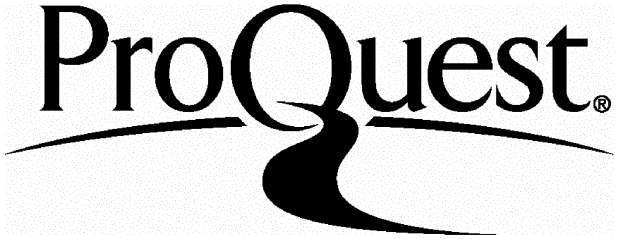
STRUCTURAL STUDIES OF A URACIL-DNA GLYCOSYLASE FROM HERPES SIMPLEX VIRUS TYPE 1

by

Renos Savva

A Thesis Submitted
in the
University of London
For the degree of
Doctor of Philosophy.

Department of Biochemistry and
Molecular Biology
University College
September 1994



ABSTRACT

The work presented in this thesis describes experiments carried out in order to determine the three-dimensional structure of a DNA repair enzyme, uracil-DNA glycosylase. An open reading frame, UL2, in the herpes simplex virus type 1 genome, is known to encode a uracil-DNA glycosylase. By sequence homology, there are three candidate start codons which might express a functional uracil-DNA glycosylase. Expression from two of these was attempted in *Escherichia coli*, using plasmids designed for high level production of recombinant proteins. The second candidate start codon produces high levels of a soluble, functional uracil-DNA glycosylase in *Escherichia coli*, both in a native form, and as part of a fusion protein. Both the fusion and the native form of the enzyme have been purified to apparent homogeneity, as has a recombinantly expressed insoluble *Escherichia coli* uracil-DNA glycosylase.

Preliminary attempts were made at deriving structural and functional information from the soluble, native recombinant herpes simplex enzyme with the use of circular dichroism. This form of uracil-DNA glycosylase has subsequently been crystallised in two ways, firstly as the free enzyme, and secondly in a complex with a single stranded DNA oligonucleotide. Extensive optimisation of the crystallisation parameters have been carried out in conjunction with modifications to the original purification protocol, and large, single crystals of both free, and DNA bound forms, suitable for X-ray diffraction studies are now readily reproduced. A systematic search for isomorphous heavy atom derivatives has been carried out for both types of crystal, and preliminary phases have been obtained for the DNA-bound form of the enzyme. This has enabled the calculation of an electron density map in which protein secondary structure features can be located. Improvement of this map will reveal the molecular structure of the enzyme/ DNA complex.

DECLARATION AND ACKNOWLEDGEMENTS

The work presented in this thesis is my own, and was carried out unaided except for the helpful assistance of those mentioned below. This project was financially supported by a SERC CASE studentship in collaboration with The Wellcome Foundation Plc.

Firstly I would like to thank my supervisors in this project.

Dr. Laurence H. Pearl for his unorthodox [almost exclusively caffeine-driven] brand of supervision and encouragement throughout, his efforts to teach me crystallography, and his excellent foresight in designing this project. The processing of much of the crystallographic data, particularly in the early stages of this work was carried out by Laurence. The processing of isomorphous heavy atom derivative data to obtain phases, and the initial refinement of the electron density map was also carried out by Laurence.

Dr. Eddie Littler, my SERC CASE supervisor, for his excellent help in developing my skills in molecular biology, and much helpful discussion at critical points in this project. Attempts to raise antibody sera against the purified recombinantly expressed enzyme produced in this project were carried out by Eddie.

I would also like to thank the following for their contributions to this project.

Dr. Duncan J. McGeogh for the kind gift of pGX23, without which . . .

Dr. Nicholas F. Totty for the N-terminal sequence analysis work.

Dr. Katherine McAuley-Hecht, and Dr. Tom Brown, for the synthesis of oligonucleotides used very successfully in co-crystallisation experiments, and phase determination.

Dr. David T. Jones for running the *optimal sequence threading* program.

Thanks are due to many members of the department during the course of this project for making helpful and timely suggestions, and for much helpful discussion, particularly Dr. John Ward, Dr. Chris Taylorson, and all those in the laboratory [full and part time]. I would also like to thank all who provided friendship along the way during the term of this project. Thanks again to Laurence Pearl and Eddie Littler for their critical proofreading of this manuscript during its preparation, as well as to Ian Sanderson and Vanessa Wheeler for the same. Finally an extra big thank you to Vanessa for help beyond the call of duty in the final editing of this work.

Lastly, and most importantly, I would like to thank my family [past, present, future] for always being there.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

The following abbreviations are used in the text. Standard abbreviations for the DNA bases, amino acids, and chemical elements are used throughout, as are commonly used units of measurement:

A_{nnn}	Absorbance (subscript figure indicates wavelength in nanometres)
Ada	*E.coli* O^6-methylguanine-DNA methyltransferase
AlkA	*E.coli* 3-methyladenine-DNA glycosylase II
AP	Apurinic or apyrimidinic
bp	base pairs of DNA
BSA	Bovine Serum Albumin
CD	Circular dichroism
CM-cellulose	Carboxy-methyl cellulose ion-exchanger
CIP	Calf intestinal phosphatase
CPM	Counts per minute [radioactivity]
CuKα	Copper K-alpha radiation.
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytidine 5'-triphosphate
DEAE	Diethylaminoethyl [DE52 cellulose ion-exchanger in this thesis]
ddNTP	Dideoxynucleotide triphosphate
dGTP	2'-Deoxyguanosine 5'-triphosphate
dITP	2'-Deoxyinosine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
5'dRpase	5'-deoxyribophosphodiesterase
dTMP	2'-Deoxythymidine 5'-monophosphate
dTTP	2'-Deoxythymidine 5'-triphosphate
DTT	Dithiothreitol
dUMP	2'-Deoxyuridine 5'-monophosphate
dUTP	2'-Deoxyuridine 5'-triphosphate
dUTPase	Deoxyuridine triphosphatase
EDTA	Ethylenediaminetetraacetic acid

EGTA	Ethylene glycol-bis(β-aminoethyl ether)
f	Atomic scattering factor
F	Atomic structure factor
FAST	Fast scanning area sensitive television detector
FPG	Formamidopyrimidine-DNA glycosylase
G3PDase	Glyceraldehyde-3-phosphate dehydrogenase
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
HPLC	High Performance Liquid Chromatography
HSV1	Herpes simplex virus type 1
HSV2	Herpes simplex virus type 2
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
kb	kilobase pairs of DNA
MES	2-[N-Morpholino]ethanesulphonic acid
MNNG	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PIPES	Piperazine-N, N'-bis[2-ethanesulphonic acid]
PMSF	Phenylmethanesulphonyl fluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SSarc	Saline sodium citrate/ sarcosine
SSC	Saline sodium citrate
TAE	Tris-acetate/ EDTA electrophoresis buffer
Tag	*E.coli* 3-methyladenine-DNA glycosylase I
TBE	Tris-borate/ EDTA electrophoresis buffer
TE	Tris-EDTA buffer
TEMED	*N*, *N*, *N'*, *N'*-Tetramethylethylenediamine
T_m	DNA melting/ annealing temperature
Tris·HCl	Tris[hydroxymethyl]aminomethane hydrochloride
UDGase	Uracil-DNA glycosylase

UGI	*B.subtilis* bacteriophage PBS1 or PBS2 encoded specific protein inhibitor of UDGase
U_L	Long Unique region of the HSV1 genome
U_S	Short unique region of the HSV1 genome
UL1	Open reading frame number 1, of the Long Unique region of the HSV1 genome
UL2	Open reading frame number 2, of the Long Unique region of the HSV1 genome
UV	Ultra-violet
w x l	Width by length (dimensions of crystals)
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

1 INTRODUCTION

Deoxyribonucleic acid (DNA), the genetic template contained within all cellular organisms, and many viruses, is highly susceptible to chemical alteration by a host of reactions which readily occur in the physiological environment. In order to maintain the fidelity of the genetic message stored as DNA, a plethora of repair systems exist to constantly restore the molecule to its normal state, as a double stranded polymer of four nucleotides, thymidylic acid, cytidylic acid, guanylic acid and adenylic acid. These processes are reviewed in section **1.1**. Sections **1.2** and **1.3** are concerned primarily with the DNA repair enzyme uracil-DNA glycosylase, its function *in vivo*, and the possible role of this enzyme in latency and reactivation of herpes simplex virus type 1 (HSV1).

1.1 DNA DAMAGE: MECHANISMS, CONSEQUENCES AND REPAIR

Deficiencies in DNA repair have been implicated as the causes of a number of disease conditions (Debenham *et al.*, 1987; Nicotera, 1991; Mazzarello *et al.*, 1992; Jiricny, 1994; Tanaka and Wood, 1994). The following sections serve to introduce many of the different types of chemical modifications that can occur in DNA and their potential consequences. For almost all of these lesions, there are one or more enzymes that serve to return the molecule to its original state. These enzymes and their actions are described briefly with reference to the type of damage that they repair. The enzymatic repair of DNA has been extensively reviewed in the past (Hanawalt *et al.*, 1979; Lindahl, 1979; Lindahl, 1982; Friedberg, 1985; Sancar and Sancar, 1988; Wallace, 1988; Lindahl, 1990; Modrich, 1991; Lindahl, 1993).

1.1.1 A BRIEF OVERVIEW OF THE CAUSES AND EFFECTS OF COMMON TYPES OF DNA DAMAGE

Damage to DNA can occur by three major routes:

1) Chemical reaction with molecules in the surrounding environment.
2) Structural alteration by the effects of ionising radiation.

3) Incorporation of incorrect nucleotides, or of nucleotide analogues, during replication events.

So called 'spontaneous' damage can also be described, which occurs under normal physiological conditions (pH 7.4 and 37°C *in vitro*, or *in vivo*) and results in similar lesions to those produced by chemical reaction or ionising radiation.

The main types of damage that result have been discussed in detail elsewhere (Lindahl, 1979; Friedberg, 1985; Lindahl, 1993), and are as follows:

1) Base loss: This occurs spontaneously under physiological conditions, with purines being released 20-fold more frequently than pyrimidines. The rate of depurination in double stranded DNA is reduced by a factor of four over the rate observed in single stranded DNA. A growing mammalian cell has been estimated to lose about 10 000 purines, and several hundred pyrimidines a day by this route. The base loss is due to hydrolytic fission of the N-glycosidic bond between the base and the deoxyribose resulting in an apurinic or apyrimidinic site. In addition, the loss and partial destruction of bases by ionising radiation adds further to the tally. Abasic sites also result from the first step in DNA base excision repair.

2) Strand breakage: This event is made more frequent by base loss, as the sugar residues of apurinic/ apyrimidinic sites that result are in equilibrium between the major (~99%) furanose (closed) form and the reactive aldehyde (open) form. The aldehyde form causes loss of a phosphate group on the 3' side of the abasic site by a process of β-elimination. The estimated lifetime of an abasic site before strand breakage is about 400 hours, and this is reduced by the presence of certain divalent metal ions, or by polyamines. In addition, strand breakage can occur more rarely by bond scission as a result of ionising radiation.

3) Base modification: Although base modifications by cellular enzymes are a normal feature of DNA, alterations to indigenous bases are by far the most prevalent form of DNA damage. The modifications which can be classed as damage are those changes that would alter the integrity of the DNA template during replication. This means that the bases would give rise to a transition (purine to purine, or pyrimidine to pyrimidine) or transversion (purine to pyrimidine, or *vice versa*) mutation, due to modified base pairing preference. Alternatively the damage modifications are those which can arrest replication completely, thus acting as lethal agents.

The modification of bases occurs spontaneously via hydrolytic deamination, with cytosine being the most prone of the four naturally present bases. The rate of this event in single stranded DNA is slightly faster than for loss of purine bases, with the half-life of a cytosine residue calculated to be about 200 years (Lindahl, 1993). In double stranded DNA, the rate is drastically lowered by a factor of about 150. The product of the deamination of cytosine is uracil, which preferentially pairs with the base adenine. Thus the deamination of cytosine, if uncorrected, will result in a transition mutation (Section **1.2.1**). The deamination of cytosine can also be brought about by reaction with bisulphite ions, which are the aqueous form of the environmental pollutant sulphur dioxide. In addition to the eventual formation of uracil by this means, there are two reaction intermediates which are also pro-mutagenic.

Reaction of cytosine with nitrous acid is another means of deamination of this base, and it is also a means of deamination of adenine and guanine to form hypoxanthine and xanthine, respectively. The deamination of adenine and guanine also occurs spontaneously, but it is a very minor reaction, with adenine deamination events some 50-fold less frequent than those of cytosine. Guanine is thought to deaminate even less frequently than adenine. Lastly, a common enzymatic modification of cytosine is its methylation at the 5 position of the pyrimidine ring. The resulting base 5-methylcytosine undergoes a deamination event four times more rapidly than cytosine.

The next commonly occurring modifications to DNA bases are those induced by ultraviolet light. These are the pyrimidine dimers, and can theoretically occur in 12 isomeric forms, though one particular from known as *cis-syn* is thought to predominate. Pyrimidine dimers are formed between adjacent pyrimidines at wavelengths around 260 nm. It has been shown that cyclobutane pyrimidine dimers containing cytosine deaminate to form uracil in a predictable fashion (Tessman *et al.*, 1994). The formation of pyrimidine dimers is most common between thymine and thymine, and least common between a pair of cytosines. In addition, the likelihood of formation of pyrimidine dimers appears to be dependent on the sequence of the flanking region. Another type of ultraviolet lesion is the pyrimidine-pyrimidine (6-4) adduct, which can be formed between T-C, C-C, and T-T sequences, but not between C-T sequences.

Ultraviolet radiation can also result in addition of water across the 5, 6 double bond in pyrimidines to form pyrimidine hydrates. This type of lesion is not thought to be significant as they have very short half lives and are readily dehydrated back to the

normal base. The di-hydroxy addition to the 5, 6 bond in thymine results in the formation of a much more stable thymine glycol. Finally, photosensitisation reactions can occur, where a compound other than the DNA absorbs the light energy and passes this onto the base which becomes modified. This type of reaction occurs between certain planar compounds and DNA bases. DNA can become cross-linked between strands by such mechanisms. In addition, it has been observed that DNA can become cross-linked to protein in ultraviolet light in the presence of alkylating agents such as β-propiolactone (Friedberg, 1985).

Another cause of base modification is ionising radiation, which can destroy bases directly, or indirectly by the excitation of molecules in the environment which become ionised or form radicals. It has been shown that the double stranded form of DNA is an effective barrier to ionising radiation damage, with single stranded DNA and lone nucleotides having far less resistance to the reactive species. Due to the large numbers of potential reactions with the many environmental compounds that can be ionised to form reactive species, the types of modifications that can occur to bases are numerous. The modifications of thymine have been well studied, and thymine glycol type compounds are the predominant result, reaction being most frequent at the 5, 6 double bond of the pyrimidine ring. These compounds can degrade further to formylpyruvylurea, urea, and N-substituted urea derivatives. Though less work has been done on cytosine, the 5, 6 double bond is again found to be the major reaction point. With the purines, ring opening and partial destruction result in formamidopyrimidine compounds, as well as the formation of 8-hydroxyguanine, which if unrepaired will lead to a transversion mutation. Again, inter-strand cross-links, and DNA-protein cross-links, are a minor type of reaction product from ionising radiation (Friedberg, 1985).

Base modification occurs most frequently due to reaction with environmental mutagens. These mutagens are constantly present in the cellular environment, and give rise to the great majority of DNA damage. Examples of this type of base modification are the base deaminations mentioned earlier, which result from reaction with bisulphite and nitrous acid. In addition, the common cellular transmethylation methyl group donor *S*-adenosylmethionine is known to methylate the DNA bases guanine and adenine, to yield 7-methylguanine and 3-methyladenine. Though 7-methylguanine is not pro-mutagenic, it may present problems for certain DNA binding proteins with possibly disastrous consequences due to erroneous cell signalling. An estimated 600 non-enzymatically

catalysed methylation (alkylation) events per day are thought to occur in DNA adenine residues in a human cell, yielding 3-methyladenine. This modified base blocks replication, and is therefore cytotoxic. This level of alkylation can also be achieved by low levels of alkylating agents such as methyl methanesulphonate.

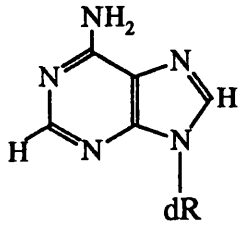
Alkylating agents such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) and methylnitrosourea are responsible for the formation of the transition mutation-forming lesion O^6-methylguanine. This modified base is also seen to be formed in stationary phase *E.coli* mutants deficient in repair capability for this particular lesion. This has led to tentative evidence which suggests that certain metabolites, common in quiescent cells, act as DNA damaging agents. It is likely that many cellular compounds, such as metabolic intermediates and reducing sugars, react with bases, or that nucleotide analogues are misincorporated by DNA polymerases during DNA synthesis events, resulting in a wide variety of lesions that are potentially mutagenic or cytotoxic. Finally, DNA strands can be cross-linked by various chemical agents, for instance, two guanines on opposite strands of DNA may be linked through their N^7 positions by nitrogen mustard. The cross-linking of DNA results in the blockage of transcription and replication, due to the strands becoming non-denaturable at the lesion. Blocks to transcription and replication can also occur if DNA-protein cross-links are present. This has been shown to occur in the presence of certain alkylating agents (Friedberg, 1985).

