH VPTE-LMIEY YTQRAS-AG- LIITEATGIT QEGLGWPYAP GIWSDEQVEA
Cpr       MVTATATAAA TAPSHQAIPL LTPYKMGKFN LSHRVVLAPL SRER--SYNN VPQP-HAVVY YSQRTS-NGG LLIAEATGVS DTAQGYPNTP GIWTKEQVEA
Mor       ---------M PDTSFSNPGL FTPLQLGSLS LPNRVIMAPL TRSR--TPDS VPGR-LQQIY YGQRAS-AG- LIISEATNIS PTARGYVYTP GIWTDAQEAG
Nem       ---------- ----MSSEKL YSPLKVGAIT AANRIFMAPL TRLRSIEPGD IPTP-LMAEY YRQRAS-AG- LIISEATQIS AQAKGYAGAP GIHSPEQIAA
Onr       ---------- ----MSAEKL FTPLKVGAVT APNRVFMAPL TRLRSIEPGD IPTP-LMGEY YRQRAS-AG- LIISEATQIS AQAKGYAGAP GLHSPEQIAA
Oxy       --------ME NGEAKQSVPL LTPYKMGRFN LSHRVVLAPL TRQR--SYGN VPQP-HAAIY YSQRTT-PGG FLITEATGVS DTAQGYQDTP GIWTKEHVEA
O23       MAKQVDNGQE EQQQNPQNPL LTPYKLGKFQ LSHRIVLAPL TRQR--SYGN VPQP-HAVLY YSQRTS-KGG LLISEAAGVS NTAQGYPMTP GIWTKEQVEA
          Mutations: Location and 3 letter code       25 Thr to Pro            46 Ala to Thr (Knock out)
```


105 115 125 135 145 155 165 175 185 195
Oye WTKIFNAIHE KKSFVWVQLW VLGWAAFPDN LARDGLRYDS ----A-SDNV FMDAFQEAKA KKANNPQHSL TKDEIKQYIK EYVQAAKNSI AAGADGVEIH
α2 β2 αLid1 βH1 αLid2 βH2 α3 β3 H191
N SI RT
Ner WKKITDGVHS AGGKIVAQIW HVGRISHTSL QPHGGQPVAP SAITAKSKTY IIN-DDGTGA FAETSEPRAL TIDDIGLILE DYRSGARAAL EAGFDGVEIH
Aad WKPVTQAVHE AGGRIILQLW HMGRTVHSSF LG-GAKPVSS SATRAPGQAH TYE------G KQDYDEARPL SADEIPRLLN DYEHAAKNAM AAGFDGVQIH
Cpr WKPIVDAVHA KGGIFFCQIW HVGRVSDSSY QPNGQAPISS ---TDKPLPP TPR-ANGLGY IEHTP-PRRL TTEELPGIVN DFRIAARNAI EAGFDGVEVH
Mor WKGVVEAVHA KGGRIALQLW HVGRVSHELV QPDGQQPVAP SALKAEGAEC FVEFEDGTAG LHPTSTPRAL ETDEIPGIVE DYRQAAQRAK RAGFDMVEVH
Nem WKKITAGVHA ENGHMAVQLW HTGRISHASL QPGGQAPVAP SALSAGTRTS LRD-ENGQAI RVETSMPRAL ELEEIPGIVN DFRQAIANAR EAGFDLVELH
Onr WKKITAGVHA EDGRIAVQLW HTGRISHSSI QPGGQAPVSA SALNANTRTS LRD-ENGNAI RVDTTTPRAL ELDEIPGIVN DFRQAVANAR EAGFDLVELH
Oxy WKPIVDAVHA KGGIFFCQIW HVGRVSNSGF QPNGKAPISC ---SDKPLMP QIR-SNGIDE ALFTP-PRRL GIEEIPGIVN DFRLAARNAM EAGFDGVEIH
O23 WKPIVDAVHA KGGVFFCQIG HVGRVSNYSY QPNGQAPISS ---TDKGLTP LYR-ANGIGY VEHTP-PRRL TTDELPGIVN DFR-AALNAI EAGFDGVEIH
141 Glu to Asp
Leu to Ile 155
163 Gly to Arg
164 Ala to Thr
154 Gly to Ser (Knock out)
205 215 225 235 245 255 265 275 285 295
Oye SANGYLLNQF LDPHSNTRTD EYGGSIENRA RFTLEVVDAL VEAIGHEKVG LRLSPYGVFN SMSGGAETGI VAQYAYVAGE LEKRAKAGKR LAFVHLVEPR
αFill α4 β4 R243 α5 β5
V
Ner AANGYLIEQF LKSSTNQRTD DYGGSIENRA RFLLEVVDAV AEEIGAGRTG IRLSPVTPAN DIFEADP--- QPLYNYVVEQ LGK-----RN LAFIHVVEGA
Aad AANGYLIDQF LRDNSNVRGD AYGGSIENRI RLLVEVTRRV AETVGAEKTG VRLSPNGDSQ GVNDSNP--- EPLFSAAAKA LDE-----IG IAHLELREP-
Cpr GAHGYLLDQF LKDQANDRTD QYGGSLENRC RLPLEVVEAV VNEVGADRVG IRLSPFADFN DCGDSNP--- LQLGLYMVNA LNK-----YN ILYCHMVEP-
Mor AANACLPNQF LATGTNRRTD QYGGSIENRA RFPLEVVDAV AEVFGPERVG IRLTPFLELF GLTDDEP--- EAMAFYLAGE LDR-----RG LAYLHFNE--
Nem SAHGYLLHQF LSPSSNHRTD QYGGSVENRA RLVLEVVDAG IEEWGADRIG IRVSP-IGTF QNTDNGPN-E EADALYLIEQ LGK-----RG IAYLHMSE--
Onr SAHGYLLHQF LSPSSNQRTD QYGGSVENRA RLVLEVVDAV CNEWSADRIG IRVSP-IGTF QNVDNGPN-E EADALYLIEE LAK-----RG IAYLHMSE--
Oxy GANGYLIDQF MKDTVNDRTD EYGGSLQNRC KFPLEIVDAV AKEIGPDRVG IRLSPFADYM ESGDTNP--- GALGLYMAES LNK-----YG ILYCHVIEA-
O23 GAHGYLIDQF LKDQVNDSTD EYGGSLENRC RFALEIVEAV SNAIGADRVG IRLSPFAGYM ESGDSNP--- EALGLYMAEA LNK-----YG ILYCHMVNP-
205 Glu to Val



End of Figure 7.10

7.2.6 Conserved residues

As expected many of the residues that have a function in Oye were found to be conserved in the alignment. The two residues His 191 and Asn 194 that are involved in positioning the substrate (See also Figure 7.9) , as well as Tyr 375 are all well conserved. Tyr 375 is also involved in positioning of the substrate (Spain, 2000). His 191, Asn 194 and Tyr 375 were indicated in red in the alignment (His 191 and Asn 194 can be found near β sheet 3).

Residues that are involved in the association of FMN to Oye were also analysed. They have been identified previously (Figure 7.11) and were located in the family alignment (Table 7.8) (Fox, 1999). As expected these residues were always found close to the tips of the β sheets that form the β barrel. The FMN is immobilised from the side that faces the β barrel, the *re* face. The *si* side must remain available for substrate interaction.

Table 7.8 FMN binding residues

The residues that are involved in the association of FMN to Oye were located in the NerA family alignment (indicated in red in Figure 7.10).

Residue in Oye	Comments about fit in alignment
P35	Fully conserved, directly upstream of β sheet number 0
T37	Highly conserved, sometimes exchanged by Ser[1]
G72	Highly conserved, usually exchanged by Ala[1], directly upstream of β sheet number 1.
Q114	Fully conserved, in β sheet number 2
R243	Fully conserved, in β sheet number 4
G324	Highly conserved, sometimes exchanged by Ser, upstream of β sheet number 6
N325	Poorly conserved
F326	Poorly conserved
G345	Highly conserved, exchanged by Ile or Val[1], in β sheet number 7.
G347	Fully conserved
R348	Fully conserved (Exchanged by Lys[1] In one occasion)
	[1] In these exchanges the replacing residue has properties that are very similar to the properties of the replaced residue.

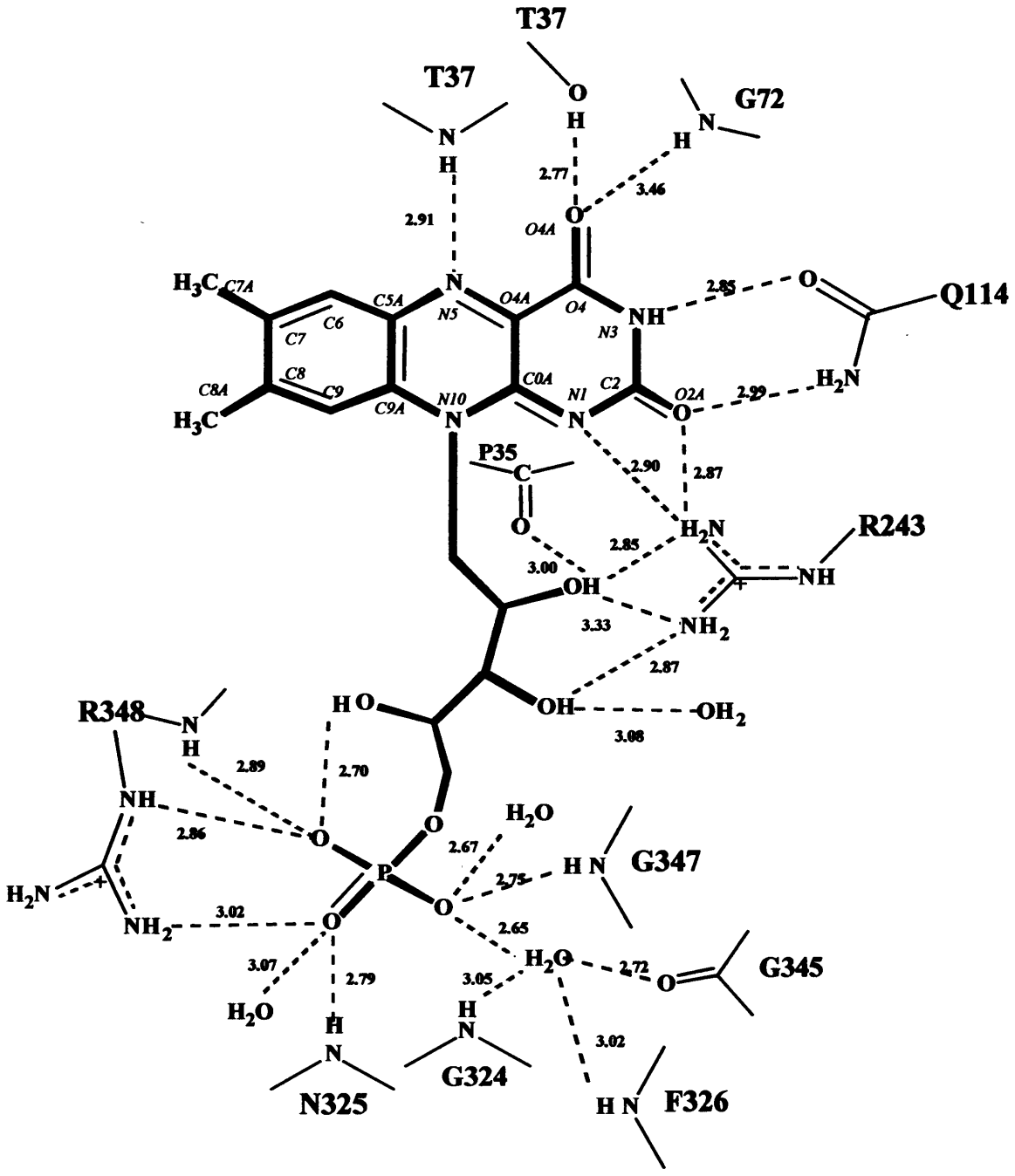


Figure 7.11 Flavin environment in Oye

The protein environment around FMN is shown with the FMN in bold lines, and the amino acid residues are represented by their one-letter code and their numbers in the Oye sequence. Only the portion of the side chain that interacts directly with FMN is shown. Hydrogen bonds between FMN atoms and protein atoms are indicated by dotted lines, and the bond distance is given in Ångstroms. Picture taken (with kind permission) from (Fox, et al. 1999).

In addition to the residues involved in FMN binding, many more conserved residues were found directly downstream of each of the β sheets that form the β barrel. This was expected as these regions extend into the active site (Figure 7.6).

7.2.7 Conserved regions

In some cases the conserved residues cluster together and form conserved regions. The presence of conserved regions further confirmed the significance of the alignment. The conserved regions indicate that there must be similarities in the secondary structure of proteins that are represented by the aligned sequences. Some of these conserved regions will be discussed here.

Directly downstream of β sheet 3 lies "α fill". Many conserved residues are found in the α helix itself as well as its direct surroundings. This is not surprising as α fill lies very close to the FMN cofactor and the active site (Figure 7.7).

Two extensive conserved regions were found in the C terminal region of the alignment: one that interrupts α helix 0 (Stack 2) and one that lies only 14 residues away from the C terminal (Stack 4) (see also right panel in Figure 7.8). In Oye these regions lie stacked on top of each other, too far away from the active site to be directly involved in catalytic action. A possible explanation for this stack could be a (conserved) need for a strong interaction that ties the C terminal to the bulk of the enzyme. This link would prevent a loose C terminal end that is vulnerable to protease degradation and it would also stabilise the C terminal loop that extends into the active site.

The conserved regions "Stack 2" and "Stack 4" were included in the analyses of secondary structures in the next section.

Part III. Secondary structure

A secondary structure motif that is very common to the Old Yellow family is the repetition of an α/β cassette (Figure 7.12, Fersht, 1999). The majority of these cassettes are involved in the formation of the β barrel and the α wreath (Figure 7.4).

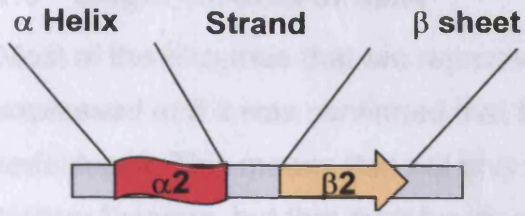


Figure 7.12 α/β Cassette

The cassette "α helix – Short Strand - β sheet" is a common feature in the Old Yellow family.

Indeed this α/β cassette can be found several times in the secondary structure diagram of Oye (Figure 7.13). The β sheets and α helixes were named according to their location in the tertiary structure of Oye (See Part I).

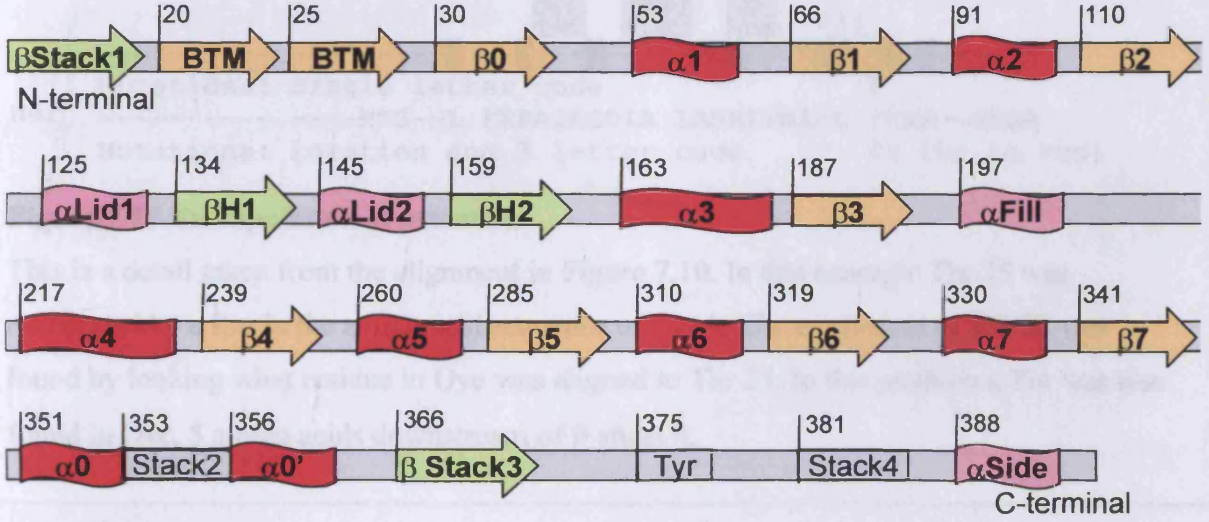


Figure 7.13 Location of Secondary structures on the primary structure of Oye

Overview of the location of the various secondary structures (See Part I in this section) in Oye. β sheets and α helixes are indicated in yellow and red respectively. β sheets and α helices that are involved in the β barrel and α wreath are indicated in dark yellow and red. BTM = βsheet forming the bottom of the barrel (indicated as "Bottom 1" and "Bottom 2" in Figure 7.3.

7.3 Single mutants of NerA

Most of the enzymes that are represented in the alignment have been cloned and expressed and it was confirmed that they are NAD(P)H dependent oxido reductases. This means that not only their primary structure is similar to that of Old Yellow Enzyme, but that their function is also similar. Based on these findings , due to lack of other evidence and the absence of better models, it was assumed that the secondary and tertiary structure of NerA is similar to that of Old Yellow enzyme.

To appreciate the effect that amino acid exchanges have on the function of NerA, the exchanges that were found in NerA were extrapolated to a position in Oye, using the amino acid alignment (Figure 7.10). An example of such an extrapolation is given in Figure 7.14.

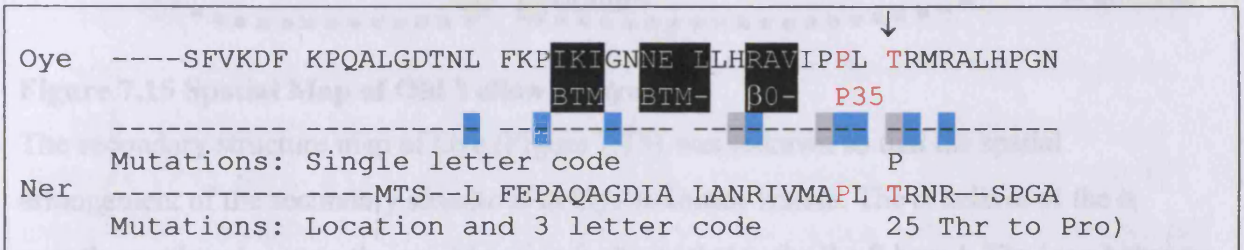


Figure 7.14 Extrapolation of mutation

This is a detail taken from the alignment in Figure 7.10. In this example Thr 25 was exchanged by a Pro in the amino acid sequence of NerA. The equivalent of Thr 25 was found by looking what residue in Oye was aligned to Thr 25. In this position a Thr was also found in Oye, 5 amino acids downstream of β sheet 0.

It is not known how the tertiary structure of NerA compares to that of Oye. Therefore no attempt was made to predict anything about the direct environment of the amino acid exchange. Instead the exchange is discussed in context of the secondary structures that lie nearby.

A spatial map of Oye was drawn to enable a better insight into the arrangement of secondary structures in the tertiary structure of Oye (Figure 7.15). Details of this map are used later to discuss the various amino acid exchanges.

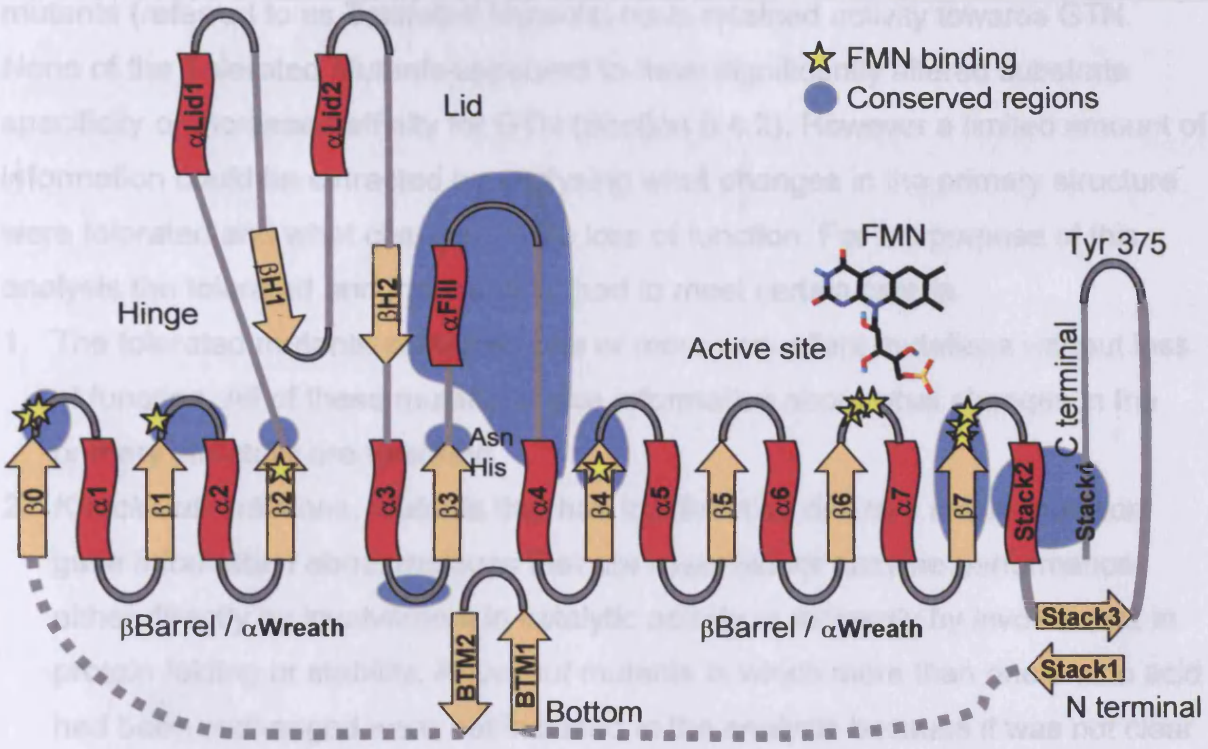