Common types of DNA base lesions are illustrated in figure **1.1.1.**

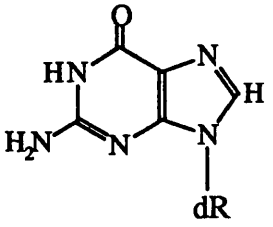
1.1.2 A BRIEF OVERVIEW OF THE AVAILABLE CELLULAR RESPONSES TO DNA DAMAGE

Considering the various types of damage that can occur to DNA and the potentially disastrous consequences this can have on the viability of the cell, it is not surprising to find that many complex and specific mechanisms exist to reduce and counteract the occurrence and effects of DNA damage. These types of enzymatic prevention and repair activities have been described extensively elsewhere, (Hanawalt, 1979; Lindahl, 1979; Lindahl, 1982; Friedberg, 1985; Sancar and Sancar, 1988; Wallace, 1988; Sakumi and Sekiguchi, 1990; Kunz and Kohalmi, 1991; Modrich, 1991) and may be grouped as follows:

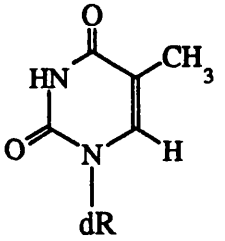
ADENINE



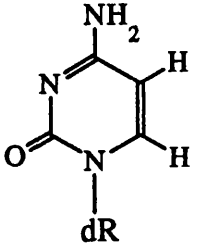
GUANINE



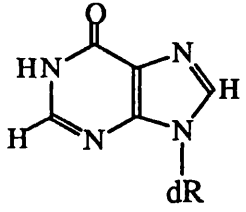
THYMINE



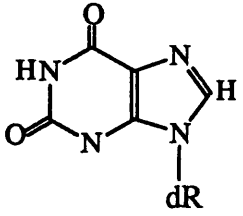
CYTOSINE



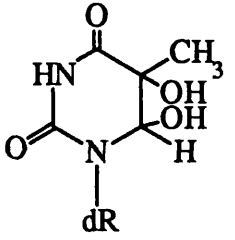
HYPOXANTHINE



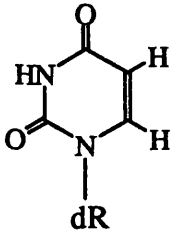
XANTHINE



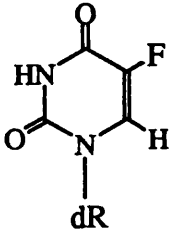
THYMINE GLYCOL



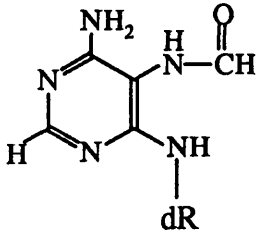
URACIL



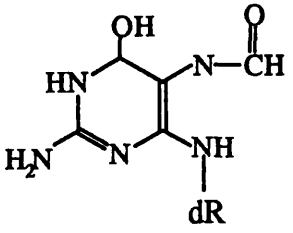
5-FLUOROURACIL



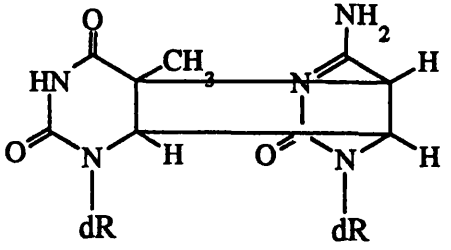
4, 6-DIAMINO-5-FORMAMIDO-PYRIMIDINE



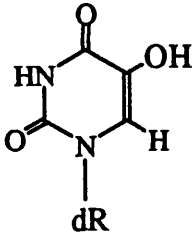
2, 6-DIAMINO-4-HYDROXY-5-FORMAMIDOPYRIMIDINE



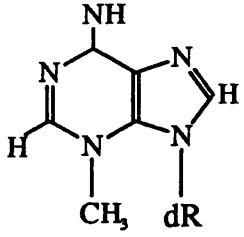
PYRIMIDINE DIMER



5-HYDROXYURACIL



3-METHYLADENINE



8-HYDROXYGUANINE

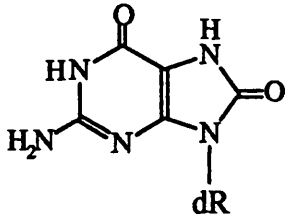


Figure 1.1.1

The normal DNA bases (top row), and the most common types of lesions associated with them. The nomenclature given is for the free bases, though the position of the N-glycosidic bond is shown in each case. The structure shown for 8-hydroxyguanine is the form taken when attached to DNA, namely 8-oxo deoxyguanidine. The column at the end shows the only other known substrates of UDGase.

1) Environmental Control: Deterrence of damage by control of free-radicals in the cellular environment, and control of the nucleotide levels in the cell.

2) Damage Reversal: Direct reversal of damage sustained.

3) Base Excision Repair: Removal of damaged bases from DNA, followed by the removal of the abasic sites left behind, and reinsertion of the correct nucleotide.

4) Nucleotide Excision Repair: Removal of damaged nucleotides from DNA, followed by repair synthesis.

5) Induced Damage Responses: Induction of multi-enzyme responses to DNA damage sites, including repair of strand breaks and cross-links.

Short descriptions of types 1 to 3 will now be given, with respect to the enzymes involved and their respective actions. Types 4 and 5 are more complex repair responses which make use of many proteins acting in concert. These responses are generally well understood in *E.coli*, and direct comparisons can be made between these systems and analogous eukaryotic systems. However, the complexity of these systems would make even a short description of them very lengthy. Instead, a brief outline of the types of enzymatic functions involved, and the circumstances under which these are invoked will be presented. More detailed discussion is presented elsewhere (Friedberg, 1985).

1) Environmental Control: Suppression of damaging species such as oxygen radicals and reactive ions is achieved by cellular enzymes such as catalase and superoxide dismutase. In eukaryotes, the compartmentalisation of the cell further protects DNA from cytoplasmic levels of reactive species. The nucleus is found to be a less oxygen rich environment than the cytoplasm, though the same is not true for mitochondria which amass higher levels of oxidative damage (Lindahl, 1993).

The levels of nucleotides in a normally functioning cell are maintained in a strict balance. The exact levels depend on the source examined, with mammalian cells possessing nucleotide pools 20-fold larger than bacterial cells. Experiments in which imbalances in the relative levels of nucleotides are introduced show that mutations associated with DNA replication are increased. In addition, if alkylation damage is inflicted upon such cells, the levels of mutation are higher than in wild type cells. Detailed studies on the types of misincorporation found in cells in which nucleotide pool imbalances exist, indicate that replication is more error prone overall, and not just with respect to the nucleotide which has had its level increased (Kunz and Kohalmi, 1991).

Nucleotide imbalances are usually due to mutations in enzymes in the synthesis or regulation pathways. One particular example is a mutation in the deoxyuridine triphosphatase enzyme. This enzyme keeps the levels of 2'-deoxyuridine 5'-triphosphate (dUTP) in the cell very low. The result is that replication past adenine residues vastly favours the incorporation of thymine, even though both thymine and uracil have similar kinetics for incorporation. When the enzyme is not present, the levels of dUTP increase and uracil is incorporated at a significant level. The effect is DNA fragmentation due to the cellular response to uracil in DNA (Section **1.2.2**).

2) Damage Reversal: The most investigated direct reversals of DNA damage are mediated by the following enzymes: DNA photolyase, O^6-methylguanine-DNA methyltransferase, DNA ligase, and purine base insertase.

The formation of *cis-syn* pyrimidine dimers has been mentioned earlier (Section **1.1.1**). This lesion can be repaired by direct reversal of the dimer into its constituent bases by a process of enzymatic photoreactivation. The enzyme that catalyses this reaction is known as DNA photolyase. Enzyme catalysed photoreactivation activities which have been well characterised have been isolated from *E.coli*, *S.griseus*, and *S.cerevisiae* (Sancar and Sancar, 1988). Photoreactivating enzyme activity has been demonstrated in plants, and in a variety of invertebrates, fish, reptiles, and marsupial mammals (Friedberg, 1985). The phenomenon of enzyme catalysed photoreactivation by mammalian (other than marsupial) enzymes related to photolyases has still not been definitively proven, owing to a failure to detect a light absorbing co-factor, and a failure to demonstrate consistent levels of production of the enzyme in isolated cells. The reaction is mediated by absorption of light with a wavelength greater than 300 nm by a loosely bound co-factor. The energy of the light thus absorbed is channelled via the enzyme to reverse the ultraviolet light induced dimer. The crystallisation of the *E.coli* photolyase was recently reported (Park *et al.*, 1993). Solution of this structure may allow a mechanism of action to be assigned to this process, as the kinetic parameters for the action of this enzyme have been thoroughly investigated (Friedberg, 1985; Sancar and Sancar, 1988).

The O^6-methylguanine methyltransferase enzyme is found to be present in bacteria and in eukaryotic cells. In *E.coli*, where the enzyme is known as the Ada protein, it is produced in response to the presence of alkylation damage of DNA. This has been shown in an adaptive response to treatment with MNNG (Sedgwick, 1987). This response is elicited once cells have been exposed to low levels of this alkylating agent. If cells which

If cells which have had this treatment are subsequently challenged with a very high dose, they are found to survive, whereas cells which have not previously been exposed are killed. The enzyme is found to catalyse the transfer of the methyl group from the damaged base onto one of its own cysteine residues, inactivating itself in the process. This 'suicide repair' mechanism is a direct reversal of the alkylation damage to the base (Lindahl *et al.*, 1982). The three-dimensional molecular structure of the *E.coli* Ada protein has recently been solved by X-ray crystallography (Moore *et al.*, 1994).

The ligation of DNA strands which have been broken by β-elimination events following depurination, or by the action of ionising radiation, constitutes the direct reversal of this type of damage. The repair is carried out by a DNA ligase. This enzyme is also recruited in other types of DNA repair, where new DNA synthesis is required as part of a multi-step repair mechanism (Figure **1.2**).

3) Base Excision Repair: This type of repair is carried out by enzymes which recognise a particular damaged base, or a particular class of base-lesion, and remove the base by scission of the N-glycosidic bond between the base and the deoxyribose. Some of these enzymes leave an abasic site behind which is subsequently removed by an AP (apurinic/ apyrimidinic) endonuclease enzyme and a 5'-deoxyribophosphodiesterase (5'-dRpase) enzyme. Other base-excision repair enzymes have their own AP lyase activity. The subsequent gap in the DNA is repaired by DNA polymerase and DNA ligase (Figure **1.2**). The enzymes involved in base removal are commonly grouped together as the DNA glycosylases, and their common characteristic is that they are highly specific monomers of molecular masses ranging from 20 000 to 30 000.

The most abundant of the cellular base excision repair enzymes is uracil-DNA glycosylase (UDGase), the major role of which is to remove uracil bases rapidly as they arise in DNA. The enzyme is apparently ubiquitous, has no associated AP lyase activity, and is the most highly conserved of the DNA repair enzymes studied so far. UDGase is discussed in greater detail in section **1.2**.

A similar DNA glycosylase activity exists to remove hypoxanthine residues from DNA, but it is not as rapid or abundant as UDGase. No known glycosylase activity exists for xanthine residues, but these may be removed by another type of repair system. A 5-hydroxymethyluracil-DNA glycosylase exists in human cells, but apparently not in *E.coli* (Hollstein *et al.*, 1984; Boorstein *et al.*, 1987). Another activity recently investigated is a mismatch-specific thymine-DNA glycosylase from HeLa cells (Neddermann and Jiricny,

1994). The repair of 3-methyladenine residues in *E.coli* is achieved by two DNA glycosylases, but these do not appear to be very highly conserved in their action, specificity, or primary structure. In *E.coli*, there are two enzymes which can remove 3-methyladenine, known as 3-methyladenine-DNA glycosylase I (Tag, the product of the *tag* 1 gene) (Steinum and Seeburg, 1986) and 3-methyladenine-DNA glycosylase II (AlkA, the product of the *AlkA* gene) (Nakabeppu *et al.*, 1984). The latter is involved in the *E.coli* adaptive response, and also removes 7-methylguanine and 3-methylguanine. The AlkA protein has been crystallised (Yamagata *et al.*, 1988). The Tag enzyme is more like the UDGase in its mode of action, as it is constitutively expressed, and has a relatively narrow substrate specificity, for 3-methyladenine, 3-ethyladenine, and 3-methylguanine in DNA. It is also end-product inhibited by the free bases. Repair of 3-methyladenine has been demonstrated in eukaryotes, though these enzymes do not show any primary structure homology with the *E.coli* Tag protein, and are able to remove other lesions, reminiscent of the *E.coli* AlkA protein.

The removal of imidazole-ring opened purines that result from damage by ionising radiation is carried out by the formamidopyrimidine-DNA glycosylase (FPG) (Boiteux *et al.*, 1987). This enzyme has been isolated from prokaryotes and eukaryotes, and is rare among the DNA glycosylases in that it contains a known DNA binding motif. This zinc-finger motif is the only well recognised DNA binding motif to be positively identified in a DNA glycosylase so far. The FPG also has an associated AP lyase activity which can incise 3' to an abasic site. If a 5' incision to the site has already been made then the sugar phosphate is released, enabling DNA polymerase to repair the gap. In addition, ring opened bases, urea and its derivatives are removed by a urea-DNA glycosylase which is in fact the same enzyme as that known to remove thymine-5, 6- hydrates, namely endonuclease III, which has an associated AP lyase activity (Asahara *et al.*, 1989). The AP lyase activity of this enzyme is to incise 3' to the abasic site, though this enzyme is unable to act if the 5' phosphodiester bond in the abasic site is broken. The AP lyase-treated abasic site can then be acted upon by 5'-dRpase and repair is completed by DNA polymerase and DNA ligase. The three-dimensional structure of this enzyme was recently determined (Kuo *et al.*, 1992), and it was unusual in that it contained an iron-sulphur center which is thought to serve to position the conserved basic residues of the enzyme so that they correctly contact the DNA phosphate backbone.

The structure report for endonuclease III implicated a catalytic mechanism for the

N-glycosylase reaction involving nucleophilic attack by an essential glutamic acid residue in a basic environment. This proposal is in agreement with the environment of the predicted catalytic site of T4 endonuclease V, also implicated as a glutamic acid residue. This observation has been proposed as a possible general mechanism for N-glycosylases.

The final DNA glycosylase of note is the pyrimidine dimer-DNA glycosylase, otherwise known as T4 endonuclease V. The structure of this enzyme was recently solved (Morikawa *et al.*, 1992), and though kinetic and sequence data had previously been used to indicate residues that might be involved in damage recognition and DNA binding, the absence of DNA in the structure makes this difficult to prove. The action of the enzyme is to hydrolytically cleave the N-glycosidic bond of the 5' base in the dimer, and to cause a 3' nick in the apyrimidinic site via its AP lyase activity. The lesion is still attached to the DNA by the other base in the dimer, and is only removed during strand displacement synthesis by DNA polymerase I.

Lastly, more than one AP endonuclease is known to be involved in base excision repair in *E.coli*. These enzymes hydrolyse the phosphodiester bond on the 5' side of the abasic site. The fission catalysed by these enzymes ensures that a 3' hydroxyl is free for subsequent DNA synthesis. Since the 3' side of the abasic site is prone to fission by spontaneous β-elimination events, the removal of abasic sites could be achieved by the action of AP endonucleases alone. However, this would be a very slow reaction, thus other enzymes exist to complete the removal of abasic sites. The 5'-dRpase enzyme, or the FPG can then complete the removal of the abasic site prior to repair synthesis. The former enzyme has been shown to be identical to the *RecJ* protein, a 5'→3' exonuclease (Lindahl, 1990; Dianov *et al.*, 1994).

The most common AP endonuclease activity in *E.coli* is exonuclease III. The name arose from the fact that the enzyme was initially identified by virtue of its 3'→5' exonuclease activity on double stranded DNA *in vitro*, which is a minor role. In addition, *E.coli* contains another AP endonuclease called endonuclease IV. This is homologous in function to the AP endonucleases isolated from other organisms ranging from bacteria through to the higher eukaryotes including humans (Friedberg, 1985; Lindahl, 1990).

4) Nucleotide Excision Repair: The multi-protein repair mechanism for short patch repair (~15 nucleotides) is well understood in *E.coli*, and may be briefly described as follows. The UvrA protein in the presence of ATP can dimerise and bind to DNA. One molecule of UvrB can interact with this dimer, and the entire complex scans the DNA until a repairable lesion is located. These may be pyrimidine dimers of any type including pyrimidine-pyrimidine (6-4) adducts, or cross-linked lesions. The UvrA dimer then dissociates with hydrolysis of ATP, and the UvrB remains at the site of the lesion until it is bound by a single molecule of UvrC. Endonucleolytic incisions are then made at the 8th phosphodiester bond 5' and the 4th or 5th phosphodiester bond 3' of the lesion. The exact spacing appears to be dependent upon the geometry of the lesion involved. Finally, UvrD and DNA polymerase I act to remove the damaged oligonucleotide section prior to repair synthesis and ligation. UvrC appears to be dissociated from UvrB by the action of UvrD (Weeda *et al.*, 1993).

In addition, long patch repair exists to replace stretches of nucleotides of the order of kilobases. This mechanism in *E.coli* involves the MutHDLS gene product system, and relies on the fact that the template strand after DNA synthesis is in a *dam* methylated state, and thus can be easily distinguished from the newly synthesised strand. If base damage is detected in this stretch of newly synthesised DNA, large tracts may be subject to repair synthesis under control of this system (Modrich, 1991). Lastly, long patch repair is also an induced response to DNA damage. This is outlined in (5) below. Analogous systems to long and short patch repair are known to exist in yeast and humans, and deficiency in humans is part of Bloom's Syndrome and Cockayne's Syndrome defects.

Finally, very short patch repair exists to replace single base mismatches. This system relies on nucleotide excision of one or very few nucleotides from the site of a mismatch. The system is flanking sequence sensitive, and can remove uracil from U·G mismatches, although not as efficiently as UDGase. This system specifically removes G·T mismatches which result from the deamination of 5-methylcytosine residues, normally found after modification of DNA by host cytosine-5 methyltransferases (Fox *et al.*, 1994).

5) Induced Damage Responses: An induction of otherwise repressed DNA repair functions is known to occur in both prokaryotic and eukaryotic cells. In this response, called the SOS response, the repressor molecule, *LexA*, is cleaved by the *RecA* protein, which releases *LexA* molecules from DNA and allows transcription of more than 20 proteins associated with the tolerance or repair of DNA damage. The *RecA* protein also

cleaves λ-repressor, though this is less efficient than its cleavage of *LexA*. The cleavage of λ-repressor results in reactivation of the phage from a lysogenic state. The *RecA* protein can be thought of as a cellular alarm, signalling that DNA damage is present, however the damage does not need to be severe for this response to occur. Since *LexA* is autoregulatory, as soon as overproduction of *RecA* protein ceases the levels of intact *LexA* soon return to normal, resulting in a fast switching-off of the response.

The types of repair that occur in the SOS response are those associated with the removal of bulky lesions. The repair with which the *RecA* protein is involved is not all strictly part of the SOS response. The repair involves DNA polymerase activity past sites of damage that have not yet been removed, and homologous recombination of identical DNA molecules to replace regions at which DNA replication is stalled. Homologous recombination involves *RecA* protein and single strand binding protein, and is also used to repair double strand breaks. In the presence of large amounts of damage, tolerance of damage sites occurs. In such circumstances, error-prone DNA synthesis can occur, where the polymerase adds bases to a site opposite a lesion. In the case of abasic sites the error rate is very high, with adenine being incorporated 59% of the time. In constitutive mutants for SOS repair, the incidence of C·G to A·T transversions is abnormally high owing to a combination of the most common deamination event (C→U) being incompatible with the error-prone replication base preference for adenine (Friedberg, 1985; Wallace, 1988).

Finally, it is known that DNA repair appears to be more widespread in actively transcribed regions of the genome. This preference for coding regions has recently been attributed to the involvement of RNA polymerases and transcription factors, which can identify sites of DNA damage such as mismatches and bulky adducts (Downes *et al.*, 1993). When these enzymes stall on the DNA, they act as molecular beacons which are only displaced by the arrival of DNA repair enzymes. It is also possible that other features associated with open reading frames (ORFs), such as upstream elements, may make them more heavily policed by the DNA repair enzymes. A general detection mechanism exists in the form of poly(ADP-ribose) polymerase which is thought to bind to nicks in DNA, where it can stall replication until repair is carried out (de Murcia and de Murcia, 1994). In addition, this enzyme can transfer ADP-ribose to certain histone proteins in order to allow relaxation of chromatin for DNA repair. Finally, it is also possible that a multienzyme repair complex associated with DNA polymerases exists. There is tentative evidence of the existence of a 'repairosome' in human cells (Seal and Sirover, 1986;

Downes *et al.*, 1993; Selby and Sancar, 1993).

1.2 URACIL-DNA GLYCOSYLASE, A UBIQUITOUS AND HIGHLY CONSERVED DNA REPAIR ENZYME OF EXTREME SPECIFICITY

UDGase is a base-excision repair enzyme which acts as the first step in a DNA repair pathway (Figure **1.2**) in which uracil, and a limited set of analogous bases, are removed from DNA. Detailed reviews concerning the properties of this enzyme have appeared in the past (Lindahl, 1979; Duncan, 1981; Mosbaugh, 1988; Tomilin and Aprelikova, 1989), but new information is constantly appearing. A full introduction to this enzyme is given in section **1.2.2**, but firstly the occurrence and consequences of uracil in DNA will be introduced.

1.2.1 URACIL IS NOT A FAVOURED COMPONENT OF DNA

Uracil, a nitrogenous base, is a molecule which is normally found as one of the four major bases of ribonucleic acid (RNA). This base differs from the analogous DNA base thymine by the absence, in uracil, of a methyl group at the 5-position of the pyrimidine ring (Figure **1.2.1 a**). Thymine is in fact formed by the methylation of 2'-deoxyuridine 5'-monophosphate (dUMP) in the uracil base by thymidylate synthase, to form 2'-deoxythymidine 5'-monophosphate (dTMP) (Figure **1.2.1 a**).

Both thymine and uracil are incorporated with similar efficiency by DNA polymerases (Brynolf *et al.*, 1978; Dube *et al.*, 1979) into a growing DNA strand to form a complementary base pair with the purine base adenine. However, the relative sizes of the deoxynucleotide triphosphate (dNTP) pools for dUTP and dTTP are tipped vastly in favour of dTTP by the action of the enzyme deoxyuridine triphosphatase (dUTPase) on dUTP to form dUMP (Figure **1.2.1 b**), and consequently dUTP is rarely incorporated into DNA in the usual cellular environment. Mutants lacking a dUTPase activity are seen to have highly elevated levels of dUTP and the DNA in such mutants has a high proportion of U:A base-pairs. It has been shown that incorporation of dUTP instead of dTTP during replication in *E.coli* and mammalian cells results in short Okazaki fragments (Brynolf *et al.*, 1978; Sedwick *et al.*, 1981) and DNA fragmentation (Wist *et al.*, 1978; Ingraham *et al.*, 1986). The DNA base-excision repair enzyme activity, UDGase, is present in most

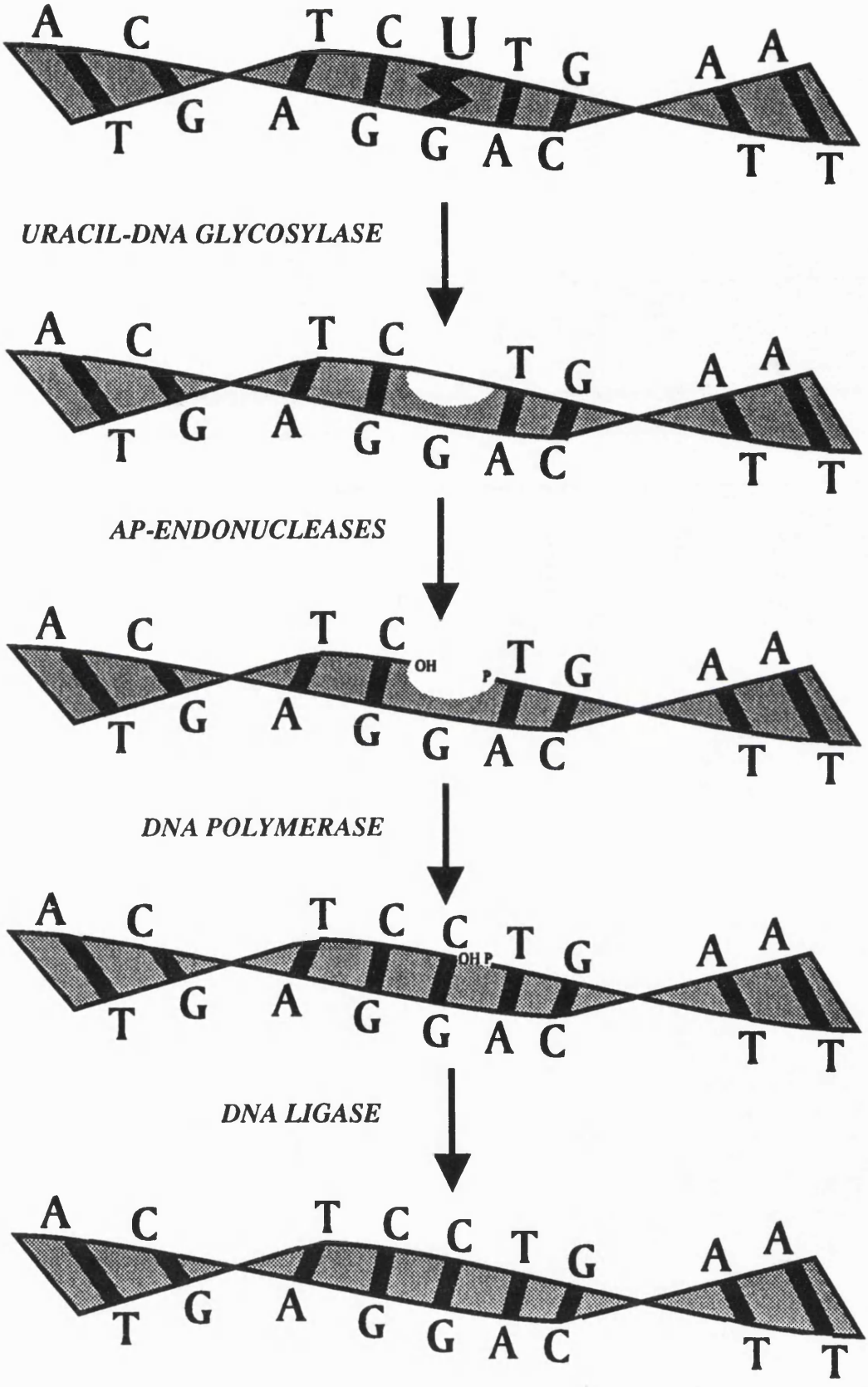


Figure 1.2

The uracil-DNA repair pathway. Uracil-DNA glycosylase acts at the first step, removing uracil by hydrolysis of the N-glycosidic bond between the deoxyribose and base. The remaining abasic lesion is subsequently repaired by the enzymes shown.

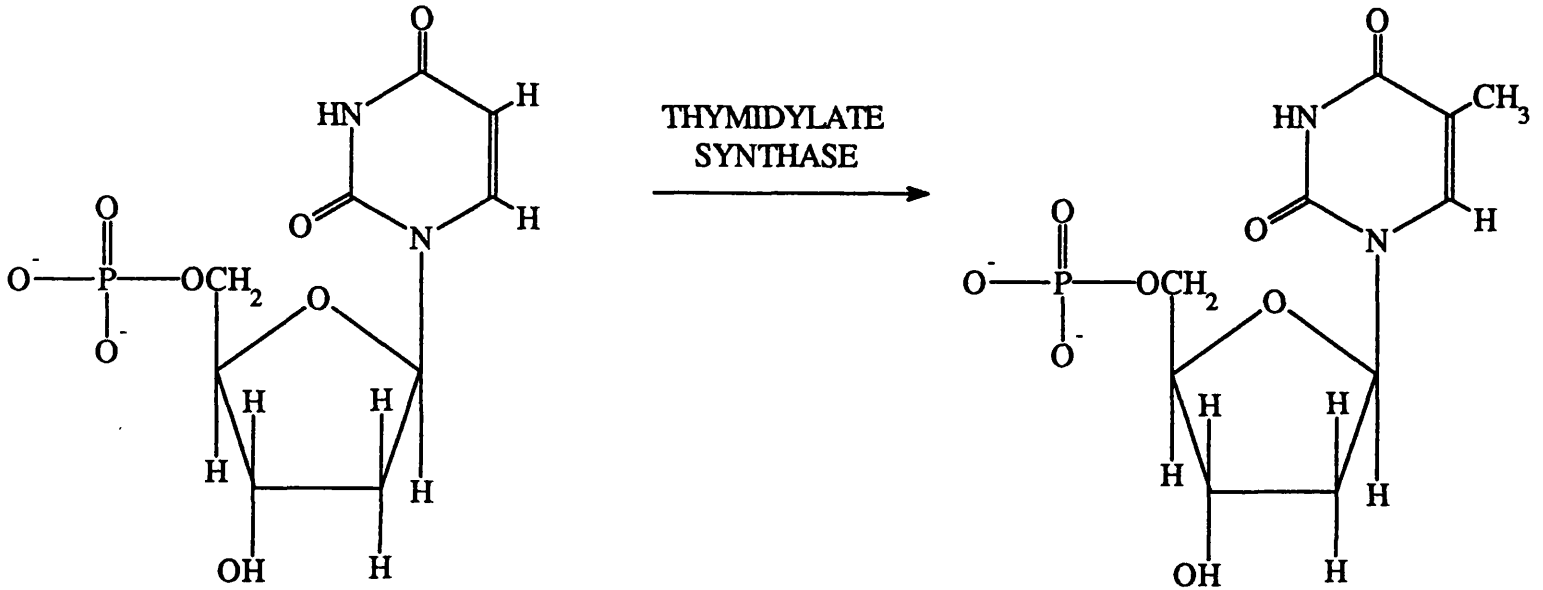


Figure 1.2.1 a
Uracil, not normally incorporated into DNA, is a precursor of thymine.

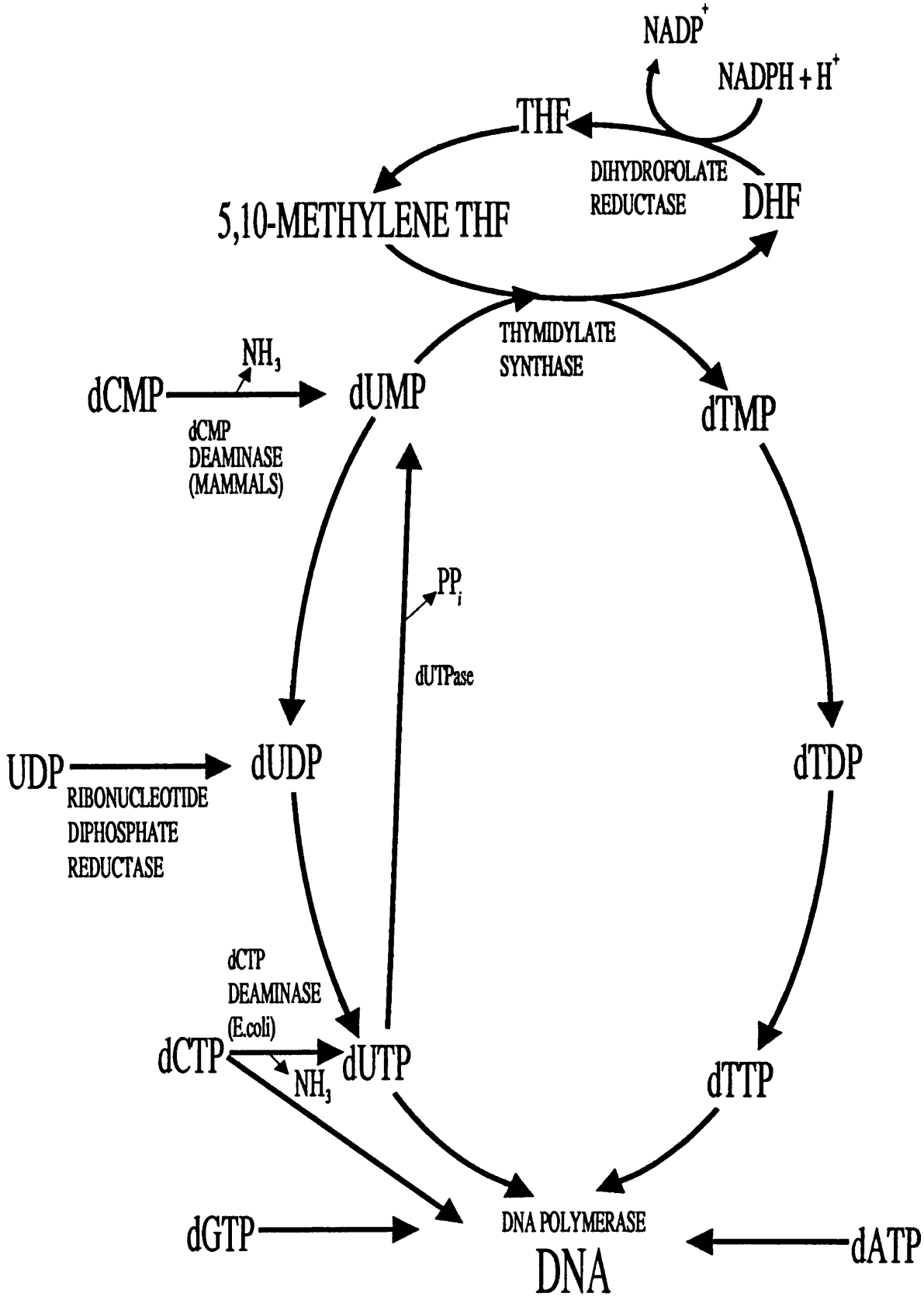


Figure 1.2.1b

The relative concentrations of dUTP and dTTP are tightly controlled, making incorporation of dUTP into a growing DNA strand a rare event.

cells (Section **1.2.2**) to immediately remove any uracil that occurs in DNA (Lindahl, 1974). This activity is likely to be a precursor to such strand breakage (Mosbaugh, 1988).

Why should there be such a bias against uracil in DNA considering it is a precursor of thymine, and as such is a metabolically 'cheaper' alternative (Figure **1.2.1 a**)? It has been shown that the U:A base pair is a thermodynamically less favourable arrangement than the T:A base pair (Carmody and Vary, 1993), and that the methyl group of thymine may be a somewhat stabilising influence on the B-DNA helix (Delort *et al.*, 1985; Ivarie, 1987). This is one possible explanation, but more convincing still is the fact that uracil in DNA is a pro-mutagen in two possible ways.