Figure 7.15 Spatial Map of Old Yellow Enzyme

The secondary structure map of Oye (Figure 7.13) was redrawn so that the spatial arrangement of the secondary structures in Oye becomes visible. The α helixes of the α wreath are placed next to the neighbouring β sheets that make the β barrel. The loop between the two β sheets that form the bottom of the β barrel are very close to the loops that connect β sheet 3 with α helix 3 and β sheet 4 with α helix 4 respectively. In the middle and top of the picture, the hinge and top-lid regions are indicated. Conserved regions that were found in the alignment (Figure 7.10) are indicated in blue. Residues involved in FMN binding are indicated with a yellow star. FMN was placed near a number of residues that are involved in the binding of FMN.

Figure 7.15 clearly illustrates that many conserved residues (indicated in blue) are found near the top surface of the β barrel.

Several mutants of NerA were isolated and non-silent mutations in the *nerA* gene have been identified (Section 6.4, a complete list of mutants can be found in Appendix A). Some of these mutants (referred to as *Knock out* Mutants or (*k.o.*)) were found to have lost the ability to reduce GTN or any other nitrate ester. Other

mutants (referred to as Tolerated Mutants) have retained activity towards GTN. None of the Tolerated Mutants appeared to have significantly altered substrate specificity or increased affinity for GTN (Section 6.4.2). However a limited amount of information could be extracted by analysing what changes in the primary structure were tolerated and what changes led to loss of function. For the purpose of this analysis the tolerated and *k.o.* mutants had to meet certain criteria.

1. The tolerated mutants contained one or more non- silent mutations without loss of function. All of these mutations give information about what changes in the primary structure are tolerated.
2. *Knock out* mutations. Mutants that had lost function due to a single mutation gave information about residues that are essential for enzyme performance, either directly by involvement in catalytic activity or indirectly by involvement in protein folding or stability. *Knockout* mutants in which more than one amino acid had been exchanged were not included in the analysis because it was not clear which of the mutations was responsible for the loss of function. Structures that had lost function due to frame-shift events were also not further analysed.

Tolerated mutations and single *knockout* mutations are discussed in Table 7.9, in order of the position of the mutation in the primary sequence, with help of the spatial map of Oye (Figure 7.15).

Table 7.9 Mutants of NerA

Tolerated mutations and single *knockout* mutations are listed in order of position of mutation in the primary sequence. The mutations listed are tolerated (green border) unless indicated as "**knock-out**" (red border). Pictures are taken from Figure 7.15, locations of exchanged amino acid are indicated with an arrow. NOTE numbers refer to the location of the mutation in the primary sequence of NerA. This location was extrapolated to the primary sequence of Oye as explained in Figure 7.14. Figures refer to secondary structures in Oye that lie nearby this extrapolated location.

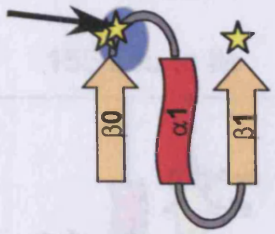
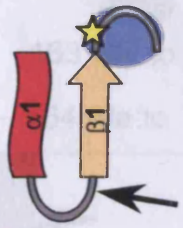
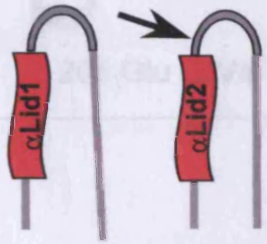
Location in NerA Mutation	Discussion
 25 Thr to Pro	Located inside conserved region downstream of β sheet 0, involved in FMN binding at the top of the β Barrel. It is remarkable that this mutation is tolerated since the OH group of Thr (and sometimes Ser) is involved in FMN binding at the top of the β barrel.
 46 Ala to Thr	***Knockout* mutation**, inside conserved region downstream of β sheet 1. It is located at the bottom of the β barrel, close to the site where the two β sheets "BTM 1" and "BTM 2" form the bottom of the β barrel. It is surprising that Thr is not tolerated at this location since a Thr residue is found in many of the family members in the alignment. It remained unclear why absence of Thr was essential for activity towards GTN.
 141 Glu to Asp	Glu to Asp. Located in the last residue of α Lid 2 in non-conserved region, probably exposed to solvent. Predictably this mutation is tolerated as it is found in a non-conserved region and the structural changes involved are minimal, since Asp is very similar to Glu. Glu might have a task in interacting with the solvent, this task can be taken over by Asp.

Table 7.9 continued.

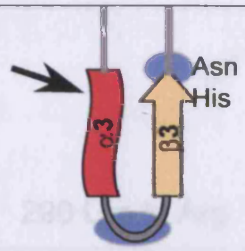
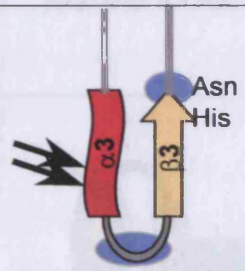
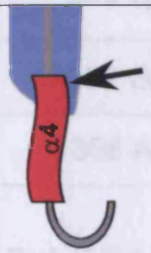
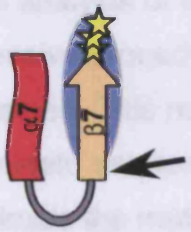
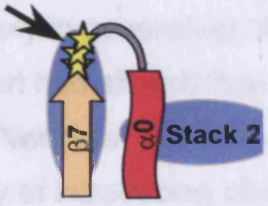
 154 Gly to Ser	***Knockout* mutation**, in hydrophobic region in α helix number 3. It is located below the hinge that is attached to the lid on the active site. The substitution of the apolar Gly with the polar Ser might destabilise the α helix or the conserved hydrophobic region.
155 Leu to Ile	Located in hydrophobic region in α helix number 3. In this region Ile Leu and Val occur often in the alignment. It is therefore not remarkable that this mutation is tolerated.
 163 Gly to Arg 164 Ala to Thr	Both mutations are found in a hydrophobic region in α helix number 3 In spite of the strong change in residue properties (small hydrophobic residue replaced by polar and large, charged residue) these mutations are tolerated. The strong change is perhaps tolerated because the region is not highly conserved.
 205 Glu to Val	Located in highly conserved region at the start of α helix 4. Probably facing the solvent. Located at the top end of the α wreath close to the hinge region. Remarkable that this mutation is tolerated since it lies in the middle of a conserved region and the acidic Glu, possibly interacting with the solvent, is replaced by a hydrophobic Val.

Table 7.9 continued.

290 Lys to Arg	Located in poorly conserved region at beginning of α helix 6. Mutation is found in an extended region with no conservation. Arg is sometimes found in the other family members. The residue Arg is very similar to residue Lys.
 322 Val to Ala	This mutation was found in all mutants of NerA that were analysed, at the beginning of β sheet 7. Ala is found in four other family members in the alignment. Because this mutation is found in all mutants, it is likely that the native NerA actually had an Ala in this position but that an error was made when the sequence of native NerA was submitted to the Swiss-Prot database. This mutation could also have been introduced in an early stage when the *nerA* gene was cloned.
 329 Lys to Arg	Located in highly conserved region directly downstream of β sheet 7. Other family members of NerA all have the Arg at this position.
352 Gln to Leu	Both mutations were found in the C-terminal loop. In this loop only Tyr 375 and the conserved region "Stack 4" seem to be of importance.
368 Ala to Ser	

End of Table 7.9

It is unfortunate that none of the mutations discussed in Table 7.9 lead to altered substrate specificity. Information about whether a mutation is tolerated or not can lead to insights into structure function relationships, but not enough mutations were obtained to perform such an analysis on NerA. It would be of interest to re-evaluate the mutations in Table 7.9 when the 3D structure of NerA becomes available.

7.4 Conclusion

7.4.1 Mutagenic PCR: Screening or sequencing?

The analysis of the characteristics of mutagenic PCR is an expensive and labour intensive procedure since the sequence of a number of library members must be obtained. This requires that each of these members is sequenced several times to eliminate sequencing errors and to detect single mutations. Every attempt to optimise the mutagenic protocol must be followed by this labour intensive analysis. It is tempting to skip this analysis and to proceed directly to the screening of the library using the argument "as long as there are winners". However the two mutagenic protocols that have been analysed in the thesis both performed very poorly in generating change. Unfortunately this was only discovered after a lot of effort had already been invested in screening and preliminary characterisation and no "winners" were found. At first glance mutagenic PCR seems a quick and easy way of introducing change but it must be emphasised that it can be biased so that only a few new structures are generated. Furthermore, mutagenic PCR has the appearance of being a cheap method but it should be taken into account that the analysis of its characteristics (for example the extent of transition / transversion bias) requires many sequence runs.

7.4.2 Family

One of the requirements to study structure and function relationships, is the availability of both the structure and function in a single model enzyme. The study into the family of NerA failed to meet this requirement because little was known about the function of structural family and attempts to find the structure of functional family failed. It would be of great interest to test the structural family members of NerA for activity against a number of nitro-organic compounds. Indeed by the time this thesis was written a report indicated that Old yellow enzyme can reduce GTN (Meah, 2001).

7.4.3 Bioremediation of explosives

A major finding was the fact that the ability to liberate NO_2^- from GTN was wide spread among micro organisms. This is not surprising now that it is known that reduced FMN alone can liberate NO_2^- from all three nitro groups of GTN (Meah, 2001). This finding implies that GTN must be readily mineralised in the environment and that it does not pose a threat as a pollutant. Since the enzyme did not appear to play a crucial role in the reduction of GTN, this is also likely to be the case for EGDN and PETN. It would be of interest to confirm that reduced FMN can liberate nitrite from EGDN and PETN, because if this were the case, PETN and EGDN would be readily degraded in the environment just like GTN.

7.4.4 Starting point for exploration