Firstly, U-G mismatches can occur in DNA due to the formation of uracil by the spontaneous hydrolytic deamination of the base cytosine, which occurs with significant frequency under normal physiological conditions (Shapiro and Klein, 1966; Frederico *et al.*, 1993). In single stranded DNA the reaction is more than two orders of magnitude more frequent (Mosbaugh, 1988). The rate of reaction is also increased by heat (Lindahl and Nyberg, 1974). Uracil can also be formed from cytosine by the action of environmental mutagens such as bisulphite, produced when sulphur dioxide dissolves in physiological fluid, and nitrous acid (Lindahl, 1979). In addition it has recently been shown that cyclobutane pyrimidine dimers, if tolerated by the cell beyond a critical time, will undergo deamination to form uracil (Tessman *et al.*, 1994). Considering that such base changes can result in a transition mutation from a C:G to U:A base pair (Figure **1.2.1 c**) and that DNA in single stranded form occurs at replication forks, it is very likely that uracil, unless removed promptly as it arises in DNA, will be mutagenic. Secondly, it has been shown that certain DNA-binding regulatory proteins have reduced affinity for, or no longer bind to, their recognition sequences when uracil is present in that sequence (Verri *et al.*, 1990; Focher *et al.*, 1992a). The uracil can either be part of a U:A base-pair or a U:G mismatch. This introduces the possibility of failure of certain cell signalling pathways with potentially lethal or transforming effects unless the uracil is removed rapidly as it arises.

1.2.2 URACIL-DNA GLYCOSYLASE

UDGase is present, with apparent ubiquity, in organisms which possess DNA genomes. The enzyme is highly efficient at removing uracil as it arises in DNA (Lindahl,

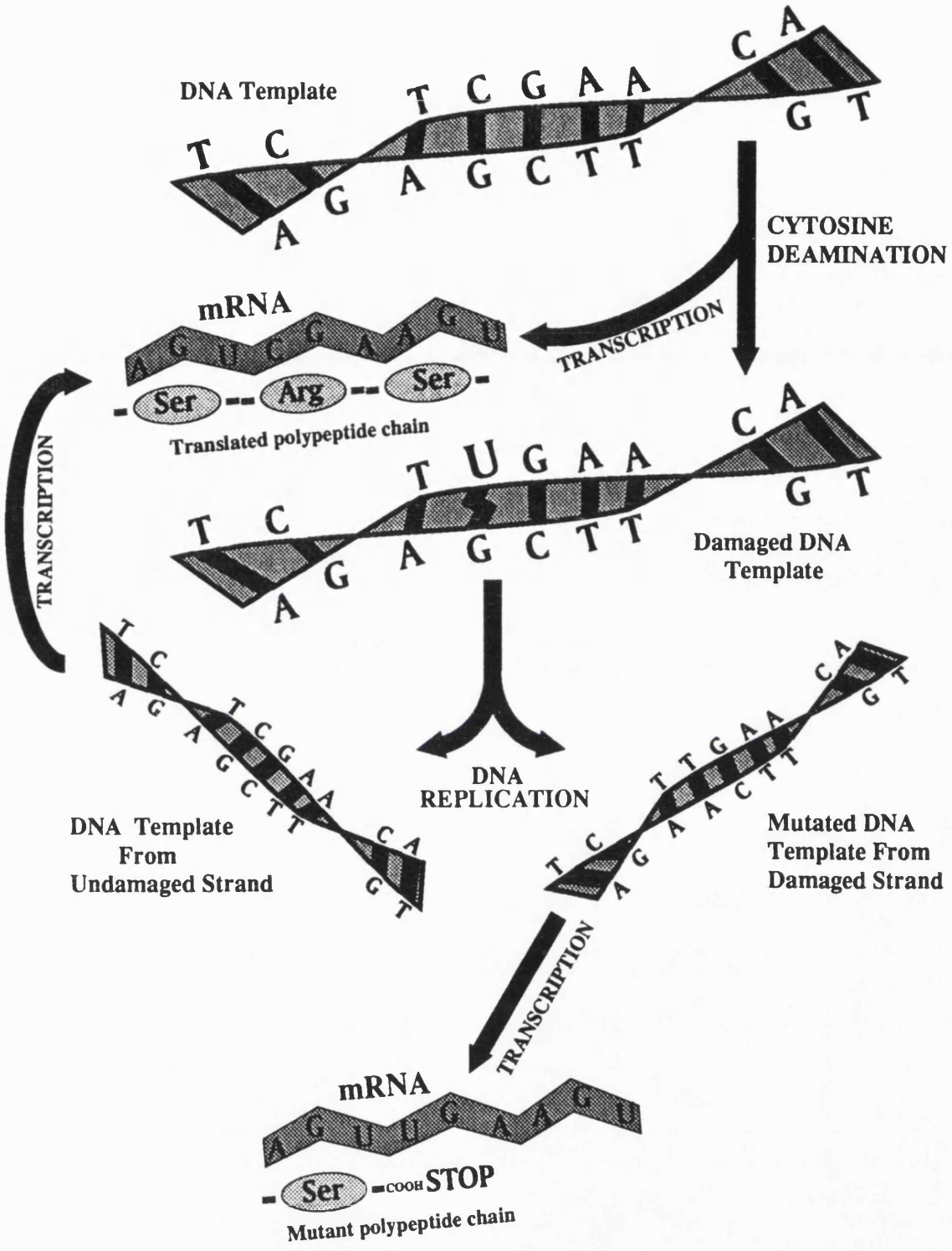
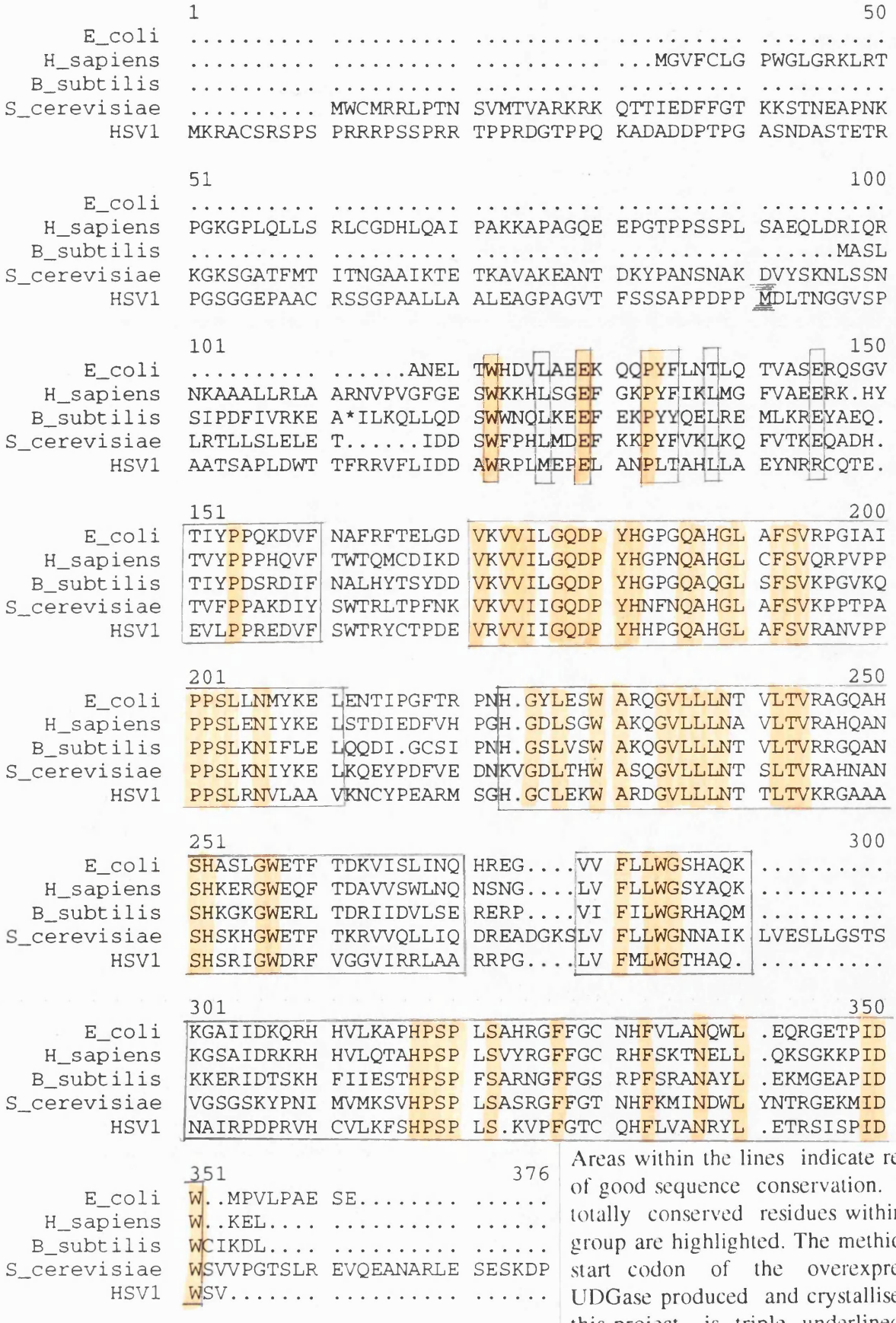


Figure 1.2.1 c

The consequences of an unrepaired cytosine deamination event. The presence of uracil in one of the template strands at replication results in a transition mutation from a C:G to U:A base pair. Resulting polypeptides from the mutant transcript may be severely affected.

1974; Lindahl *et al.*, 1977). It has also been shown to remove 5-fluorouracil (Mauro *et al.*, 1993), and 5-hydroxyuracil (Hatahet *et al.*, 1994), from DNA at a reduced rate compared to that of uracil removal. The enzyme is very highly conserved between all types of organisms in which it has been studied (Figure **1.2.2**). The mode of action of this enzyme has been shown to be the hydrolytic fission of the N-glycosidic bond between uracil base and the deoxyribose sugar (Lindahl *et al.*, 1977) with the subsequent production of an abasic site in the DNA and free uracil base (Figure **1.2**). UDGase activity has been demonstrated in a wide variety of organisms, and the enzyme has been studied extensively (Cone *et al.*, 1977; Lindahl *et al.*, 1977; Talpaert-Borlé *et al.*, 1979; Leblanc *et al.*, 1982; Crosby *et al.*, 1981; Kaboev *et al.*, 1981; Blaisdell and Warner, 1983; Colson and Verly, 1983; Kaboev *et al.*, 1985; Guyer *et al.*, 1986; Bensen and Warner, 1987; Caradonna *et al.*, 1987; Morgan and Chlebek, 1989; Olsen *et al.*, 1989; Williams and Pollack, 1990; Masters *et al.*, 1991; Bones, 1993; Upton *et al.*, 1993; Winters and Williams, 1993). The enzyme, from all sources so far studied, has been shown to be monomeric and active in the absence of divalent metal ions (and in the presence of chelators such as EDTA and EGTA) or other co-factors. There is a degree of variability in molecular mass, although the larger enzymes tend to be eukaryotic and may contain N-terminal localisation sequences (Olsen *et al.*, 1989; Slupphaug *et al.*, 1993). When N-terminal signal sequences have been accounted for, the molecular mass of a typical UDGase is around 23 000 to 28 000. In general, UDGase has been shown to remove uracil from DNA, but not from RNA, neither does it remove uracil from dUMP or dUTP (Talpaert-Borlé *et al.*, 1979; Caradonna and Cheng, 1980; Talpaert-Borlé *et al.*, 1982; Domena and Mosbaugh, 1985; Krokan and Wittwer, 1981).

The kinetics of the enzyme from *E.coli* have been thoroughly studied both on single and double stranded DNA and various oligonucleotides (Lindahl *et al.*, 1977; Varshney and van de Sande, 1990; Eftedal *et al.*, 1993). This enzyme has been shown to bind DNA in a non-specific manner and act processively with an average scan length of 1.5 to 2 kilobases before dissociation (Higley and Lloyd, 1993). In an investigation carried out on the kinetics of uracil excision from oligonucleotides by the *E.coli* enzyme, it was noted that the shortest substrate is a dimeric oligonucleotide phosphorylated at both ends with the uracil at the 5' end, or a trinucleotide phosphorylated at the 5' end only, with the uracil at the 5' end (Varshney and van de Sande, 1990). These experiments also showed that uracil at the 3' end is generally a poor substrate and that uracil at the 5' end is not



Areas within the lines indicate regions of good sequence conservation. The totally conserved residues within the group are highlighted. The methionine start codon of the overexpressed UDGase produced and crystallised in this project is triple underlined and overlined.

Figure 1.2.2
Multiple sequence alignment of species distinct examples of main class UDGases, done using default settings in the PILEUP program of the GCG package.

removed unless that end is phosphorylated. A study carried out on UDGases from *E.coli*, human and HSV1, showed that uracil is removed from U:G base pairs more efficiently than from U:A base pairs (Verri *et al.*, 1992), though an independent study carried out on the UDGases from calf thymus and *E.coli* showed the opposite effect (Eftedal *et al.*, 1993). However, the result of the latter investigation was not in accord with the results of a third investigation of this phenomenon using the *E.coli* UDGase (Neddermann and Jiricny, 1994). The investigation by Eftedal *et al.* also stated that uracil from C:G rich environments is removed less efficiently than from A:T rich environments. This same study indicated that a thymine immediately 3' to a uracil slowed the rate of removal of that uracil significantly. The rate is slowed further when a cytosine or guanine is located directly 5' to the uracil. This data seems to indicate that some local unwinding may be occurring, and it is a general observation that UDGases remove uracil approximately ten times faster from single stranded DNA on average. One notable exception to this trend, observed in an early study on the enzyme from calf thymus (Talpaert-Borlé *et al.*, 1979) has since been thrown into question (Eftedal *et al.*, 1993).

UDGase has been shown to be product inhibited in a non-competitive manner by free uracil base (Lindahl *et al.*, 1977, Caradonna and Cheng, 1980) and competitively by abasic DNA sites (Domena *et al.*, 1988). An apparent inhibition of UDGase activity by free thymine, dTMP, and deoxythymidine residues in DNA (Talpaert-Borlé *et al.*, 1979) has never since been observed. Some base-analogues of uracil have been shown to be inhibitory, though less so than uracil (Krokan and Wittwer, 1981). Recently, novel inhibitors based on uracil have been synthesised. Such inhibitors can be used to selectively inhibit a UDGase from HSV1 while affecting the human host enzyme to a far lesser extent (Focher *et al.*, 1993; Botta *et al.*, 1994). Protein inhibitors of the UDGases from *E.coli* and *B.subtilis* have been identified. An early product from bacteriophage T5 infections is one of these (Warner *et al.*, 1980), though it is not well understood why this phage should encode such a function as its genome normally contains no uracil. In addition, a specific protein inhibitor (UGI) of the *B.subtilis* enzyme, produced in bacteriophage PBS1 or PBS2 infections, has been isolated and also shown to inhibit UDGases from other sources (Anderson and Friedberg, 1980; Karran *et al.*, 1981; Wang and Mosbaugh, 1988; Wang and Mosbaugh, 1989; Winters and Williams, 1990). Both of the phages PBS1 and PBS2 have DNA genomes which contain uracil instead of thymine, and the inhibitor of UDGase completely inactivates the host enzyme within 4 minutes

post-infection. The interaction of the [PBS2 UGI:*E.coli* UDGase] complex has been studied (Wang *et al.*, 1991) and a kinetic scheme has been proposed for the mode of inhibition of the enzyme by this inhibitor (Bennett *et al.*, 1993). The binding is irreversible except by incubation at 70°C in the presence of SDS or 8 M urea, but the mode of binding is not thought to be covalent (Bennett and Mosbaugh, 1992; Bennett *et al.*, 1993).

UDGase has been shown to be unable to bind DNA following PBS2 UGI binding, and UGI seems to be more efficiently bound than DNA. This suggests that the DNA binding site is the target of the inhibitor protein, or that a forced conformational change is imposed upon UDGase by UGI. The inhibition, by this protein, of UDGases from sources other than *E.coli* suggests that the UGI binding site is a highly conserved region in all these enzymes. Even so, the binding of the PBS2 encoded UGI can be used to kinetically distinguish two UDGases from different sources (Winters and Williams, 1990). Considering the high specificity of the enzyme for uracil, it is likely that the enzyme from any two sources will have a very similar tertiary structure arrangement, even though there is a degree of variability in molecular mass.

Attempts to determine the tertiary structure of a UDGase have not been reported other than by this author (Savva and Pearl, 1993). The tertiary structure, especially in complex with an oligonucleotide, would give a very good insight into the interaction of a protein with DNA, and also reveal a very highly conserved reaction mechanism which is activated by uracil, but not by thymine (Figure **1.2.1 a**).

Recently, there have been reports of enzymes that appear to be UDGases in a kinetic sense, but share little or no sequence homology with the main class of UDGases introduced above. One such report claims that the monomeric subunit of human glyceraldehyde-3-phosphate dehydrogenase (G3PDase), a tetrameric enzyme, is in fact a UDGase (Meyer-Siegler *et al.*, 1991). This enzyme displayed UDGase activity when isolated and characterised (Vollberg *et al.*, 1989), and was thought to be the human UDGase. Another enzyme from humans, which is part of the main class of UDGases was isolated at a similar time (Olsen *et al.*, 1989). The human G3PDase monomer has a molecular mass of about 37 000. Some G3PDases from other sources have been assayed, and show no UDGase activity. There is about 55 to 80% sequence homology between these G3PDases and the human enzyme, but it may be that there is a somewhat modified tertiary structure in the monomer of the human enzyme that allows a dual activity dependent on subunit association.

If this glycolytic enzyme is really also a UDGase, then it is an incredible example of convergent molecular evolution resulting in a single set of folds which have two completely distinct enzymatic activities on totally different substrates. In fact, G3PDase has many functions attributed to it (Meyer-Siegler *et al.*, 1991 and references therein), one of which is a single stranded DNA binding activity potentially involved in transcription regulation (Perucho *et al.*, 1977; Morgenegg *et al.*, 1986). It has been shown that binding of NAD^+ inhibits DNA binding by this enzyme. Interestingly, it appears that this enzyme acting as a UDGase is inhibited by uracil and its analogues 5-fluorouracil and 5-bromouracil in a different way (Seal *et al.*, 1987) to the human UDGase that is homologous to the main class (Krokan and Wittwer, 1981). There is also tentative evidence that this form of UDGase is part of a multi-enzyme complex comprised of at least two enzymes, itself and the 70 000 dalton catalytic subunit of human DNA polymerase α (Seal and Sirover, 1986).

This enzyme is also implicated as the UDGase that is modified in activity (Seal *et al.*, 1991) and immunogenic recognition (Vollberg *et al.*, 1987; Seal *et al.*, 1988; Seal *et al.*, 1990) in the inherited disorder Bloom's Syndrome (Nicotera, 1991). It is still in dispute whether total cellular activity of UDGase, which would include all types present in humans, is altered in Bloom's Syndrome (Vilpo and Vilpo, 1989). Expression of the gene encoding this G3PDase is dependent on the proliferative state of the cell, and coincides with peaks of UDGase activity (Meyer-Siegler *et al.*, 1992). It is not clear whether this or another type of human UDGase is the activity that is abnormally regulated in Bloom's Syndrome cells (Yamamoto and Fujiwara, 1986), as the human enzyme of the main class also exhibits cell-cycle dependent regulation (Slupphaug *et al.*, 1991).

The second reported UDGase activity showing no sequence homology to the main class is, again, isolated from human cells and shares some sequence homology with the cyclins. It also differs from the predominant human UDGase, the gene for which is located on chromosome 12 (Aasland *et al.*, 1990), in that it is encoded by a split gene containing two exons located on chromosome 5. All mammalian UDGases of the usual type so far isolated are encoded by genes that do not contain introns. However, an antibody that is specific only to this recently identified type of UDGase also cross reacts with cell extracts from simian species, but not rodents, indicating that it is a recently acquired activity in evolutionary terms. This particular type of new UDGase is also found to be a cell cycle dependent activity (Muller and Caradonna, 1991; Muller and Caradonna, 1993). This

proliferative dependency of expression has also been reported for other UDGases (Yamamoto and Fujiwara, 1986; Hernandez and Gutierrez, 1987; Cool and Sirover, 1989; Slupphaug *et al.*, 1991; Dudley *et al.*, 1992; Meyer-Siegler *et al.*, 1992; Bones, 1993; Weng and Sirover, 1993).

1.3 URACIL-DNA GLYCOSYLASE IS ENCODED BY VIRAL GENOMES: THE POSSIBLE ROLE OF THIS ENZYME IN LATENCY AND REACTIVATION OF HSV1

The finding that UDGase is encoded by viral genomes (Caradonna *et al.*, 1987; Upton *et al.*, 1993) has further increased interest in the apparently indispensable role of this enzyme in any organism that carries a DNA genome. Though an organism can be shown to be viable without this enzyme (Makino and Munakata, 1977; Duncan *et al.*, 1978; Burgers and Klein, 1986; Mullaney *et al.*, 1989; Chen and Lacks, 1991), it is contended that the fidelity of the genetic message is decreased in its absence with ultimately deleterious effects (Mazzarello *et al.*, 1992; Verri *et al.*, 1990; Focher *et al.*, 1992a; Focher *et al.*, 1992b). The following sections serve to define the herpesviruses as a group, and then to expand on the life-cycle of HSV1. The importance of viral UDGases in viral viability will also be introduced.

1.3.1 THE DEFINITION OF HERPESVIRUSES AND THEIR CLASSIFICATION

The following information is a description of what is meant by the term 'herpesvirus', and is presented in greater detail elsewhere (Honess and Watson, 1977; Roizman, 1991; Littler and Powell, 1992).

A virus can be included in the group classification 'herpes' by virtue of both structural and biological properties. The structure of a herpesvirion is based around the capsid, which appears to be invariant throughout members of the group. Structural information at 40Å resolution has become available for HSV1, which is very much the model herpesvirus. This data was obtained using cryo-electronmicroscopy (Schrag *et al.*, 1989) which is a non-disruptive technique, thus data on apparently almost intact virus particles has been obtained. Within the capsid is a double stranded DNA in the form of a torus. This is linear when extracted from virions but is immediately circularised upon

release from the capsid into infected cell nuclei. The DNA is of variable length, ranging from 120 to 230 kb, and the G+C content is also highly variable from 31 to 75%. The polyamines, spermine and spermidine, are also present, presumably to neutralise the negative charge on the DNA and enable tight packing within the capsid. The capsid is surrounded by an amorphous, fibrous 'tegument'. Finally, the entire virion is surrounded by a trilaminar envelope derived from the host cell that displays spikes of viral glycoprotein on its surface. These spikes are generally more abundant, and of shorter length (about 8 nm), than in other virus families that display surface glycoproteins.

In a biological sense, the herpesviruses can be broadly grouped as having the following features: infected cells are destroyed upon production of infectious virus progeny; the synthesis of viral DNA and the assembly of capsids takes place in the infected cell nucleus; finally, latency in the host is a common feature among all herpesviruses studied to date. The viral genome is present as a closed circular episomal DNA during this time, with only a few viral genes being expressed (Section **1.3.4**). Finally, a large range of enzymes involved in nucleic acid metabolism, modification, synthesis and repair are encoded by the viral genome.

Herpesviruses can be sub-divided into six main groups, designated by the letters A to F, which are based on the arrangement of sequences of DNA in the genome, and are currently classified into three types α-, β-, and γ- based on specific details of the biology of the virus. Groups A to E are characterised by repeated sequences, mainly terminal repeats, of more than 100 bp which subdivide the genome into shorter regions. HSV1 is a type α, group E herpesvirus. The structure of the HSV1 genome is depicted in figure **1.3.2**. The classification α entails that latency is most often established in sensory ganglia when it takes place, that the lysis of infected cells is particularly efficient, that in culture the virus spreads rapidly, and that there is a relatively short reproductive cycle and a variable range of hosts.

Finally, a classification exists, established primarily for herpes simplex viruses, concerning the order of production of proteins in herpes virus infections. The designation is again denoted by α-, β-, and γ-. The α (immediate early) proteins are those that are produced prior to *de novo* viral gene expression, the β (early) proteins are those that are produced after α proteins but are not dependent on viral DNA synthesis, and the γ (late) proteins are produced after the former two types, and to a greater or lesser extent rely on viral DNA synthesis (Littler and Powell, 1992).

1.3.2 THE STRUCTURE AND ORGANISATION OF THE HSV1 GENOME

The following is a brief overview of the detailed genomic layout of the HSV1 genome. More thorough treatments of this material are given elsewhere (Honess, 1984; McGeogh 1987; McGeogh *et al.*, 1991; Roizman and Sears, 1991).

Genomes of HSV1 strains are approximately 150 kb in length. Strain 17 which has been completely sequenced has a genome of 152,260 bp (McGeogh *et al.*, 1988). Typically for a herpes virus, the DNA is a linear, double-stranded toroid in the capsid, and is immediately circularised in the host. Figure **1.3.2** depicts the layout of a typical HSV1 genome. A terminal repeat sequence, termed a, is present at both ends. It is also repeated internally, and may be reversed, a', and may also be present in multiple tandem arrays, a_n. Other repeat sequences are found internally. These are termed b and c, and are repeated as inverted sequences, b' and c'. The internal repeats c and c' divide the unique regions of the genome into two pieces, U_L and U_S , known as the long and short unique regions respectively. These regions can be inverted so four possible isomers of the genome are possible, P (prototype), I_L (inverted L), I_S (inverted S), and I_{SL} (inverted S and L). All of these isomers are found in virus populations, and deletions of the internal repeats and parts of the unique regions, or other interruptions of the internal repeats can be carried out without affecting viral viability in cell culture. Such deletion mutants are frozen in their respective isomeric arrangements (Roizman and Sears, 1991).

The nucleotide sequence of the entire HSV1 genome has been obtained, allowing thorough assignments of G+C rich regions, overlaps, repeats and ORFs (McGeogh *et al.*, 1985; McGeogh *et al.*, 1986; McGeogh *et al.*, 1988). There are 72 ORFs, specifying 70 polypeptides; 2 ORFs are repeated. Of these ORFs, 56 are in U_L , 12 in U_S , and one each in b, b', c, and c'. The possibility of further genes which are either highly spliced, extensively overlapping, or which occur in repeat regions of the genome is yet to be shown (McGeogh *et al.*, 1988). The assignment of ORFs was done by examining codon usage, triplet periodicity in base frequencies, and looking for transcription signals associated with 3' polyadenylation of mRNA, located appropriately with respect to the ORF (McGeogh *et al.*, 1988). The exact translation start sites for some of the identified ORFs is not clear.

The ORF UL2, encoding the UDGase, is a prime example of such a sequence. The possibility that this enzyme function is produced from more than one start codon, or is

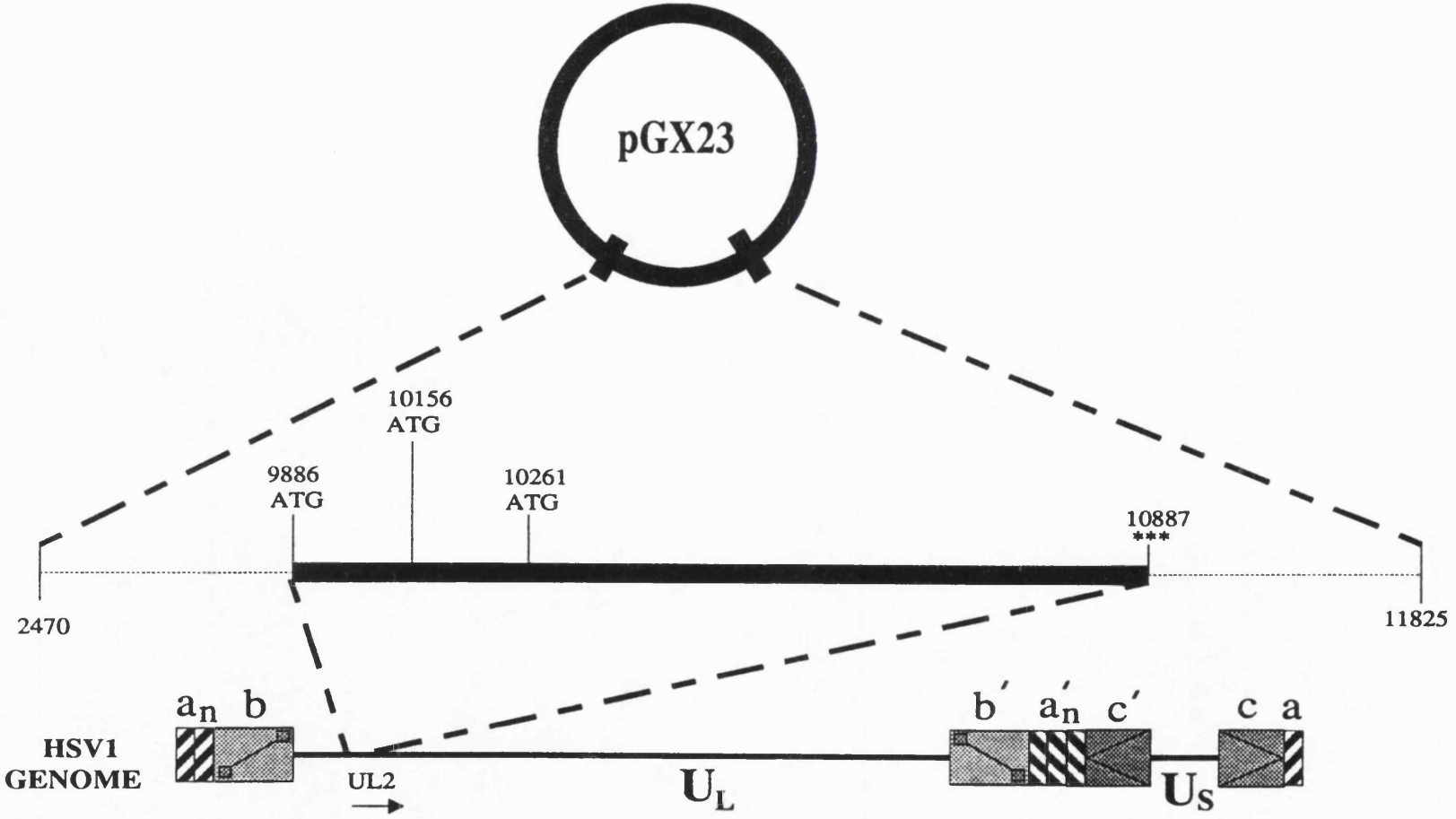


Figure 1.3.2

The position and orientation of UL2 in the HSV1 genome. The relative positions and co-ordinates, in bp of U_L , of the putative translation start sites (ATG) and stop point (***) of the UL2 open reading frame are highlighted. This region of the HSV1 genome is contained as a cloned inserted DNA in the plasmid pGX23.

processed post-translationally has been observed (Caradonna *et al.*, 1987; Mullaney *et al.*, 1989). The UL2 ORF was not one of those suggested to include an N-terminal signal sequence (McGeogh *et al.*, 1988) however. In the light of this information the observation from work carried out in this thesis is that, of the possible start codons, the second ATG in the ORF is a possible start point of translation (Figure **1.3.2**). The sequence from this start point directs translation of a functional UDGase of typical size (Section **1.2.2**) in a prokaryotic expression system, whereas the sequence from the first ATG in this ORF does not (Savva and Pearl, 1993) (Section **4.2.3**, Section **4.2.4**, Section **4.2.5**, Section **7.1.2**). UL2 is one of 27 rightward oriented ORFs in the U_L region, lying approximately 10 kb from the internal repeat region b, and has been shown to be an early β gene (Mullaney *et al.*, 1989). Detailed maps of the layout of HSV1 in terms of ORF locations, transcription signal sites and gene products have been published previously (McGeogh *et al.*, 1988; Roizman and Sears, 1991).

1.3.3 THE REPLICATIVE CYCLE OF HSV1: AN OVERVIEW

An extensive description of the replication of herpes simplex viruses is given elsewhere (Roizman and Sears, 1991). The following section is an introduction to the replicative cycle of HSV1.

The viral replicative cycle is now reasonably well understood (Roizman and Sears, 1991), and it is clear that the entire process is tightly regulated with the synthesis of viral gene products sequentially co-ordinated in a cascade fashion. The host RNA polymerase II carries out transcription of all viral genes throughout the replicative cycle. Attachment and penetration of the virion are key points in a productive infection. Virions which cannot attach and fuse to the plasma membrane are internalised and degraded in endocytic vesicles. In fully permissive tissue culture cells the total time for the infective/ replicative cycle using whole virions is about 20 hours.

The replicative cycle of HSV1 can be summarised by the following events. The virus attaches to specific cell receptors. This is followed by fusion of the viral envelope with the plasma membrane. Two viral proteins are released in this process, VHS which stops host-protein synthesis, and αTIF which is transported to the nucleus. The capsid is also transported to the nucleus where the DNA is immediately released and becomes circularised.

The transcription of 'immediate early' α genes by host enzymes is induced by αTIF, and the resulting mRNAs are transported to the cytoplasm and translated into the five α proteins. These are then transported back to the nucleus where transcription of the β genes begins. The β proteins are produced as a result of this. These are largely the nucleic acid associated proteins and are produced in two phases, early and late β proteins. At this point the host chromatin is degraded and the nucleoli disaggregate.

Viral DNA synthesis then ensues, the genomes being formed by a rolling circle mechanism resulting in concatamerisation. The transcription of the γ genes and translation of the γ proteins takes place at this time. These proteins are mainly virion structural and coat proteins. Capsid proteins then begin to assemble into empty capsids which slightly precedes the co-ordinated cleavage and packaging of viral DNA into the pre-formed capsids as toroids. Capsids that are successfully loaded with DNA acquire a new protein (Section **1.3.2**) and then bind to patches of accumulated viral glycoproteins and tegument proteins on the cellular membranes and become enveloped. The enveloped capsids amass in the endoplasmic reticulum and are then transported out of the cell.

1.3.4 HSV1 IN LATENCY

HSV1 can undergo latency in the nuclei of neuronal cells of various dorsal root ganglia, where it exists in one or more copies as an episomal closed circular DNA (Mellerick and Fraser, 1987; Deshmane and Fraser, 1989). Reviews of the molecular and clinical aspects of latency in herpes simplex viruses can be found elsewhere (Roizman and Sears, 1987; Roizman and Sears, 1991).