NerA was chosen as a starting point for the exploration of the protein landscape because of the assumption that NerA is a rare enzyme and that it would prove difficult to find other enzymes with similar function (section 1.2.5). This assumption proved to be wrong and perhaps NerA was therefore not the best starting point. Now that it has been reported that Oye can also reduce GTN, it might make a better starting point because the crystal structure for Oye has been resolved. However, NerA has the advantage that, unlike Oye, it is not readily oxidised by oxygen. It is therefore not necessary to maintain anaerobic conditions when assaying NerA and that makes procedures less complicated. Furthermore NerA has a preference for the cheaper NADH cofactor (compared to NADPH) so that recycling of cofactor in the assay is not so much of an issue.

Besides NerA and Oye there are numerous members in the Old Yellow family that might have activity towards GTN. Any of these members might prove to be a more suitable starting point for exploration. Family shuffling could take all of them into consideration.

7.4.5 Enzymatic conversion of xenobiotics

Initial degradation studies of GTN were dependent on the relatively laborious HPLC methods to detect glycerol dinitrate. When it became clear that NO_2^- was released and that a reductase was involved, the assay to detect activity towards GTN was greatly simplified. Activity could be detected with help of the azo dyes or kinetic photospectometry for the detection of consumption of NADH. As soon as the simplified assay became available it could have been used to test cell extract of a number of well-studied micro organisms, like *E. coli* or *Bacillus subtilis*. If activity was found, the identifying and cloning of the reductases involved would have been facilitated because of the available genome sequence. This is also true for any other study that involves degradation of xenobiotic substrates. If an assay is available, popular lab strains should be tested first.

Bibliography

Achtnich, C. & Lenke, H. (2001). Stability of immobilized 2,4,6-trinitrotoluene metabolites in soil under long-term leaching conditions. *Environmental Toxicology and Chemistry* **20**(2), 280-283.

Aita, T. & Husimi, Y. (1996). Fitness spectrum among random mutant on mount Fuji-type fitness landscape. *Journal of Theoretical Biology* **182**, 469-485.

Altamirano, M. M., Blackburn, J. M., Aguayo, C. & Fersht, A. R. (2000). Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* **403**(6770), 617-22.

Alvarez, M. A., Kitts, C. L., Botsford, J. L. & Unkefer, P. J. (1995). *Pseudomonas-aeruginosa* strain Mao1 aerobically metabolizes the aminodinitrotoluenes produced by 2,4,6-trinitrotoluene nitro-group reduction. *Canadian Journal of Microbiology* **41**(11), 984-991.

Arnold, F. H. (1996). Directed evolution: Creating biocatalysts for the future. *Chemical Engineering Science* **51**(23), 5091-5102.

Arnold, F. H. (1997). Directed evolution of pNB esterase: Activation and stabilization of an industrial enzyme. *Abstracts of Papers of the American Chemical Society* **213**(Pt1), 13-BIOT.

Bayman, P., Ritchey, S. D. & Bennett, J. W. (1995). Fungal interactions with the explosive RDX (hexahydro-1,3,5-trinitro- 1,3,3-triazine). *Journal of Industrial Microbiology* **15**(5), 418-423.

Blehert, D. S., Fox, B. G. & Chambliss, G. H. (1999). Cloning and sequence analysis of two *Pseudomonas* flavoprotein xenobiotic reductases. *Journal of Bacteriology* **181**(20), 6254-6263.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & M., S. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**, 365-370.

Bonneau, P. R., Martin, R., Lee, T., Sakowicz, R., Martichonok, V., Hogan, J. K., Gold, M. & Jones, J. B. (1996). Enzymes in organic synthesis. Present and future. *Journal of the Brazilian Chemical Society* **7**(5), 357-369.

Boopathy, R. & Kulpa, C. F. (1994). Biotransformation of 2,4,6-Trinitrotoluene (Tnt) By a *Methanococcus* Sp (Strain-B) isolated from a lake sediment. *Canadian Journal of Microbiology* **40**(4), 273-278.

Bornscheuer, U. T., Altenbuchner, J. & Meyer, H. H. (1998). Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnology and Bioengineering* **58**(5), 554-559.

Brown, B. J., Deng, Z., Karplus, P. A. & Massey, V. (1998). On the active site of Old Yellow Enzyme. Role of histidine 191 and asparagine 194. *Journal of Biological Chemistry* **273**(49), 32753-32762.

Bruhlmann, F. & Chen, W. (1999). Tuning biphenyl dioxygenase for extended substrate specificity. *Biotechnology and Bioengineering* **63**(5), 544-551.

Cadwell, R. G. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR methods and Applications* **2**, 28-33.

Chakrabarty, A. M. & Mongkolsuk, S. (1992). Microbial-degradation of synthetic toxic-Chemicals. *Journal of the Science Society of Thailand* **18**(1), 3-25.

Chen, W., Bruhlmann, F., Richins, R. D. & Mulchandani, A. (1999). Engineering of improved microbes and enzymes for bioremediation. *Current Opinion in Biotechnology* **10**(2), 137-141.

Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A. & Pedersen, A. H. (1999). Directed evolution of a fungal peroxidase. *Nature Biotechnology* **17**(4), 379-384.

Christians, F. C., Scapozza, L., Crameri, A., Folkers, G. & Stemmer, W. P. C. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nature Biotechnology* **17**(3), 259-264.

Cooley, M. (1991). Electroporation of *Escherichia coli*. *Biorad Gene Pulser Electroprotocol*.

Costa, V., Boopathy, R. & Manning, J. (1996). Isolation and characterization of a sulfate-reducing bacterium that removed TNT (2,4,6-trinitrotoluene) under sulfate- and nitrate- reducing conditions. *Bioresource Technology* **56**(2-3), 273-278.

Crameri, A., Dawes, G., Rodriguez, E., Silver, S. & Stemmer, W. P. C. (1997). Molecular evolution of an arsenate detoxification pathway DNA shuffling. *Nature Biotechnology* **15**(5), 436-438.

Crameri, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**(6664), 288-291.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **14**(3), 315-319.

Dafforn, T., Hawarani, A., Moreton, K. & Holbrook, J. J. (1995). Engineering an alteration in substrate-specificity of Ldh By forced evolution. *Protein Engineering* **8**(SS), 5.

Davis, E. P., Boopathy, R. & Manning, J. (1997). Use of trinitrobenzene as a nitrogen source by *Pseudomonas vesicularis* isolated from soil. *Current Microbiology* **34**(3), 192-197.

Di Giulio, M. (1989). Some aspects of the organization and evolution of the genetic code. *Journal of Molecular Evolution* **29**(3), 191-201.

Donnelly, K. C., Chen, J. C., Huebner, H. J., Brown, K. W., Autenrieth, R. L. & Bonner, J. S. (1997). Utility of four strains of white-rot fungi for the detoxification of 2,4,6-trinitrotoluene in liquid culture. *Environmental Toxicology and Chemistry* **16**(6), 1105-1110.

Eilers, A., Rungeling, E., Stundl, U. M. & Gottschalk, G. (1999). Metabolism of 2,4,6-trinitrotoluene by the white-rot fungus *Bjerkandera adusta* DSM 3375 depends on cytochrome P-450. *Applied Microbiology and Biotechnology* **53**(1), 75-80.

Esteve-Nunez, A., Lucchesi, G., Philipp, B., Schink, B. & Ramos, J. L. (2000). Respiration of 2,4,6-trinitrotoluene by *Pseudomonas* sp strain JLR11. *Journal of Bacteriology* **182**(5), 1352-1355.

EsteveNunez, A. & Ramos, J. L. (1998). Metabolism of 2,4,6-trinitrotoluene by *Pseudomonas* sp. JLR11. *Environmental Science & Technology* **32**(23), 3802-3808.

Ewing, D. (1995). The directed evolution of radiation-resistance in *Escherichia-coli*. *Biochemical and Biophysical Research Communications* **216**(2), 549-553.

Fernando, T. & Aust, S. D. (1991). Biodegradation of munition waste, Tnt (2,4,6-Trinitrotoluene), and RDX (Hexahydro-1,3,5-Trinitro-1,3,5-Triazine) by *Phanerochaete- chrysosporium*. *ACS Symposium Series* **468**, 214-232.

Fersht, A. (1999). *Biochemical Society Meeting 671, University of Leeds, UK*.

Fiorella, P. & Spain, J. (1997). Transformation of 2,4,6-trinitrotoluene by *Pseudomonas pseudoalcaligenes* JS52. *Applied and Environmental Microbiology* **63**(5), 2007-2015.

Fishwick, S. (2000). *Collation of toxicological data and development of guideline values for explosive substances*, Environment Agency, Bristol.

Fox, K. M. & Karplus, P. A. (1994). Old Yellow Enzyme at 2 A resolution ± overall structure, ligand-binding, and comparison with related flavoproteins. *Structure* **2**, 1089-1105.

Freeland, S. J. & Hurst, L. D. (1998). The genetic code is one in a million. *Journal of Molecular Evolution* **47**(3), 238-48.

French, C. E. & Bruce, N. C. (1995). Bacterial morphinone reductase is related to Old Yellow Enzyme. *Biochemical Journal* **15**(312), (Pt 3):671-8.

French, C. E., Nicklin, S. & Bruce, N. C. (1996). Sequence and properties of pentaerythritol tetranitrate reductase from *Enterobacter-Cloacae* Pb2. *Journal Of Bacteriology* **178**(22), 6623-6627.

French, C. E., Nicklin, S. & Bruce, N. C. (1998). Aerobic degradation of 2,4,6-trinitrotoluene by *Enterobacter cloacae* PB2 and by pentaerythritol tetranitrate reductase. *Applied and Environmental Microbiology* **64**(8), 2864-2868.

Gabrielson, J., Hart, M., Jarelov, A., Kuhn, I., McKenzie, D. & Mollby, R. (2002). Evaluation of redox indicators and the use of digital scanners and spectrophotometer for quantification of microbial growth in microplates. *Journal of Microbiological Methods* **50**(1), 63-73.

George, S. E., Huggins-Clark, G. & Brooks, L. R. (2001). Use of a Salmonella microsuspension bioassay to detect the mutagenicity of munitions compounds at low concentrations. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis* **490**(1), 45-56.

Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998). Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* **95**(22), 12809-12813.

Goel, A., Kumar, G., Payne, G. F. & Dube, S. K. (1997). Plant cell biodegradation of a xenobiotic nitrate ester, nitroglycerin. *Nature Biotechnology* **15**(2), 174-177.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15), 2714-23.