The virus enters sensory nerves that innervate the site of infection and is delivered, as the unenveloped capsid, to the nuclei by slow axonal transport (Lycke *et al.*, 1984; Kristensson *et al.*, 1986). Occasionally the virus reactivates and is carried back down the axon where it re-infects cells close to the original site of infection (Cook and Stevens, 1973). Latency in α-herpes viruses such as HSV1 can be divided into three main stages: 1) Infection of epithelial cells and replication of virus, and entry into sensory nerve endings. It is unclear whether more than one virus infects any one neuronal cell, though it has been observed that the bigger the initial infected epithelial area, the more frequently re-activation and re-infection occur. This would suggest that the number of neuronal cells harbouring virus is the main determinant in severity of subsequent attacks (Roizman and

Sears, 1987).

2) Latency as an episomal closed circular DNA. Replication of the viral genome may occur when DNA synthesis is carried out in ganglional cells, which may result in multiple latent copies of virus DNA in any one neuronal cell.

3) Multiplication of virus and re-infection of tissues innervated by the sensory ganglia harbouring the viral progeny, accompanied by destruction of the neuronal cells involved.

It is thought that once a latent infection results, viruses at newly infected sites are not able to re-infect neuronal cells due to intervention by the host immune system; thus it would appear that replication of episomal viral genomes provides all the virus for subsequent re-activations (Centifanto-Fitzgerald *et al.*, 1982). Owing to the quiescent state of the neuronal cells in which these episomal viral genomes are found during latency (Section **1.3.5**), it is possible that chemotherapy against the viral UDGase may provide a useful anti-viral therapy (Section **7.1.1**). Even though the UL2 gene has been shown to be dispensable for viral viability in cell culture (Mullaney *et al.*, 1989), new evidence indicates that the HSV1 UDGase is important in the maintenance of viral viability during latency (Pyles and Thompson, 1994).

Little is known about the molecular events responsible for herpes simplex viruses entering latency, or remaining in latency, or the exact molecular events that lead to reactivation. A model has been proposed (Roizman and Sears, 1987; Roizman and Sears, 1991), which suggests that the environmental differences in neuronal cells, such as large physical distances between the portal of entry and nucleus, favour a non-replicative state for the virus. The suggestion is that multiplication of viral genomes occurs, but only during trauma that is known to precede re-activation. It is also suggested that multiplication of viral genomes does not lead to reactivation until a threshold number is reached, again dependent on environmental factors. The only things that are known for sure about latency of this virus are that the replicative cycle is arrested sometime after infection of neuronal cells resulting in episomal viral DNA in the nuclei, and a number of variously spliced non-polyadenylated latency associated RNA transcripts, or LATs, which are thought to preclude the synthesis of the immediate early protein $\alpha 0$ (Roizman and Sears, 1991).

1.3.5 URACIL-DNA GLYCOSYLASE AND VIRAL LATENCY

The environment of the human neuronal cells is a quiescent one. The DNA polymerases responsible for replication are present in diminishing quantities in development and are usually absent by birth (Focher *et al.*, 1990; Focher *et al.*, 1993). The same is true for UDGase in these cells. The neuronal cells do contain a polymerase active during repair synthesis, DNA polymerase β, which incorporates uracil and thymine with comparable efficiency. Neuronal cells also contain normal levels of dUTPase, thus incorporation of uracil will not be very frequent, but will nonetheless be statistically likely. The lack of cellular UDGase activity will mean that uracil arising from misincorporation and from cytosine deamination will not be repaired, unless very short patch repair is present, resulting in a lifelong progressive accumulation of uracil in the neuronal cell genome. This process has been implicated in neuronal cell aging (Mazzarello *et al.*, 1990; Focher *et al.*, 1992b). The presence of a viral UDGase would then be of some advantage in maintaining the fidelity of the viral genome, which may be in a latent state for many months or years before reactivation (Focher *et al.*, 1992a).

Though it has been shown previously that the HSV1 UDGase is dispensable for replication of the virus during infection in cell culture (Mullaney *et al.*, 1989), it has recently been proposed that the HSV1 UDGase is necessary for the maintenance of viral viability during latency (Pyles and Thompson, 1994). This is discussed in greater detail in section **7.1.1.**

Recently, it has been shown that UDGase is one of the many DNA modification enzymes encoded by poxvirus genomes (Stuart *et al.*, 1993; Upton *et al.*, 1993). These viruses replicate in the cytoplasm of infected cells, apparently completely independently of the host DNA modification enzymes which are in the nucleus. There is evidence to suggest that the poxvirus UDGase is essential in the viral lifecycle (Stuart *et al.*, 1993; Millns *et al.*, 1994). It appears then, that the removal of uracil from DNA is an absolute requirement in any organism with a DNA genome. From data on rapidly replicating organisms such as viruses, it becomes apparent that damage to genomes that are direct precursors of progeny can result in disastrous effects on that population. Again, from this point of view, it becomes an increasingly attractive notion to think in terms of anti-viral therapy directed against viral UDGases. These possibilities are discussed further in section **7.1.1.** A UDGase encoded by HSV1 strain 17 has been produced in large quantities in a

prokaryotic expression system in an attempt to determine the three dimensional structure (Savva and Pearl, 1993). Such information will be invaluable in the design of specific anti-viral compounds.

2 TECHNICAL DISCUSSION, METHODS, AND MATERIALS

PART 1:

DNA, PROTEIN, AND MICROBIAL TECHNIQUES

This chapter is the first of two in which all the techniques used in the course of this project are discussed. The methods and materials used in the manipulation of DNA, the culture and storage of bacteria, and the purification and crystallisation of proteins are discussed, along with a brief discussion of each technique.

General laboratory equipment and reagents were used throughout. Specific materials relevant to particular parts of each method are referred to in the text. Unless otherwise stated, all reagents were obtained from SIGMA Chemical Company Ltd., BDH Laboratory Supplies, or Fisons.

2.1 DNA MANIPULATION

The manipulation of DNA is a crucial first step in any strategy that is designed to overproduce a gene product. DNA manipulation is also essential for any experiment that requires DNA to be present in a pure form, for example in assays for a DNA binding protein such as UDGase, or in the co-crystallisation of a DNA-protein complex. This section describes all the DNA manipulation techniques utilised during this project, and the rationale behind the use of each technique.

2.1.1 THE POLYMERASE CHAIN REACTION

The polymerase chain reaction (Mullis *et al.*, 1986; Mullis and Faloona, 1987) is a technique that is used extensively in molecular biology, diagnostics and forensic science. It is a very effective way of producing large quantities of a DNA of interest from only picomole quantities of the DNA template, and has recently been exploited as a means of synthesising entire genes *de novo* (Prodromou and Pearl, 1991; Di Donato *et al.*, 1993). The theory of the technique, and its many uses, have been presented elsewhere (Gibbs,

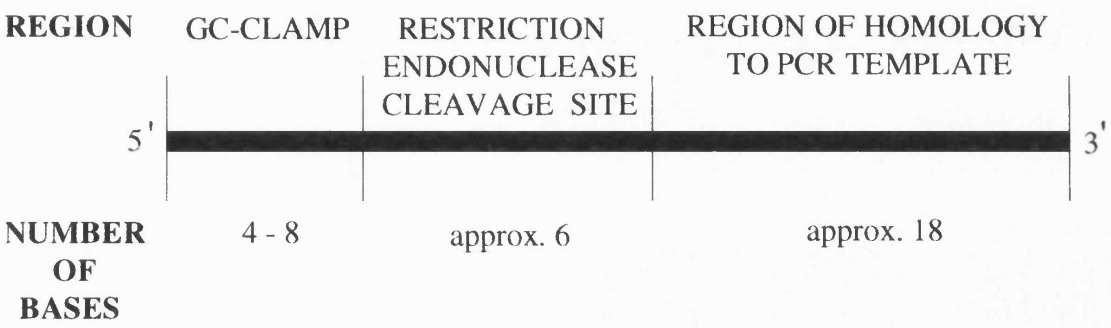
1990; Bej *et al.*, 1991; Timmer and Villalobos, 1993), so this section will deal with the use made of PCR in this project.

PCR was carried out in this project to amplify two genes (Section **4.1.1**) as a preliminary step in the cloning of these genes into expression vectors for overexpression in *E.coli*. In addition, PCR was used to produce a substrate for UDGase by the incorporation of tritiated uridine into the amplified DNA (Section **2.1.16**). The differing nature of the template DNA, and primer sequences used in the two PCR reactions means that the parameters for the reactions were different.

In seeking to amplify a gene for insertion into an expression vector, the use of PCR is an advantage, as convenient restriction enzyme sites can be designed into the primers. Thus a gene can be amplified to be inserted into a plasmid of choice. Considerations in primer design include a good degree of overlap with the gene of interest, giving a melting temperature (Figure **2.1.1**), T_m , of the primer/ template overlap region in the region of 55 to 65°C. In addition, the primer should have the restriction site outside this region, and a guanine:cytosine base-pair rich 'clamp' region at the very end (Figure **2.1.1**) to aid in restriction digestion efficiency, as some restriction endonucleases cut only very poorly at the ends of a DNA strand (Section **2.1.3**).

A Techne PHC2 programmable PCR machine was used to carry out thermal cycling. Factors for consideration in setting up the PCR program include the following. The T_m of the initial primer/ template overlap region must be calculated in order to set a sensible primer/ template annealing temperature. The length of the DNA template and the percentage of guanine:cytosine base pairs in it can affect the amount of time required at the denaturing temperature. Finally, the expected length of the PCR amplified product is related to the length of time that should be set at the extension temperature. A good start point for the annealing temperature is about 2°C higher than the calculated primer/ template overlap T_m; and in general the higher the G:C ratio, the longer the denaturing time. Also, as a rule of thumb, one minute extension time should initially be set per kilobase of the expected product length.

Lastly, the addition of each component of the reaction into the reaction tube should be done so that risk of cross contamination is minimal. Liquid paraffin is layered onto the surface of the solution to prevent evaporation during the high-temperature cycling that will occur. The success of a PCR cannot be judged just by having an amplification product



HSV1 Uracil-DNA glycosylase
Coding Strand Primer incorporating
Bam HI restriction endonuclease cleavage site

GGC GGG ATC CAT GAA GCG GGC CTG CAG C

Non Coding Strand Primer incorporating
Bam HI restriction endonuclease cleavage site

CGC CGG ATC CTC AAA CCG ACC AGT CGA TGG

***E.coli* (*ung*) Uracil-DNA glycosylase**
Coding Strand Primer incorporating
Nco I restriction endonuclease cleavage site

GAG AGC GCC ATG GCT AAC GAA TTA ACC TGG

Non Coding Strand Primer incorporating
*Hin*d III restriction endonuclease cleavage site

CCC GCA AGC TTA CTC ACT CTC TGC CGG GAA TAC

Figure 2.1.1
Diagrammatic representation of the design of a typical PCR primer. Also shown are the sequences of the PCR primers used in this project. Bases in red are the GC-rich clamp region. Blue shows the restriction enzyme site, and green shows the region of homology to the target gene.

of the correct size. The amplified product should always be checked for its fidelity by restriction digestion and/ or sequencing, as some thermostable DNA polymerases are error prone (Dunning *et al.*, 1988). Even with a proof-reading thermostable polymerase, this analysis should still be carried out (Karlovsky, 1990).

A stock mixture of four dNTPs at 25 mM was made up from individual dNTPs supplied at 100 mM by Pharmacia LKB Biotechnology. Thermophilic DNA polymerases were used: a) the 3'→5' proofreading Vent® (New England Biolabs), with the relevant buffer (1x buffer: 10 mM KCl, 20 mM Tris·HCl [pH 8.8 at 25°C], 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1 % Triton X-100) and additional magnesium sulphate solution (Section **4.1.1**). b) the non-proofreading thermostable *Taq* DNA polymerase (Perkin Elmer Cetus), with a relevant buffer (1x buffer: 50 mM KCl, 10 mM Tris·HCl [pH 9 at 25°C], 0.1 % Triton X-100, 10 mM $MgCl_2$) (Section **4.1.1**). In this project genomic DNA (extracted from the *E.coli* strain JM109), and plasmid DNA (pGX23 and pLig1), were used as templates (Section **4.1.1**). Sequences of the primers used in this project are shown in figure **2.1.1**.

2.1.2 AGAROSE GEL ELECTROPHORESIS

The technique of agarose gel electrophoresis, described extensively by Sambrook *et al.* (1989), is used routinely in the analysis of DNA to check its size and physical state. Migration of linear DNA through agarose gels is size based, with fragment sizes being identifiable in terms of their length in bp. The percentage of agarose in the gel determines the size resolution; 1 % is good for fragments in the range of 0.5 to 4.5 kb. Higher percentage gels resolve smaller bands better while lower percentage gels resolve higher molecular mass bands. Reference markers, produced by the restriction digestion of bacteriophage λ DNA, with restriction enzymes, such as *Pst* I, were used to estimate the approximate lengths of migrated bands in adjacent lanes.

Samples were loaded with glycerol-rich loading buffer which lends density. A dye such as orange-G was added to this loading buffer. This migrated with the very low molecular mass DNA, acting as a marker to show when the run should be ended. Migration was achieved by applying a constant voltage (Power Pac 300, Bio-Rad Laboratories Inc.) across the gel (Mini-Sub agarose gel electrophoresis system, Bio-Rad

Laboratories Inc.). DNA was visualised under ultraviolet light in the presence of an intercalating agent such as ethidium bromide, which fluoresces with greater intensity when bound to DNA. Other types of gel electrophoresis of DNA are discussed in sections **2.1.15** and **4.1.2.**

Stocks used were Tris·acetate buffer (TAE), stored as a 50x stock, and used at 1x concentration where Tris·acetate is at 0.04 M and EDTA is at 0.001 M. An alternative buffer used was Tris·borate buffer (TBE), stored as a 5x stock, and used at 0.5x concentration where Tris·borate is at 0.045 M and EDTA is at 0.001 M. Ethidium bromide was stored as a 10 mg/ml stock in a light-proof bottle. Agarose used throughout this project was Type I-A, Low EEO grade, supplied by Sigma Chemical Company Ltd.

2.1.3 RESTRICTION ENZYME DIGESTION OF DNA

The use of restriction digestion in molecular biology has increased with the availability of an ever-increasing range of enzymes that recognise and cleave a variety of different sequences. The more commonly used enzymes recognise 4, 6, or 8 bp stretches, and then cleave them with or without single strand protrusions left over at the 3' or 5' ends. Consequently, the number of potential cut and join operations that can be done on a DNA molecule to modify it are very large. These operations are carried out to change reading frames, and the proximity of genes to promoters, or to create hybrid DNA sequences. Other uses include preparation of size calibration markers for electrophoresis, and mapping sections of DNA.

Digestion was achieved by incubation of a purified DNA sample, with appropriate enzyme(s) and a buffer specified by the supplier (New England Biolabs), at the optimal temperature for the enzyme(s) being used, usually 37°C for 1 or 2 hours. Most enzymes are fairly tolerant of impurities accompanying the DNA, such as proteins, RNA and trace amounts of organic solvents, as well as slight perturbations from the ideal temperature and buffer (New England Biolabs Catalogue 1993/4, Appendices). Restriction enzymes in general, work well in the presence of ribonuclease and deoxyribonuclease enzymes as well as Klenow DNA polymerase and calf-intestinal phosphatase. Restriction enzymes are generally stored in 50 % glycerol at -20°C, and the final glycerol concentration in the reaction should not be more than 5 to 10 % v/v.

2.1.4 OTHER DNA MODIFYING ENZYMES USED

Other enzymes used in this project include T4 DNA ligase (to ligate the 5'-terminal phosphate and 3'-terminal hydroxyl ends of linear or nicked double stranded DNA molecules by forming a phosphodiester linkage), T4 RNA ligase (as an enhancer of T4 DNA ligase activity in DNA ligations), Mung Bean Nuclease (to remove 5' single strand DNA overhangs), Ribonuclease A (to remove RNA from crude DNA preparations prior to restriction digestion), *E.coli* Klenow DNA polymerase (to fill in 5' single stranded DNA overhangs, also to remove 3' single stranded DNA overhangs), calf-intestinal phosphatase (to remove 5' terminal phosphate groups from DNA molecules making ligation impossible unless it is to a molecule which possesses a 5' terminal phosphate group), Sequenase® v2.0 (in sequencing a DNA molecule), Vent® DNA polymerase (in PCR, section **4.1.6**), *Taq* DNA polymerase (in PCR), T4 polynucleotide kinase (to exchange 5' terminal phosphate groups with ^{32}P[ATP] in solution thus radiolabelling molecules of DNA) and T4 DNA polymerase (to remove 3' single strand DNA overhangs). Unless otherwise referenced, the uses of all these enzymes are more fully described elsewhere (Sambrook *et al.*, 1989).

2.1.5 SMALL-SCALE PLASMID DNA PREPARATION

This technique, known as the miniprep, was carried out by the alkaline lysis method, adapted from Birnboim and Doly (1979), and Ish-Horowicz and Burke (1981), as presented in Sambrook *et al.* (1989). This procedure enables the isolation of plasmid DNA from bacteria in quantities and purity sufficient for most enzymatic modification. Some enzymes can be intolerant of DNA prepared in this way, and it is necessary to use this DNA with a further purification step such as Geneclean® (Bio-101 Inc.) (Section **2.1.9**), or instead to prepare the DNA by the large-scale method (Section **2.1.6**), or using Qiagen® (Diagen GmbH) columns (Section **2.1.7**).