Gulick, A. M. & Fahl, W. E. (1995). Forced Evolution of glutathione-s-transferase to create a more efficient drug detoxication enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **92**(18), 8140-8144.

Guo. (1997). DNA-binding determinants of sigma 54 as deduced from libraries of mutations. *Journal of Bacteriology* **179**(4), 1239-45.

Hall, B. G. & Malik, H. S. (1998). Determining the evolutionary potential of a gene. *Molecular Biology and Evolution* **15**(8), 1055-1061.

Hannink, N., Rosser, S. J., French, C. E., Basran, A., Murray, J. A., Nicklin, S. & Bruce, N. C. (2001). Phytodetoxification of TNT by transgenic plants expressing a bacterial nitroreductase. *Nature Biotechnology* **19**(12), 1168-7112.

Hansson, L. O., BoltonGrob, R., Massoud, T. & Mannervik, B. (1999). Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. *Journal of Molecular Biology* **287**(2), 265-276.

Harris, J. L. & Craik, C. S. (1998). Engineering enzyme specificity. *Current Opinion in Chemical Biology* **2**(1), 127-132.

Hawari, J., Beaudet, S., Halasz, A., Thiboutot, S. & Ampleman, G. (2000). Microbial degradation of explosives: biotransformation versus mineralization. *Applied Microbiology and Biotechnology* **54**(5), 605-618.

Hughes, J. B., Wang, C. Y., Bhadra, R., Richardson, A., Bennett, G. N. & Rudolph, F. B. (1998). Reduction of 2,4,6-trinitrotoluene by *Clostridium acetobutylicum* through hydroxylamino-nitrotoluene intermediates. *Environmental Toxicology and Chemistry* **17**(3), 343-348.

Invitrogen. (1999). *pTrcHis2 Topo TA Cloning, Cataloge no. K4400-01, K4400-40.*

Iuchi, S., Yamaguchi-Shinozaki, K., Urao, T., Terao, T. & Shinozaki, K. (1996). Novel drought-inducible genes in the highly drought-tolerant cowpea: cloning of cDNAs and analysis of the expression of the corresponding genes. *Plant Cell Physiology* **38**(8), 1073-82.

Jackson, M. M., Hou, L. H., Banerjee, H. N., Sridhar, R. & Dutta, S. K. (1999). Disappearance of 2,4-dinitrotoluene and 2-amino,4,6-dinitrotoluene by *Phanerochaete chrysosporium* under non-ligninolytic conditions. *Bulletin of Environmental Contamination and Toxicology* **62**(4), 390-396.

Johnson, M. S., Holladay, S. D., Lippenholz, K. S., Jenkins, J. L. & McCain, W. C. (2000). Effects of 2,4,6-trinitrotoluene in a holistic environmental exposure regime on a terrestrial salamander, *Ambystoma tigrinum.* *Toxicologic Pathology* **28**(2), 334-341.

Kamei, H., Shimazaki, K. & Nishi, Y. (2001). Computational 3-D modeling and site-directed mutation of an antibody that binds Neu2en5Ac, a transition state analogue of a sialic acid. *Proteins* **45**(4), 285-96.

Kim, H. Y. & Song, H. G. (2000). Comparison of 2,4,6-trinitrotoluene degradation by seven strains of white rot fungi. *Current Microbiology* **41**(5), 317-320.

Kitts, C. L., Green, C. E., Otley, R. A., Alvarez, M. A. & Unkefer, P. J. (2000). Type I nitroreductases in soil enterobacteria reduce TNT (2,4,6-trinitrotoluene) and RDX (hexahydro-1,3,5-trinitro- 1,3,5-triazine). *Canadian Journal of Microbiology* **46**(3), 278-282.

Kohno, M., Enatsu, M., Funatsu, J., Yoshiizumi, M. & Kugimiya, W. (2001). Inprovement of the optimum temperature of lipase activity for *Rhizopus niveus* by random mutagenesis and its structural interpretation. *Journal of Biotechnology* **87**, 201-210.

Kuchner, O. & Arnold, F. H. (1997). Directed evolution of enzyme catalysts. *Trends in Biotechnology* **15**(12), 523-530.

Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T. & Furukawa, K. (1998). Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nature Biotechnology* **16**(7), 663-666.

Lanio, T., Jeltsch, A. & Pingoud, A. (1998). Towards the design of rare cutting restriction endonucleases: Using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude. *Journal of Molecular Biology* **283**(1), 59-69.

Lehmann, J. (2000). Physico-chemical constraints connected with the coding properties of the genetic system. *Journal of Theoretical Biology* **202**(2), 129-44.

Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., Fujii, C. Y., Mason, T., Bowman, C. L., Barnstead, M., Feldblyum, T. V., Buell, C. R., Ketchum, K. A., Lee, J., Ronning, C. M., Koo, H. L., Moffat, K. S., Cronin, L. A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L. J., Gill, J. E. & Venter, J. C. (1999). Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature* **402**(6763), 761-8, 732-1.

Lin-Goerke, J. L., Robbins, D. J. & Burczak, J. D. (1997). PCR-based random mutagenesis using manganese and reduced dntp concentration. *BioTechniques* **23**, 409-412.

Litchfield. (1968). The determination of the di- and mono-nitrates of ethylene glycol and propylene glycol in blood by colorimetric and gas chromatographic methods. *Analyst* **93**, 653-659.

Lönn, A., Gardonyi, M., van Zyl, W., Hahn-Hagerdal, B. & Otero, R. C. (2002). Cold adaptation of xylose isomerase from *Thermus thermophilus* through random PCR mutagenesis. Gene cloning and protein characterization. *European Journal of Biochemistry* **269**(1), 157-63.

Lovering, A. L., Hyde, E. I., Searle, P. F. & White, S. A. (2001). The structure of *Escherichia coli* nitroreductase complexed with nicotinic acid: three crystal forms at 1.7 A, 1.8 A and 2.4 A resolution. *Journal of Molecular Biology* **309**(1), 203-13.

Lucero, M. E., Mueller, W., Hubstenberger, J., Phillips, G. C. & O'Connell, M. A. (1999). Tolerance to nitrogenous explosives and metabolism of TNT by cell suspensions of *Datura innoxia*. *In Vitro Cellular & Developmental Biology-Plant* **35**(6), 480-486.

Luo, L. & Li, X. (2002). Construction of genetic code from evolutionary stability. *Biosystems* **65**(2-3), 83-97.

Marshall, S. J. & White, G. F. (2001). Complete denitration of nitroglycerin by bacteria isolated from a washwater soakaway. *Applied and Environmental Microbiology* **67**(6), 2622-2626.

Martin, J. L., Comfort, S. D., Shea, P. J., Kokjohn, T. A. & Drijber, R. A. (1997). Denitration of 2,4,6-trinitrotoluene by *Pseudomonas savastanoi*. *Canadian Journal of Microbiology* **43**(5), 447-455.

Matsumura, I. & Ellington, A. D. (1999). In vitro evolution of thermostable p53 variants. *Protein Science* **8**(4), 731-740.

Matsuura, T., Miyai, K., Trakulnaleamsai, S., Yomo, T., Shima, Y., Miki, S., Yamamoto, K. & Urabe, I. (1999). Evolutionary molecular engineering by random elongation mutagenesis. *Nature Biotechnology* **17**(1), 58-61.

McPherson, M. (1999). *Biochemical Society Meeting 671, University of Leeds, UK.*

Meah, Y., Brown, B. J., Chakraborty, S. & Massey, V. (2001). Old yellow enzyme: reduction of nitrate esters, glycerin trinitrate, and propylene 1,2-dinitrate. *Proceedings of the National Academy of Sciences of the United States of America* **98**(15), 8560-8565.

Medina, V. F., Larson, S. L., Bergstedt, A. E. & McCutcheon, S. C. (2000). Phyto-removal of trinitrotoluene from water with batch kinetic studies. *Water Research* **34**(10), 2713-2722.

Mee-Hie Cho, C., Mulchandani, A. & Chen, W. (2002). Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. *Applied and Environmental Microbiology* **68**(4), 2026-30.

Meharg, A. A., Dennis, G. R. & Cairney, J. W. G. (1997). Biotransformation of 2,4,6-trinitrotoluene (TNT) by ectomycorrhizal basidiomycetes. *Chemosphere* **35**(3), 513-521.

Meijer, P. J., Holmberg, N., Grundstrom, G. & Bulow, L. (1996). Directed evolution of a type I antifreeze protein expressed in *Escherichia coli* with sodium chloride as selective pressure and its effect on antifreeze tolerance. *Protein Engineering* **9**(11), 1051-1054.

Meng, M., Sun, W. Q., Geelhaar, L. A., Kumar, G., Patel, A. R., Payne, G. F., Speedie, M. K. & Stacy, J. R. (1995). Denitration of glycerol trinitrate by resting cells and cell-extracts of *Bacillus-thuringiensis cereus* and *Enterobacter-agglomerans*. *Applied and Environmental Microbiology* **61**(7), 2548-2553.

Mikes, O., Rychlik, I. & Sorm, F. (1977). Table of amino acid replacements in proteins possible as a result of one-point mutations. *International Journal of Peptide and Protein Research* **10**(2), 176-7.

Millar, T. M., Stevens, C. R., Benjamin, N., Eisenthal, R., Harrison, R. & Blake, D. R. (1998). Xanthine oxidoreductase catalyses the reduction of nitrates and nitrite to nitric oxide under hypoxic conditions. *Febs Letters* **427**(2), 225-228.

Miura, K., Tomioka, Y., Suzuki, H., Yonezawa, M., Hishinuma, T. & Mizugaki, M. (1997). Molecular cloning of the *nemA* gene encoding N-ethylmaleimide reductase from *Escherichia coli*. *Biological & Pharmaceutical Bulletin* **20**(1), 110-112.

Montpas, S., Samson, J., Langlois, E., Lei, J. Y., Piche, Y. & Chenevert, R. (1997). Degradation of 2,4,6-trinitrotoluene by *Serratia marcescens*. *Biotechnology Letters* **19**(3), 291-294.

Moore, J. C. & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology* **14**(4), 458-467.

Oh, B. T., Sarath, G. & Shea, P. J. (2001). TNT nitroreductase from a *Pseudomonas aeruginosa* strain isolated from TNT-contaminated soil. *Soil Biology & Biochemistry* **33**(7-8), 875-881.

Oue, S., Okamoto, A., Yano, T. & Kagamiyama, H. (1999). Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *Journal of Biological Chemistry* **274**(4), 2344-2349.

Pak, J. W., Knoke, K. L., Noguera, D. R., Fox, B. G. & Chambliss, G. H. (2000). Transformation of 2,4,6-trinitrotoluene by purified xenobiotic reductase B from *Pseudomonas fluorescens* I-C. *Applied and Environmental Microbiology* **66**(11), 4742-4750.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* **85**(8), 2444-8.

Proba, K., Worn, A., Honegger, A. & Pluckthun, A. (1998). Antibody scFv fragments without disulfide bonds made by molecular evolution. *Journal Of Molecular Biology* **275**(2), 245-253.

Ramos, J. L., Haidour, A., Duque, E., Pinar, G., Calvo, V. & Oliva, J. M. (1996). Metabolism of nitrate esters by a consortium of two bacteria. *Nature Biotechnology* **14**(3), 320-322.

Reddy, G., Chandra, S. A. M., Lish, J. W. & Qualls, C. W. (2000). Toxicity of 2,4,6-trinitrotoluene (TNT) in hispid cotton rats (Sigmodon hispidus): Hematological, biochemical, and pathological effects. *International Journal of Toxicology* **19**(3), 169-177.

Renoux, A. Y., Sarrazin, M., Hawari, J. & Sunahara, G. I. (2000). Transformation of 2,4,6-trinitrotoluene in soil in the presence of the earthworm Eisenia andrei. *Environmental Toxicology and Chemistry* **19**(6), 1473-1480.

Riefler, R. G. & Smets, B. F. (2002). NAD(P)H:flavin mononucleotide oxidoreductase inactivation during 2,4,6- trinitrotoluene reduction. *Applied and Environmental Microbiology* **68**(4), 1690-1696.

Riegert. (2001). Altering catalytic properties of 3-chlorocatechol-oxidizing extradiol dioxygenase from Sphingomonas xenophaga BN6 by random mutagenesis. *Journal of Bacteriology* **183**(7), 2322-30.

Robidoux, P. Y., Hawari, J., Thiboutot, S., Ampleman, G. & Sunahara, G. I. (2001). Chronic toxicity of octahydro-1,3,5,7-tetranitro-1,3,5,7- tetrazocine (HMX) in soil determined using the earthworm (Eisenia andrei) reproduction test. *Environmental Pollution* **111**(2), 283-292.

Robidoux, P. Y., Svendsen, C., Caumartin, J., Hawari, J., Ampleman, G., Thiboutot, S., Weeks, J. M. & Sunahara, G. I. (2000). Chronic toxicity of energetic compounds in soil determined using the earthworm (Eisenia andrei) reproduction test. *Environmental Toxicology and Chemistry* **19**(7), 1764-1773.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning*, 1,2,3, Cold Spring Harbour Laboratory Press.

Samson, J., Langlois, E., Lei, J., Piche, Y. & Chenevert, R. (1998). Removal of 2,4,6-trinitrotoluene and 2,4-dinitrotoluene by fungi (Ceratocystis coerulescens, Lentinus lepideus and Trichoderma harzianum). *Biotechnology Letters* **20**(4), 355-358.

Schaller, F., Biesgen, C., Mussig, C., Altmann, T. & Weiler, E. W. (2000). 12-Oxophytodienoate reductase 3 (OPR3) is the isoenzyme involved in jasmonate biosynthesis. *Planta* **210**(6), 979-84.

Schlegel, H. G. & Schmidt, K. (1990). Microorganisms and the environment. Trans. Press, C. U. In *General Microbiology* 6 edit., pp. 517,518.

Schuster, P. (1997). Landscapes and molecular evolution. *Physica D* **107**, 351-365.

Sella, G. & Ardell, D. H. (2002). The impact of message mutation on the fitness of a genetic code. *Journal of Molecular Biology* **54**, 638-651.

Servent, D., Ducrocq, C., Henry, Y., Guissani, A. & Lenfant, M. (1991). Nitroglycerin metabolism by *Phanerochaete-chrysosporium* - evidence for nitric-oxide and nitrite formation. *Biochimica Et Biophysica Acta* **1074**(2), 320-325.

Shafikhani, S., Siegel, R. A., 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.

Shelley, M. D., Autenrieth, R. L., Wild, J. R. & Dale, B. E. (1996). Thermodynamic analysis of trinitrotoluene biodegradation and mineralization pathways. *Biotechnology and Bioengineering* **51**(2), 198-205.

Siciliano, S. D., Gong, P., Sunahara, G. I. & Greer, C. W. (2000). Assessment of 2,4,6-trinitrotoluene toxicity in field soils by pollution-induced community tolerance, denaturing gradient gel electrophoresis, and seed germination assay. *Environmental Toxicology and Chemistry* **19**(8), 2154-2160.

Snape, J. R., Walkley, N. A., Morby, A. P., Nicklin, S. & White, G. F. (1997). Purification, properties, and sequence of glycerol trinitrate reductase from *Agrobacterium radiobacter*. *Journal Of Bacteriology* **179**(24), 7796-7802.

Spain, J. C., Hughes, J. B. & Knackmuss, H. J. (2000). *Biodegradation of nitroaromatic compounds and explosives*, Lews Publishers.

Stemmer, W. P. C. (1994a). DNA shuffling by random fragmentation and reassembly - In-vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**(22), 10747-10751.

Stemmer, W. P. C. (1994b). Rapid evolution of a protein in-vitro by DNA shuffling. *Nature* **370**(6488), 389-391.

Sun, W. H., Horst, G. L., Drijber, R. A. & Elthon, T. E. (2000). Fate of 2,4,6-trinitrotoluene in axenic sand culture systems containing smooth bromegrass. *Environmental Toxicology and Chemistry* **19**(8), 2038-2046.

Svendsen, A., Clausen, I. G., Patkar, S. A., Borch, K. & Thellersen, M. (1997). Protein engineering of microbial lipases of industrial interest. *Methods in Enzymology* **284**, 317-340.

Svetlov, V. & Cooper, T. G. (1998). Efficient PCR-based random mutagenesis of sub-genic (100 bp) DNA fragments. *Yeast* **14**(1), 89-91.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**(22), 4673-80.

Tope, A. M. & Jamil, K. (2001). Isolation of TNT tolerant *Pseudomonas* species (Strain KA) from TNT contaminated soil-biotransformation of TNT. *Defence Science Journal* **51**(2), 141-146.

Trakulnaleamsai, S., Yomo, T., Yoshikawa, M., Aihara, S. & Urabe, I. (1995). Experimental sketch of landscapes in protein-sequence space. *Journal of Fermentation and Bioengineering* **79**(2), 107-118.

Trinquier, G. & Sanejouand, Y. H. (1998). Which effective property of amino acids is best preserved by the genetic code? *Protein Engineering* **11**(3), 153-69.

Turner, W. B. (1980). Micriobial products, Nitro compounds. In *Handbook of Microbiology* (Laskin & Lechevalier, eds.), Vol. 3, pp. 445-447.

Uhrig, J. F. & Schwenn, J. D. (1997). Cloning of an old-yellow-enzyme homolog from a sulfate starved *Catharanthus roseus* cell suspension culture. *EMBL submission O233777.*

Van Aken, B., Cameron, M. D., Stahl, J. D., Plumat, A., Naveau, H., Aust, S. D. & Agathos, S. N. (2000). Glutathione-mediated mineralization of C-14-labeled 2-amino- 4,6-dinitrotoluene by manganese-dependent peroxidase H5 from the white-rot fungus *Phanerochaete chrysosporium*. *Applied Microbiology and Biotechnology* **54**(5), 659-664.

Van Aken, B., Hofrichter, M., Scheibner, K., Hatakka, A. I., Naveau, H. & Agathos, S. N. (1999). Transformation and mineralization of 2,4,6-trinitrotoluene (TNT) by manganese peroxidase from the white-rot basidiomycete *Phlebia radiata*. *Biodegradation* **10**(2), 83-91.

Vanderberg, L. A., Perry, J. J. & Unkefer, P. J. (1995). Catabolism of 2,4,6-trinitrotoluene By *Mycobacterium-Vaccae*. *Applied Microbiology and Biotechnology* **43**(5), 937-945.

Vartanian, J. P., Henry, M. & WainHobson, S. (1996). Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Research* **24**(14), 2627-2631.

Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z. G. (2001). Computationally focusing the directed evolution of proteins. *Journal of Cellular Biochemistry* **Supplement 37**, 58-63.

Vorbeck, C., Lenke, H., Fischer, P., Spain, J. C. & Knackmuss, H. J. (1998). Initial reductive reactions in aerobic microbial metabolism of 2,4,6- trinitrotoluene. *Applied and Environmental Microbiology* **64**(1), 246-252.

Wackett, L. P. (1998). Directed evolution of new enzymes and pathways for environmental biocatalysis. *Annals of the New York Academy of Sciences* **864**, 142-152.

White, G. F., Snape, J. R. & Nicklin, S. (1996). Bacterial biodegradation of glycerol trinitrate. *International Biodeterioration & Biodegradation* **38**(2), 77-82.

Wikstrom, P., Andersson, A. C., Nygren, Y., Sjostrom, J. & Forsman, M. (2000). Influence of TNT transformation on microbial community structure in four different lake microcosms. *Journal of Applied Microbiology* **89**(2), 302-308.

Williams, R. W. & Bruce, N. C. (2002). New uses for an old Enzyme' – the Old Yellow Enzyme family of flavoenzymes. *Microbiology* **148**, 1607–1614.

Xu, H., Petersen, E. I., Petersen, S. B. & el-Gewely, M. R. (1999). Random mutagenesis libraries: optimization and simplification by PCR. *BioTechniques* **27**(6), 1106-1108.

Yano, T., Oue, S. & Kagamiyama, H. (1998). Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proceedings of the National Academy of Sciences of the United States of America* **95**(10), 5511-5515.

Yinon, J. (1990). *Toxicity and metabolism of explosives*, CRC press, Florida.

Yinon, J. & Zitrin, S. (1993). *Modern methods and applications in analysis of explosives*, John wiley & sons, West Sussex.

Yokobayashi, Y., Ikebukuro, K., McNiven, S. & Karube, I. (1996). Directed evolution of trypsin inhibiting peptides using a genetic algorithm. *Journal of the Chemical Society-Perkin Transactions 1*(20), 2435-2437.

You, L. & Arnold, F. H. (1996). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Engineering* **9**(8), 78.

Zaccolo, M. & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: A study on TEM-1 beta-lactamase. *Journal of Molecular Biology* **285**(2), 775-783.

Zenno, S., Koike, H., Kumar, A. N., Jayaraman, R., Tanokura, M. & Saigo, K. (1996a). Biochemical characterization of NfsA, the *Escherichia coli* major nitroreductase exhibiting a high amino acid sequence homology to Frp, a *Vibrio harveyi* flavin oxidoreductase. *Journal of Bacteriology* **178**(15), 4508-4514.

Zenno, S., Koike, H., Tanokura, M. & Saigo, K. (1996b). Gene cloning, purification, and characterization of NfsB, a minor oxygen-insensitive nitroreductase from *Escherichia coli*, similar in biochemical properties to FRase I, the major flavin reductase in Vibrio fischeri. *Journal of Biological Chemistry* **120**(4), 736-44.

Zhang, J. H., Dawes, G. & Stemmer, W. P. C. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proceedings of the National Academy of Sciences of the United States of America* **94**(9), 4504-4509.

Zhang, Y. Z., Sundaram, S. T., Sharma, A. & Brodman, B. W. (1997b). Biodegradation of glyceryl trinitrate by *Penicillium corylophilum* Dierckx. *Applied and Environmental Microbiology* **63**(5), 1712-1714.

Zhao, H. M. & Arnold, F. H. (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Research* **25**(6), 1307-1308.

Appendix A. The *nerA* gene

Translation of *nerA*

The nucleotide sequence, as well as the amino acid sequence of NerA are displayed below. Boxes indicate what regions align to the secondary structures that are found in Oye (See Table 7.10 and Figure 7.13). The alterations that were found in various mutants of *nerA* are indicated below the DNA sequence. The name of the mutant and the effect on the amino acid sequence are displayed nearby. Four of the important residues that are thought to be involved in the catalytic action of NerA are displayed in reversed colour. The corresponding position of these important residues in Old Yellow Enzyme is also displayed ("OYE" followed by position).