Volumes of 1 ml to 5 ml of a saturated *E.coli* culture were pelleted in a 1.5 ml tube in a microcentrifuge. The pellet was washed carefully with distilled water and as much liquid as possible was removed. To the clean pellet, 100 µl of Alkaline Lysis Solution 1 was added and the pellet was resuspended by vortexing briefly. To this was added 200 µl of freshly prepared Alkaline Lysis Solution 2 and this was mixed gently with inversion and turning of the tube so that all inner surfaces of the tube were contacted.

This resulted in cell lysis. The tube was left on ice for five minutes, after which 150 µl of Alkaline Lysis Solution 3 was added. The tube was mixed immediately but gently in order to minimise shearing of chromosomal DNA, and left on ice for 10 minutes. At this stage many proteins and the bulk of the chromosomal DNA are precipitated. The tube was spun in a microcentrifuge at >12 000 *g* and 4°C for 10 minutes.

The supernatant was retrieved and mixed with an equal volume of phenol (equilibrated to pH 8 by washing with Tris·HCl buffer): chloroform: isoamyl alcohol (25:24:1 ratio by volume respectively) by vortexing. This was partitioned into two phases by centrifugation and the upper (aqueous) phase was kept while avoiding proteins that had precipitated at the interface between the phases. This step has the effect of removing proteins by organic precipitation and the chloroform also tends to sequester the majority of the phenol in the organic phase. This step was often repeated a couple of times depending on how much precipitate was observed at the interface on each occasion. Following the final extraction, 0.1 volumes of 3 M sodium acetate pH 5.2 were added to aid DNA precipition, which was accomplished by adding 2.5 volumes of ice-cold ethanol and mixing well. This was kept at -20°C for 30 minutes, or at -70°C for 15 minutes. The tube was then centrifuged at >12 000 *g* and 4°C for 30 minutes, after which the ethanol was drained off and 200 µl of 70 % ethanol was added. The tube was then recentrifuged for 5 minutes at 4°C and the ethanol drained off. The pellet was dried by keeping the tube inverted over tissue paper for 15 minutes, by incubating in a 55°C oven for 5 minutes, or by vacuum in a rotary evaporator for 2 minutes.

The pellet was then resuspended in de-ionised water or TE buffer, and was stored with or without the presence of ribonuclease A at 1 µg/ml at -20°C indefinitely.

Stocks used were: Alkaline Lysis Solution 1, which is 50 mM glucose/ 25 mM Tris·HCl/ 10 mM EDTA pH 8.0; Alkaline Lysis Solution 2, 0.2 M sodium hydroxide/ 1 % SDS, made up just prior to use; Alkaline Lysis Solution 3, which contains 3 M potassium/ 5 M acetate made up as a solution of 60 % v/v 5 M potassium acetate/ 11.5 % v/v glacial acetic acid/ 28.5 % v/v de-ionised water. Also used were phenol pre-equilibrated with Tris·HCl pH 8.0, with the optional addition of 0.1 % v/v 8-hydroxyquinoline (an antioxidant and yellow colorant of the organic phase), a chloroform: iso-amyl alcohol (24:1 by volume) mixture, 100 % and 70 % ethanol, TE buffer at pH 8.0, which contains 10 mM Tris·HCl pH 8.0/ 1 mM EDTA pH 8.0, and finally a stock of ribonuclease A at 10 mg/ml.

2.1.6 LARGE-SCALE ISOLATION OF PLASMID DNA

This method was generally carried out by caesium chloride density gradient purification, and enables plasmid DNA to be isolated in a very pure form from large volume *E.coli* cultures. The procedure presented here is an adaptation of the alkaline lysis method and a polyethylene glycol (PEG) plasmid precipitation method by R.Treisman as presented in Sambrook *et al.* (1989). Briefly, the culture of *E.coli* from 500 ml to 2 litres, was grown by inoculation from a small volume pre-culture, 5 to 30 ml grown at 37°C with agitation, which was at log-phase to saturation (Absorbance at 650 nm > 0.7 and <1.0). The large volume culture was then grown at 37°C with agitation and monitored by taking the absorbance at 650 nm until this was 0.4 to 0.6, or early log phase/ mid log phase.

At this point, chloramphenicol was added, to a final concentration of 170 µg/ml, to prevent protein synthesis in the bacteria by its binding to the 50S ribosomal subunit. Replication of plasmids that have pMB1 or ColE1 replicons ("relaxed" replicating plasmids) continues in the absence of host cell protein synthesis. As a result, the plasmid copy number increases considerably with little or no increase in other cellular components that rely on protein synthesis. This step is necessary with plasmids that have a low copy number, and is useful with high copy number plasmids because the amount of impurity that has to be removed from the plasmid DNA is minimised thus making the preparation easier. The chloramphenicol step is incompatible with plasmids that replicate in a "stringent" fashion, that is relying on host protein synthesis for replication; such plasmids contain replicons such as pSC101.

The bacterial culture was then grown for a further 15 hours at 37°C with agitation, and harvested by centrifugation at 4 000 *g* for 10 minutes at 4°C. The pellet was washed carefully with distilled water and as much liquid as possible was removed. The alkaline lysis procedure was carried out as detailed in section **2.1.5**, with 20 ml of Alkaline Lysis Solution 1, 50 ml of freshly prepared Alkaline Lysis Solution 2, and 50 ml of Alkaline Lysis Solution 3. The subsequent centrifugation was at 8 500 *g* for 30 minutes at 4°C.

The resulting supernatant was then transferred through 4 layers of sterile gauze into a fresh tube and 50 ml of a 50 % w/v solution of PEG 6 000 was added with gentle mixing by swirling the tube. The tube was held on ice for 1 to 2 hours during which time the plasmid DNA precipitated. This DNA was pelleted by centrifugation at 13 500 *g* for

30 minutes at 4°C. The DNA pellet was resuspended in 5 ml of TE buffer pH 8.0, mixed with an equal volume of phenol: chloroform: isoamyl alcohol (Section **2.1.5**) by vortexing, and partitioned into two phases by centrifugation. The upper (aqueous) phase was kept while avoiding proteins that had precipitated at the interface. This was then repeated twice. To the resulting aqueous solution, 1.1 g of caesium chloride was added per ml. This was followed by the addition of 150 µl of a 10 mg/ml solution of ethidium bromide.

The resulting solution was dispensed into 2 Beckman quick-seal polyalomer ultracentrifuge tubes and ultracentrifuged at 50 000 rpm for 20 hours in a Beckman L-7 ultracentrifuge at 20°C using a 70.1Ti rotor. The end of the run was preceded by a drop to 40 000 rpm for 30 minutes in order to slightly relax the gradient formed. Braking was then applied down to 800 rpm. The plasmid DNA was removed using a hypodermic syringe inserted into each tube just below the plasmid band which is visible under long wave ultraviolet light (this is less damaging to DNA than short wave ultraviolet light). The usual volume of the extracted DNA was about 1 to 3 ml, and 10 volumes of TE buffer were added to this to prevent precipitation of the caesium chloride. Ethidium bromide was then removed by several vigorous mixes with caesium chloride-saturated isopropanol. This partitions rapidly on standing, and the upper isopropanol layer was removed each time until there had been two or three rounds where no pink colour was visible in the aqueous layer. The aqueous phase was then ethanol precipitated by the method described in section **2.1.5**. The DNA pellet was resuspended in 1 to 2 ml of TE buffer or de-ionised water. Ribonuclease addition was unnecessary as the DNA was of very high purity. The DNA was stored at -20°C.

Stocks used were chloramphenicol at 34 mg/ml in ethanol, Alkaline Lysis Solutions 1, 2, and 3, a 50 % w/v solution of PEG 6 000, pre-equilibrated phenol, 24:1 chloroform: iso-amyl alcohol mixture, caesium chloride, 10 mg/ml ethidium bromide, and TE buffer pH 8.0.

2.1.7 ISOLATION OF PLASMID DNA USING QIAGEN® COLUMNS

The Qiagen® (Diagen GmbH) tip 20, 100 and 500 columns yield DNA of caesium chloride density gradient quality (Section **2.1.6**). The protocol is supplied with the kit. The initial stages follow a variation on the alkaline lysis method (section **2.1.5**) until the centrifugation step following the addition of Alkaline Lysis Solution 3. After this, the

supernatant was applied to the column which is packed with a resin that preferentially binds DNA. Then there were washing and elution steps before the DNA was finally precipitated, using 0.7 volumes of isopropanol. The tube was centrifuged for 30 minutes at >12 000 *g* at 4°C to pellet the DNA. The pellet was washed with 70 % ethanol and recentrifuged for 5 minutes. The pellet was then dried by one of the methods described in section **2.1.5**, and was resuspended in TE or de-ionised water without the requirement of ribonuclease A. The DNA was stored at -20°C.

2.1.8 EXTRACTION AND PURIFICATION OF CHROMOSOMAL DNA FROM *E.COLI*

This method is adapted from that of Chater *et al.* (1981). Due to the size of *E.coli* chromosomal DNA, it is very prone to shearing forces during manipulation. However, the number of chromosomes from 10 ml of saturated cultures is enormous, and thus breaks in the DNA will not damage all copies of a region of interest. Thus, handling of this DNA for the purposes of PCR from the chromosome does not need to be delicate.

Briefly, 10 ml of *E.coli* JM109 were cultured to saturation at 37°C with agitation. The cells were pelleted at 4 000 *g* for 10 minutes. The pellet was washed carefully with distilled water and as much liquid as possible was then removed. The pellet was resuspended in 1 ml of Alkaline Lysis Solution 1 and left at room temperature for 5 minutes. After this time 100 µl of a 10 % w/v solution of SDS was added and the tube was inverted several times to mix. To the lysate was added 500 µl of a 10 mg/ml stock of proteinase K (Sigma Chemical Company Ltd), followed by the addition of 4 ml TE buffer containing ribonuclease A at 20 µg/ml. The mixture was left at 37°C for an hour, at which point a further 5 ml of TE buffer with ribonuclease A was added, and the mixture was left at 37°C for a further 2 hours.

After this time, 1 ml of 5 M sodium chloride is added and the solution was mixed before layering 15 ml of ice cold ethanol onto the surface. The chromosomal DNA was spooled out with a Pasteur pipette and resuspended with vortexing in 5 ml of TE buffer. This was mixed with an equal volume of phenol: chloroform: isoamyl alcohol (Section **2.1.5**) and left on a shaking table for an hour. The phases were separated by centrifugation, and the upper aqueous layer was removed, carefully avoiding precipitated matter at the interface of the phases, and subjected to two more such organic solvent

extractions. Ethanol precipitation of the DNA was then carried out as outlined in section **2.1.5.** The pellet was then resuspended in 1 ml of TE buffer or de-ionised water with or without the presence of ribonuclease A. This was then stored at -20°C.

The stock solutions were Alkaline Lysis Solution 1, 10 % w/v SDS solution, 10 mg/ml proteinase K, TE buffer pH 8.0, ribonuclease A at 10 mg/ml, 5 M sodium chloride, 100 % ethanol, pre-equilibrated phenol, and finally 24:1 v/v chloroform: iso-amyl alcohol.

2.1.9 GENECLEAN® ISOLATION OF DNA

This method is for use with the kits Geneclean® and Geneclean® II (Bio 101 inc.). The protocol is based on the properties of an amorphous silica resin called Glassmilk® to bind DNA from 0.15 to 10 kb long under high salt conditions, and to elute it only under very low salt aqueous conditions. The procedure has been used in this project both to isolate DNA from agarose gels and to clean up samples of DNA after they have been isolated by miniprep or undergone an enzymatic reaction. All RNA, very small DNA (less than 100 bp), nucleotides, proteins, salts etc. are removed in this procedure, leaving highly purified DNA. Such DNA is readily acted upon in enzymatic modification reactions, the DNA being of similar purity to DNA isolated from a caesium chloride density gradient.

Briefly, 3 volumes of a 6 M sodium iodide solution were added to the DNA. If the DNA was to be extracted from a gel then TAE buffer was used to run it, and the volume of sodium iodide added to such a sample was 3 times the mass in grams of the gel slice containing the DNA. If the DNA was being extracted from a gel then the tube containing the sample with sodium iodide was incubated at 50°C with occasional agitation, until the gel had dissolved. The incubation time at 50°C was kept to a minimum. Once the sample was all in solution, the Glassmilk® was added. About 5 µl per 5 µg was added, if there was more than this then an additional 1µl per 1µg should be added. The tube was vortexed to mix, though shearing became a problem with DNA larger than 5 to 10 kb. This was incubated for 5 minutes on ice, the tube was then pulsed at >12 000 *g* for 5 seconds in a microcentrifuge. The supernatant was discarded and 500µl of New-Wash® solution was added and the pellet mixed into this thoroughly using an air-displacement pipette. The tube was pulsed as before and the supernatant discarded. The New-Wash® step was repeated twice more, with a further spin for 30 seconds at >12 000 *g* after the final supernatant had been discarded. The residual supernatant around the pellet was then

carefully removed with a pipette before the elution step.

Elution was achieved in two or three steps. It is estimated (information from literature supplied with the kit) that the first elution recovers some 80 % of the bound DNA, the second elution about 15 %, and the third recovers between 1 % to 2 %. Thus, the third step is often omitted where the DNA is available in abundance. To elute, half or a third of the final required volume of the eluted DNA was added as de-ionised water or TE buffer. The pellet was resuspended in this and incubated at 50°C for 3 to 5 minutes. The tube was then pulsed at 12 000 *g* for 30 seconds and the supernatant transferred to a fresh tube. The elution was repeated one or two times as required, with the supernatant in each case added to the first. The tube containing the pooled supernatants was then pulsed for 30 seconds, this often revealed a small carry-over of Glassmilk®. The bulk of the supernatant was removed and stored in a fresh tube, leaving the glassmilk behind. The DNA is of high purity and was stored in the absence of ribonuclease at -20°C.

Stocks used in this procedure were saturated sodium iodide solution, Glassmilk®, New-Wash® solution, and (optional) TE buffer pH 8.0.

2.1.10 RAPID ISOLATION OF DNA FROM AGAROSE GELS BY CENTRIFUGATION THROUGH SILICONISED GLASS WOOL

This method is adapted from that of Koenen (1989), and it allows recovery of DNA with reasonable purity and yield. Briefly, a slice of gel was removed with a scalpel and inserted into a 0.5 ml microcentrifuge tube which had a hole in the bottom made with a hypodermic needle. This tube contained a small plug of glass wool covering this hole, such that the bottom third of the tube was taken up by the glass wool. The small tube was inserted into a 1.5 ml microcentrifuge tube. This tube assembly was spun for 10 minutes at 5 000 *g* in a microcentrifuge tube, and the small tube was discarded. The supernatant in the large tube was treated with phenol: chloroform: isoamyl alcohol, and then ethanol precipitated as described in section **2.1.5**. The pellet of DNA was resuspended in de-ionised water or TE buffer, and stored at -20°C.

2.1.11 ENZYME REMOVAL FROM DNA USING STRATACLEAN™ RESIN

Strataclean™ resin (Stratagene) is a silica based resin that removes proteins in a similar way to phenol, except that the resin is a solid, and simply needs to be centrifuged to isolate it, and the proteins it binds, from the DNA sample. A list of enzymes is supplied with the product, and each needs a particular number of extractions with a particular volume of the resin in order to be removed.

After adding the resin and vortexing to mix, the mixture was left for one minute at room temperature. The tube was then spun at >12 000 *g* for one minute, and the supernatant was kept and re-mixed with fresh resin as many times as indicated for the particular enzyme. After this the DNA was ethanol precipitated and resuspended (Section **2.1.5**) in the absence of ribonuclease A. The DNA was stored at -20°C or used in a subsequent manipulation.

2.1.12 RAPID SIZING OF PLASMIDS FROM CRUDE CELL LYSATES BY AGAROSE GEL ELECTROPHORESIS

This technique is described in detail by Barnes (1977), and is a method of comparing the size of plasmids with ColE1 replicons from bacterial colonies picked from a nutrient agar plate. The colonies were rapidly lysed at 70°C in TAE electrophoresis buffer containing SDS, and loaded directly onto a gel. The plasmids migrated between a low molecular weight smear of RNA and protein, and a high molecular weight chromosomal DNA smear close to the wells (Figure **4.1.2 b**). The relative migration of supercoiled plasmids is more reliable in the absence of ethidium bromide, thus the gel was stained after the run is ended (Section **2.1.2**). The pellet from a 1 ml to 3 ml saturated culture may be used instead of a single colony to increase the intensity of the bands on the gel. The method allows the screening of hundreds of colonies a week for relative changes in plasmid size. The technique was successfully used to identify 1 kb and 750 bp increases in size of plasmids of 2.9 kb and 4.3 kb in size respectively on 1 % agarose gels (Section **4.1.3**, Section **4.1.4**).

The stock solutions used were TAE buffer, 10 mg/ml ethidium bromide, and agarose. Also required were Quick-Lysis Buffer, which is a solution containing 5 % v/v of 10 % v/v SDS, 50 % v/v glycerol, 20 % v/v 0.5 M EDTA pH 8.0, 2 % v/v of 10 %

w/v Orange G, 23 % v/v de-ionised water.