```
                                        LID1                    LID2
1     Met Thr Ser Leu Phe Glu Pro Ala Gln Ala Gly Asp Ile Ala Leu   15 Amino acid
1     ATG ACC AGT CTT TTC GAA CCG GCA CAG GCC GGC GAT ATC GCA CTC   45 Base pair
                  C PM4   Silent

              β0                              Pro
16    Ala Asn Arg Ile Val Met Ala Pro Leu Thr Arg Asn Arg Ser Pro   30
46    GCC AAC CGC ATC GTC ATG GCT CCT CTC ACC CGC AAT CGC TCC CCG   90
                              A PM11  Silent          T PM3  Silent
                                      C PM15

                                    α1
31    Gly Ala Ile Pro Asn Asn Leu Asn Ala Thr Tyr Tyr Glu Gln Arg   45
91    GGC GCC ATT CCC AAC AAT CTC AAC GCC ACC TAT TAC GAG CAG CGG   135

      Thr         β1
46    Ala Thr Ala Gly Leu Ile Val Thr Glu Gly Thr Pro Ile Ser Gln   60
136   GCG ACA GCC GGC CTG ATC GTC ACT GAA GGC ACG CCC ATT TCC CAG   180
      A NM4
        A NM5 Silent
```

```
                                                                  α2  Val
61    Gln Gly Gln Gly Tyr Ala Asp Val Pro Gly Leu Tyr Lys Arg Glu   75
181   CAG GGC CAG GGT TAC GCC GAT GTT CCC GGT CTT TAT AAG CGG GAA  225
                                      G PM13 Silent          T NM2
```

```
                                                          Tyr| α2
76    Ala Ile Glu Gly Trp Lys Lys Ile Thr Asp Gly Val His Ser Ala   90
226   GCG ATT GAA GGC TGG AAA AAG ATC ACT GAC GGC GTT CAT TCG GCG   270
                                                  T NM2 A PM11  Silent
```

```
            β2
91    Gly Gly Lys Ile Val Ala Gln Ile Trp His Val Gly Arg Ile Ser   105
271   GGC GGC AAG ATC GTT GCC CAA ATC TGG CAC GTG GGC CGA ATT TCC   315
        T PM8   Silent
```

```
                                          Ala
106   His Thr Ser Leu Gln Pro His Gly Gly Gln Pro Val Ala Pro Ser   120
316   CAC ACC TCG TTG CAG CCG CAT GGT GGC CAG CCG GTC GCC CCC TCG   360
                                  A PM14 Silent
                                        A PM11 Silent
                                          C NM5
```

```
121   Ala Ile Thr Ala Lys Ser Lys Thr Tyr Ile Ile Asn Asp Asp Gly   135
361   GCC ATC ACC GCC AAA TCG AAG ACC TAT ATC ATC AAT GAT GAC GGC   405
```

```
                        Asp                                  α3
136   Thr Gly Ala Phe Ala Glu Thr Ser Glu Pro Arg Ala Leu Thr Ile   150
406   ACC GGT GCC TTT GCC GAA ACC TCC GAG CCC CGT GCC CTC ACC ATC   450
                          C PM11                  T NM5 Silent
```

```
                Ser Ile                              Arg Thr
151   Asp Asp Ile Gly Leu Ile Leu Glu Asp Tyr Arg Ser Gly Ala Arg   165
451   GAC GAT ATC GGC CTC ATC CTC GAG GAT TAC CGC AGT GGC GCG CGC   495
                    A PM4                            C PM14
                A NM1                                  A PM4
```

α3 β3 OYE191

166 Ala Ala Leu Glu Ala Gly Phe Asp Gly Val Glu Ile His Ala Ala 180

496 GCG GCA CTG GAG GCC GGT TTC GAC GGC GTC GAA ATC CAT GCC GCC 540

OYE194 OYE196

181 Asn Gly Tyr Leu Ile Glu Gln Phe Leu Lys Ser Ser Thr Asn Gln 195

541 AAT GGT TAC CTG ATC GAG CAG TTC CTG AAA TCC AGC ACC AAC CAG 585

T NM2 Silent

α4 Val

196 Arg Thr Asp Asp Tyr Gly Gly Ser Ile Glu Asn Arg Ala Arg Phe 210

586 CGC ACG GAT GAT TAC GGT GGC TCC ATC GAA AAC CGC GCT CGC TTC 630

A NM5 C All Silent T PM9

Silent

α4

211 Leu Leu Glu Val Val Asp Ala Val Ala Glu Glu Ile Gly Ala Gly 225

631 CTG CTG GAA GTC GTC GAC GCG GTG GCG GAA GAG ATC GGC GCC GGC 675

T PM3 Silent T NM3 Silent

D NM6 Frame shift

β4 Ser

226 Arg Thr Gly Ile Arg Leu Ser Pro Val Thr Pro Ala Asn Asp Ile 240

676 CGC ACC GGC ATC CGC CTC TCG CCA GTT ACC CCT GCC AAC GAT ATT 720

T NM5

α5

241 Phe Glu Ala Asp Pro Gln Pro Leu Tyr Asn Tyr Val Val Glu Gln 255

721 TTC GAG GCT GAT CCG CAA CCG CTC TAT AAT TAC GTC GTC GAG CAG 765

α5 β5

256 Leu Gly Lys Arg Asn Leu Ala Phe Ile His Val Val Glu Gly Ala 270

766 CTG GGC AAG CGC AAT CTC GCC TTC ATC CAT GTC GTT GAA GGT GCA 810

```
271   Thr Gly Gly Pro Arg Asp Phe Lys Gln Gly Asp Lys Pro Phe Asp   285
811   ACG GGC GGT CCG CGC GAT TTC AAG CAG GGC GAC AAG CCT TTC GAT   855
                                        T NM4 Silent

                      Arg      Val  α6|          Thr          β6|
286   Tyr Ala Ser Phe [Lys Ala Ala Tyr] Arg Asn Ala Gly Gly [Lys Gly]  300
856   TAC GCG TCC TTC AAG GCG GCT TAT CGT AAT GCC GGC GGC AAG GGC   900
    Silent A PM7       G PM14        C PM4 Silent
                                T NM3          A NM5

            Asn                              |α7
301   [Leu Trp Ile] Ala Asn Asn Gly Tyr Asp [Arg Gln Ser Ala Ile Glu]  315
901   CTG TGG ATC GCC AAC AAT GGT TAC GAC AGG CAG AGC GCC ATC GAA   945
      A PM12  Silent          T PM13 Silent
              A NM3

                                Ala                    β7|     Arg
316   Ala Val Glu Ser Gly Lys [Val Asp Ala Val Ala Phe] Gly Lys [Ala]  330
946   GCC GTG GAA AGT GGC AAG GTT GAT GCC GTT GCT TTC GGG AAG GCC   990
                                C All                          G PM14


331   Phe Ile Ala Asn Pro Asp Leu Val Arg Arg Leu Lys Asn Asp Ala   345
991   TTC ATC GCC AAT CCT GAT CTG GTG CGC CGC CTG AAA AAC GAC GCG   1035
                A All silent                      G NM2 Frame shift


                                Leu          OYE375
346   Pro Leu Asn Ala Pro Asn Gln Pro Thr Phe Tyr Gly Gly Gly Ala   360
1036  CCG CTG AAC GCA CCG AAC CAG CCG ACC TTC TAT GGC GGC GGC GCC   1080
                                T PM11                      T PM4   Silent

                                    Ser
361   Glu Gly Tyr Thr Asp Tyr Pro Ala Leu Ala Gln End
1081  GAA GGT TAC ACC GAC TAT CCG GCC CTC GCC CAA TAG   1116
                T NM1  Silent     T PM9
      D   A NM3
```

Mutants of *nerA* by name

23 clones obtained from two library screening experiments were sequenced and analyzed for single mutations.

Table A.1. Tolerated Mutations

Mutant name	Mutations	Silent mutations
All	Val322 to Ala	2
PM1		-
PM2		-
PM3		2
PM4	Leu155 to Ile, Ala 164 to Thr	3
PM5		-
PM6		-
PM7		1
PM8		1
PM9	Glu205 to Val, Ala368 to Ser	-
PM10		-
PM11	Glu141 to Asp, Gln352 to Leu	2
PM12		1
PM13		3
PM14	Gly163 to Arg, Lys290 to Arg, Lys329 to Arg	1
PM15	Thr25 to Pro	-

Two additional mutants were analysed, but they were found to be identical to PM13 and PM14 (see Section 7.1.6).

Table A.2. Knock-out Mutations

Mutant name	Mutations
NM1	154 Gly to Ser,
NM2	75 Glu to Val, 88 His to Tyr Insert G @bp1024, Frame shift after 341aa
NM3	292 Ala to Val, 303 Ile to Asn, Deletion @bp1081
NM4	46 Thr to Ala
NM5	117 Ala to Val, 236 Pro to Ser, 296 Ala to Thr
NM6	Deletion @bp637, Frame shift after 212aa

Appendix B. Media and Buffers

Media

Anaerobic selection medium: Phosphate buffer (30 mM), Salt stock (1x, see below), Glycerol (25 mM), Fumarate (25 mM), Rezasurin (2.5 mg/l), Na_2S $6H_2O$ (0.4 g/l, added freshly), NH_4Cl (50 μM) Nitro organic compound or Nitrite (0.2 mM).

LB medium: Trypton (10 g/l), Yeast extract (5 g/l), NaCl (5 g/l)

2xLB medium: Trypton (20 g/l), Yeast extract (10 g/l), NaCl (5 g/l, Salt concentration is not doubled)

LB and 2xLB medium were supplemented with Ampicillin (50 mg/l) as required

Salt stock (10x): NaCl (5 g/l), MgSO4 (1.2 g/l), trace amount of B, Fe, Co, Cu, Mn, Zn, Mo.

Nitrogen free medium: Phosphate buffer (30mM), Salt stock (1 x), Glycerol (1% v/v), Nitro organic compound or Nitrite (0.2 mM).

SOB: Sodium chloride (0.5 g/L), Tryptone (20 g/L), Yeast extract (5 g/L)

SOC: Sodium chloride (0.5 g/L), Tryptone (20 g/L), Yeast extract (5 g/L), Glucose (25 mM)

Buffers

Electroporation buffer: Glycerol (10% v/v) Hepes (1 mM, Ph7). Hepes stock solution (100 mM) was titrated to Ph7 using acetic acid (Cooley, 1991).

Gel filtration buffer: KP50 supplemented with NaCl (100 mM).

KP50: Potassium Phosphate buffer (50 mM, Ph7).

Ligation buffer (1x): Tris HCL (50 mM pH 7.5), $MgCl_2$ (10 mM), dithiothreitol (10 mM), ATP (1 mM), BSA (25 μg/ml)

PCR Reaction buffer (1x) : Tris/HCl (10 mM, pH = 9), KCl (50 mM), Triton X-100 (0.1% w/v).

Reagent A: sulphanilic acid (6 mM) in 60% v/v Acetic Acid.

Reagent B: naphthylethyldiamine (4 mM) in 30% v/v Acetic Acid.

TK buffer: Tris /Hcl (Ph8, 33 mM), KCl (100 mM).

Stock solutions

Substance	Stock
Ampicillin	5% w/v
Carbenicillin	10% w/v
IPTG	1M

Appendix C Program to generate table3.2

In MS Quickbasic

```
DECLARE SUB writetbl ()
DECLARE FUNCTION class! (bp1, bp2 AS INTEGER)
DECLARE SUB main ()
DECLARE SUB mutatecodon (bp1, bp2, bp3 AS INTEGER)
DECLARE SUB main2 ()
DECLARE SUB origen ()
DECLARE SUB compare (aa1 AS STRING, aa2 AS STRING)
DECLARE SUB main1 ()
DECLARE SUB loadaaname ()
DECLARE SUB loadaaproperties ()
DECLARE SUB loaddata ()
DECLARE SUB loadcodondata ()
DECLARE SUB laodaaproperties ()
DECLARE SUB loadcodon ()

aadata: ' list of one-leter code abbreviation for amino acids. 1 =
stopcodon amber, 2 = ochre, 3 = opal
DATA A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y,1,2,3
codondata: 'list of 64 codons; UUU=0 (first on the list), last basepair
=0..3, middle basepair=0..3*4, first basepair=0..3*16
DATA
5,5,10,10,16,16,16,16,20,20,22,21,2,2,23,19,10,10,10,10,13,13,13,13,7,7,14
,14,15,15,15,15,8,8,8,11,17,17,17,17,12,12,9,9,16,16,15,15,18,18,18,18,1,1
,1,1,3,3,4,4,6,6,6,6


DIM SHARED codon(4, 4, 4) AS INTEGER    ' Codon (basepair1, basepair2,
basepair3) carries reference number to aminoacid.
DIM SHARED nuc(4) AS STRING * 1
nuc(1) = "U": nuc(2) = "C": nuc(3) = "A": nuc(4) = "G"' define nucleotides
DIM SHARED aa(23) AS STRING * 1   '20 aminoacids, 3=stopcodons
DIM SHARED tbl(23, 23) AS INTEGER 'Table to compare all mutagenic events.
' In the above table all mutagenic classed are stored in a binary fashion
by assigning exponents of 2
' to each class. Class1 = 2, Class2=4, Class3=8 etc..

loaddata' prepare arrays codon and aa

OPEN "m64x9.dat" FOR OUTPUT AS #1'matrix 64x9
main'evaluate all 64*9 single mutation events
CLOSE

writetbl

FUNCTION class (bp1, bp2 AS INTEGER)

IF bp1 = bp2 THEN
        PRINT "ERROR in function CLASS: bp1=bp2"
        END
END IF
```

```
SELECT CASE bp1
        CASE 1
                SELECT CASE bp2
                CASE 2' U to C
                        class = 1
                CASE 3' U to A
                        class = 3
                CASE 4' U to G
                        class = 5
                END SELECT
        CASE 2
                SELECT CASE bp2
                CASE 1' C to U
                        class = 2
                CASE 3' C to A
                        class = 6
                CASE 4' C to G
                        class = 4
                END SELECT
        CASE 3
                SELECT CASE bp2
                CASE 1' A to U
                        class = 3
                CASE 2' A to C
                        class = 5
                CASE 4' A to G
                        class = 1
                END SELECT
        CASE 4
                SELECT CASE bp2
                CASE 1' G to U
                        class = 6
                CASE 2' G to C
                        class = 4
                CASE 3' G to A
                        class = 2
                END SELECT
        CASE ELSE
        PRINT "ERROR in function CLASS: bp overflow"
        END SELECT

END FUNCTION

SUB compare (aa1 AS STRING, aa2 AS STRING)
IF aa1 = aa2 THEN PRINT #1, "= silent": EXIT SUB' silent mutation
PRINT #1, "= "; aa1; "-"; aa2
END SUB

SUB loadaaname ' load the names of the aminoacids

RESTORE aadata' list with names

FOR i = 1 TO 23
        READ n$
        aa(i) = n$
NEXT
END SUB

SUB loadcodondata
RESTORE codondata
```

```
FOR bp1 = 1 TO 4: FOR bp2 = 1 TO 4: FOR bp3 = 1 TO 4
        READ aa
        codon(bp1, bp2, bp3) = aa
NEXT: NEXT: NEXT
END SUB

SUB loaddata
loadaaname'fill array with one-letter aa-name
loadcodondata' fill array codon with aa-number
END SUB

SUB main
DIM bp1, bp2, bp3 AS INTEGER

FOR bp1 = 1 TO 4
'*****************************************************
  FOR bp2 = 1 TO 4       '* These three nested loops run trough all 64
codons *
    FOR bp3 = 1 TO 4
'*****************************************************
                mutatecodon bp1, bp2, bp3' mutate that particular codon
NEXT: NEXT: NEXT
END SUB

SUB mutatecodon (bp1, bp2, bp3 AS INTEGER)
DIM bp(3)
DIM mut(3) AS INTEGER
DIM originalcodon AS INTEGER
DIM OriginalAA AS STRING * 1
DIM newtbl AS INTEGER
bp(1) = bp1: bp(2) = bp2: bp(3) = bp3' convert three loose values into
array.
originalcodon = codon(bp1, bp2, bp3) ' Newly mutated codon must always be
compared to original codon.
OriginalAA = aa(originalcodon)
'##

FOR position = 1 TO 3 ' run through all three positions in the codon
mut(1) = bp(1): mut(2) = bp(2): mut(3) = bp(3)' reset codon before next
mutation

        FOR mutation = 1 TO 3' mutate every position three times
                mut(position) = 1 + ((bp(position) + mutation - 1) MOD 4)
                ' MOD command makes the cycle A,C,T,G endless (A comes
after G etc)
                ' change one basepair in the codon. (mut(position)
                singlemutant = codon(mut(1), mut(2), mut(3))
                ' the other two mut() have not been changed
                PRINT #1, nuc(bp(position)); "-"; nuc(mut(position)); " ";
' write mutagenic event

                newtbl = tbl(originalcodon, singlemutant) OR (2 ^
class(bp(position), mut(position)))' binary addition
                ' the bit that represents a cetrain class inside the tbl
variabel is set high
                ' using the OR statement. If it was already high, nothing
changes.
                tbl(originalcodon, singlemutant) = newtbl

                compare OriginalAA, aa(singlemutant)'
        NEXT
```

```
NEXT
END SUB

SUB writetbl
OPEN "Table.dat" FOR OUTPUT AS #1
' Write class - table to file

FOR i = 1 TO 23
FOR j = 1 TO 23

a$ = ""
IF tbl(i, j) = 0 OR i = j THEN
        a$ = "------|"
ELSE
        FOR k = 1 TO 6
         IF tbl(i, j) AND 2 ^ k THEN ' binary "ONE" detected
                a$ = a$ + CHR$(48 + k)
         ELSE
                a$ = a$ + "+"
         END IF

        NEXT k
        a$ = a$ + "|"

END IF

PRINT #1, a$;

NEXT j
PRINT #1,
NEXT i

CLOSE
END SUB
```