2.1.13 DETECTION OF THE PRESENCE OF KNOWN DNA SEQUENCES BY COLONY HYBRIDISATION:NON-RADIOACTIVE METHOD

This method allows bacterial colonies on nutrient agar plates to be screened for a DNA of interest *in situ* by replica plating. This procedure is adapted from that of Grunstein and Hogness (1975). The DNA to be used as a probe is non-radioactively labelled. The non-radioactive system used was the Boehringer Mannheim Non-radioactive DNA labelling and detection kit. The system used by this kit is random-hexamer primed incorporation of deoxyuridine triphosphate (dUTP). The dUTP is linked via a spacer arm to the steroid hapten digoxigenin. Detection of successfully hybridised probes is via an enzyme-linked immunoassay using an antibody conjugate, anti-digoxigenin alkaline phosphatase conjugate, followed by an enzyme-catalysed colour reaction with 5-bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium salt.

Non-radioactive probes were prepared by taking the whole gene of interest, denatured and rapidly chilled on a water-ice/ sodium chloride slurry, or an oligonucleotide complementary to part of the gene. To this was added a hexanucleotide mixture, dNTP labelling mixture and Klenow DNA polymerase. The mixture was then incubated for an hour at 37°C followed by ethanol precipitation (Section **2.1.5**) before being resuspended in TE buffer pH 8.0.

The colony hybridisation procedure was carried out as follows. The *E.coli* colonies on a nutrient agar plate were replica plated using Amersham 'Hybond-C extra' nitrocellulose. This was done by taking a circle of the membrane large enough to cover the surface of the agar and cutting a notch in two places on the very edge. The membrane was placed on the surface of the agar and left for 1 minute to allow adsorption of the bacterial colonies. The notched regions were marked with indelible ink on the corresponding regions of the base of the plate, to permit correct orientation of the developed reaction and the original colonies.

The filter was then removed with tweezers and pressed colony side down onto a fresh Hybond membrane, and notches cut into the second membrane corresponding to those of the first. The second membrane was then placed colony side up on a nutrient agar

plate. Both this plate and the original plate were re-incubated at 37°C for 6 hours to allow the colonies to recover, after which they were stored at 4°C.

The first Hybond membrane was then placed colony side up on a piece of Whatman 3MM filter paper that had been soaked through with a 0.2 M sodium hydroxide/ 1 % SDS solution. This was left until the solution had soaked into the membrane, and had the effect of lysing the bacterial colonies, and denaturing the DNA. The membrane was then carefully transferred, colony side up, using tweezers, onto a second piece of Whatman 3MM paper soaked through with a 1 M Tris·HCl pH 7.4 solution. This was left for 2 minutes and had the effect of neutralising the reagents from the previous step. The membrane was then carefully transferred colony side up using tweezers onto a third piece of Whatman 3MM paper that had been soaked through with a 1.5 M sodium chloride/ 0.5 M Tris·HCl pH 7.4 solution. Again the filter was left for 2 minutes; this step had the effect of continuing the neutralisation process while providing the salt which would allow the DNA from the colonies to be fixed onto the membrane.

The membrane was finally transferred onto a dry piece of Whatman 3MM paper and allowed to air dry. A larger section of Whatman 3MM paper was then folded over such that a half inch spine was created at the fold. The dry membrane was then carefully inserted into this 'envelope' colony side up such that this side did not make contact with the Whatman paper. The envelope was then left in a vacuum oven at 80°C for 2 hours. The baked membrane was then transferred with tweezers into a polythene bag filled with a solution of 6xSSC (20xSSC stock is 3 M sodium chloride/ 0.3 M sodium citrate pH 7.0)/ 0.25 % w/v Marvel skimmed milk powder, and sealed while avoiding the inclusion of air pockets. This bag was then submerged in a 65°C water bath for at least 8 hours. This pre-hybridisation step had the effect of preparing the membrane for hybridisation and lowering the affinity of the membrane to the DNA probe by the interaction of the skimmed milk powder constituents, thus keeping the background low.

The pre-hybridisation solution was then thoroughly expelled from the bag and replaced immediately with 4 to 5 ml of hybridisation solution (5xSSC, 0.5 % w/v blocking agent, 0.1 % w/v N-lauroylsarcosine, 0.02 % w/v SDS) and the labelled DNA probe at a final concentration of 50 to 60 ng/ml. The bag was then re-sealed and placed in a water bath at 65°C for a minimum of 6 hours, during which time hybridisation took place. The plastic bag was then breached and the membrane placed in a plastic sandwich box where two 5 minute washes were carried out at room temperature using 50 ml of 2xSSC, 0.1 %

w/v SDS each time. This was followed by two 15 minute washes at 68°C with 50 ml of 0.1 % w/v SDS, 0.1xSSC. The box was agitated intermittently during these washes.

The following steps outline the antibody and colour reactions. Two 1 minute washes with 40 ml of 100 mM Tris·HCl/ 150 mM sodium chloride pH 7.5 were carried out at room temperature with continuous agitation. The antibody conjugate was added to the membrane as a 1 in 5 000 stock dilution in the buffer used for the previous step, and this was incubated at room temperature with gentle agitation for 30 minutes. After this time, excess unreacted antibody conjugate was removed by two 15 minute washes with 100 ml of 100 mM Tris·HCl/ 150 mM sodium chloride pH 7.5 at room temperature with gentle agitation. The membrane was then equilibrated at room temperature with gentle shaking in the presence of 100 mM Tris·HCl, 100 mM sodium chloride, 50 mM magnesium chloride pH 9.5. Finally the filter was incubated in the dark at room temperature in 10 ml of 100 mM Tris·HCl, 100 mM sodium chloride, 50 mM magnesium chloride pH 9.5, 35 µl of 5-bromo-4-chloro-3-indolyl phosphate and 45 µl of nitroblue tetrazolium salt.

The colour reaction can be overdeveloped and thus must be monitored closely. When the reaction is complete it must be stopped by a 5 minute wash in 50 ml of TE buffer pH 8.0.

2.1.14 DETECTION OF THE PRESENCE OF KNOWN DNA SEQUENCES BY COLONY HYBRIDISATION: RADIOACTIVE METHOD

This procedure, adapted from that of Grunstein and Hogness (1975), is an alternative method to that discussed in section **2.1.13**. Apart from the probe labelling and signal visualisation steps, the two systems are very similar and so the discussion will be brief. In this case, the DNA to be used as a probe is radioactively labelled, by phosphorylation at the 5' end using polynucleotide kinase and γ-^{32}P-labelled ATP.

The radioactive probe was prepared by incubating an oligonucleotide homologous to part of it with ^{32}P-labelled ATP (~10 µCi/µl) in the presence of polynucleotide kinase at 37°C for one hour. The *E.coli* colonies on a nutrient agar plate were replica plated using Amersham 'Hybond N^{+}' nylon membranes using the same method as that described in section **2.1.13**. The membrane was transferred onto several changes of Whatman 3MM

paper as described in section **2.1.13**. The Whatman papers in this case were soaked with, in order of transfer:

1) 1.5 M sodium chloride, 0.5 M sodium hydroxide

2) & 3) 1.5 M sodium chloride, 0.5 M Tris·HCl/ 1 mM EDTA pH 7.2.

The membrane was then washed in 2xSSC and placed colony side up on a piece of Whatman 3MM paper soaked with 0.4 M sodium hydroxide and left there for an hour. The membrane was then rinsed for less than a minute in 5xSSC before undergoing pre-hybridisation at room temperature for 30 minutes in SSarc (7 % w/v N-Lauroylsarcosine in 6xSSC) contained in a sealed polythene bag. The radiolabelled probe was then added and the bag resealed and placed at 65°C for 90 minutes, during which time hybridisation occurred.

The bag was then breached and the membrane pre-washed in SSarc at room temperature before being washed in pre-warmed SSarc (at 5°C less than the T_m of the probe) for 4 minutes. The membrane was then covered in cling film and placed in an X-ray film cassette incorporating an intensifying screen with X-ray film, in the dark. The cassette was sealed and stored at -70°C. Before the film was developed in the dark, the X-ray cassette was allowed to equilibrate to room temperature to avoid condensation interfering with the development process.

2.1.15 SEQUENCING DNA FROM A DOUBLE-STRANDED TEMPLATE

Sequencing of DNA was done by the dideoxy-mediated chain-termination method adapted from Sanger *et al.* (1977) using double stranded plasmid DNA. The method relies on the use of alkaline denaturation of the double stranded molecule, annealing of the sequencing primer, and reaction in the presence of Sequenase® v2.0, a modified bacteriophage T7 DNA polymerase.

The method for template preparation was as follows. From a highly pure sample of plasmid DNA, prepared for instance using a Qiagen column, about 2 µg was diluted to a final volume of 18 µl in a microcentrifuge tube, to which 2 µl of a solution of 2 M sodium hydroxide/ 2 mM EDTA was added. This was left at room temperature for 5 minutes and then placed on ice. Immediately, 8 µl of 0.22 µm filtered 5 M ammonium acetate pH 7.5 was added and topped up with 100 µl of 100 % ethanol pre-chilled to

-20°C. The tube was then immediately transferred to -70°C and left for 30 minutes. The tube was then spun at 4°C for 15 minutes at >12 000 *g*. The pellet was washed with 500 µl of 70 % ethanol and spun for another 5 minutes before being thoroughly drained of supernatant and vacuum dried. The template was ready for sequencing at this point and was stored at -20°C.

The annealing of the primer to the template was carried out by dissolving the pellet in 6 µl of de-ionised water and adding 2 µl of primer and 2 µl of reaction buffer. Annealing was carried out in a water bath at 65°C for 5 to 10 minutes, and the water was then allowed to cool slowly to less than 30°C.

The reaction mix was then prepared in a microcentrifuge tube. To 14 µl of de-ionised water, 1µl of a dNTP mixture containing dATP, dCTP, dTTP, and either dGTP or dITP was added and the contents of the tube were mixed rapidly. The dITP was used where a guanine:cytosine rich region made it difficult to read the sequence clearly due to compression of the bands on the sequencing gel in that region. This is usually due to regions with many guanine:cytosine base pairs or regions with dyad symmetry not being in a fully denatured state during electrophoresis. The dITP forms a two hydrogen bond link to cytosine, thus weakening the base interaction from the three bonds that occur normally in the guanine:cytosine base pair. If more than 14 µl of de-ionised water is present in this mixture, then smaller fragments, closer to the primer, will be favoured in the reaction and if the volume is smaller then larger fragments are favoured.

A volume of 7.5 µl was transferred to another tube and 3.5 µl of 0.1 M DTT, 1 µl of ^{35}S labelled dATP, and 1 µl of Sequenase® v2.0 were added. From this mixture, 4 µl were transferred to the annealed primer/ template mixture and a stopwatch was started. This part of the reaction took place at room temperature and the dNTP mixes were at low concentrations and contained no dideoxynucleotide triphosphates (ddNTPs). During this part of the reaction, the primer was extended slowly with much incorporation of radiolabelled dATP.

Before 5 minutes had elapsed, the mixture was transferred into four microtiter plate wells as 3.5 µl aliquots. Also, the four dNTP/ ddNTP mixes were carefully transferred as aliquots of 2.5 µl onto the side walls of each well, one dNTP/ ddNTP mix per well. When the stopwatch reached 5 minutes, the plate was tapped sharply on the bench to allow the dNTP/ ddNTP mixes to drop into the reaction at the bottom of each well. If the dNTP/

ddNTP mixes were added much earlier than 5 minutes, then smaller fragments were favoured. The plate was then floated in a 37°C water bath for 5 minutes to allow the reaction rate to increase and the chain termination events to occur. If sequence very close to the primer is required, then use of a buffer containing Mn^{2+} ions rather than Mg^{2+} ions will make incorporation of ddNTPs more favourable than normal and thus increase the number of termination events, resulting in an increase in the proportion of shorter fragments.

When 5 minutes had elapsed, the reactions were stopped by the addition of 4 µl of stop mix to each reaction. Once again, this was done simultaneously for each reaction by adding the stop mix carefully to the walls of each well and sharply tapping the plate on the bench to allow the droplets to sink into the reactions at the bottom of each well. Reactions were be stored at -20°C, but their useful lifetime was limited by the half-life of the isotope used, which in the case of ^{35}S is 88 days.

The gel used to resolve the DNA sequence was a 6 % acrylamide gel containing urea, and was poured with a TBE gradient from 0.5xTBE to 2.5xTBE from top to bottom. The gel was pre-warmed at 1 500 volts constant voltage for 30 minutes prior to the loading of sample. If the gel is too hot or cold during the run, the bands tend to smear, thus a metal plate was fixed across the glass plate during the run to act as a heat conductor, ensuring a homogeneous temperature. The temperature was checked occasionally during the run to ensure that the gel did not overheat.

The samples for loading were 2 µl aliquots of each reaction, pre-heated to 80°C or more for 2 minutes. The stop mix contained both bromophenol blue and xylene cyanol dyes, thus the progress of the run could be monitored as to how far the sequence has gone. This enabled multiple loadings of the gel, where small fragments were allowed to migrate off the gel in first few lanes and were retained in other lanes that had been loaded later. Thus some lanes allowed easy reading of the larger fragments while others showed smaller fragments. The gel was removed after the run, and placed in a tank containing a solution of 5 % acetic acid/ 15 % methanol in order to remove the urea. The gel was lifted out of the tank by adhering it to Whatman 3MM filter paper. Any air pockets between the paper and gel were removed by gently waving the paper-backed gel in the tank until the air came out at one of the edges. The gel was then vacuum dried and put in a light-tight X-ray film cassette with a piece of X-ray film, and left at room temperature for 12 hours.

2.1.16 PREPARATION OF URACIL-RICH DNA CONTAINING TRITIATED URACIL, FOR THE DETECTION OF URACIL-DNA GLYCOSYLASE ACTIVITY

The action of UDGase is the hydrolytic fission of the N-glycosidic bond between the uracil base and the deoxyribose in the DNA strand, thus liberating free uracil and leaving behind an apyrimidinic site in the DNA (Section **1.2.2**). Assays have been devised which measure either free uracil (Friedberg *et al.*, 1975), or the unwinding of supercoiled DNA which has been nicked, by fluorometry (Morgan and Chlebek, 1987). The assay used in this project relies on the measurement of free uracil after the reaction is stopped, thus the uracil base must be labelled in order to be detected.

Radiolabelled (250µCi, 21Ci/mmol) 2'-deoxyuridine 5'-triphosphate (dUTP), tritiated at the 5 position in the pyrimidine ring, was obtained from Amersham International Plc as an aqueous solution in 50 % ethanol. The solution was carefully transferred from the vial into a 1.5 ml microcentrifuge tube. A cotton wool plug was inserted at the open end, and the lid was clipped over this. The lid had previously been punctured a few times with a hypodermic needle. Finally, some Nescofilm™ was used to bind the lid to the tube around the edge to prevent the lid from opening. This tube was then placed in a vacuum rotary evaporator, evaporated to dryness, and 5 µl of deionised water added to bring the concentration of the dUTP to 2.4 mM. To this tube, were added 1 µl each of 100 mM solutions of dATP, dGTP, and dCTP, and 2 µl of a 45 mM solution of non-radiolabelled dUTP. This brought the respective concentrations of the dNTPs to 10 mM, and the proportion of radiolabelled uracil to 10 % of total uracil.

The dNTP mixture was added in 0.5µl aliquots to PCR reaction tubes which contained the appropriate primers and template DNA to amplify the *ung* gene of *E.coli.* The PCR was performed in a total of 25µl using *Taq* DNA polymerase for 30 cycles under conditions described in section **4.1.1**, after which another 1µl of *Taq* DNA polymerase was added to each tube and the PCR repeated for a further 30 cycles. This repeat reaction was performed in order to use up as much of the radiolabelled nucleotide as possible, since the limiting factor in PCR is the half-life of the enzyme. Once this was replenished, the reaction could proceed until the other components become limiting thus optimising the yield.

The PCR reactions were pooled and run on a 1 % agarose gel to separate the

substrate DNA from spurious bands, template DNA, primers, enzyme and dNTPs, and the gel was stained with ethidium bromide. The band at ~700 bp was cut out with a scalpel blade and chopped into portions that had masses of 0.4 g or less. The DNA was purified from each portion by the Geneclean® procedure, and the resulting DNA in deionised water from each extraction was pooled. The DNA was quantified by adding 1 µl to 3 ml of 'Ecoscint A' scintillation fluid (National Diagnostics) and counting it in a liquid scintillation counter (Packard Tri-Carb™ 1600-TR Liquid Scintillation Analyzer) in the tritium window, assuming 67 radiolabelled uracil bases per DNA molecule (334 of the 708 bp normally contain thymine). The DNA was stored at -20°C.

In parallel with the preparation of the substrate, a control DNA was also prepared in an identical manner. The difference was that dUTP was not used, and 2'-deoxythymidine 5'-triphosphate (250 µCi, 46 Ci/mmol; Amersham International plc), tritiated in the methyl group at the 5 position of the pyrimidine ring, was used instead. The DNA was prepared to include 100 % radiolabelled thymidine, such that any thymidine release whatsoever could be detected (Section **4.3.2**). This DNA was purified, characterised and stored in an identical manner to the substrate DNA just described.

2.2 MANIPULATION OF *E.COLI*

The preparation of bacteria for the uptake of a DNA of interest, and the subsequent growth and maintenance of the stocks of bacterial clones thus produced is the link between the preparation of DNA outlined in section **2.1** and the isolation of the gene product of interest (Section **2.3** and Section **4.3**). The following procedures were used in this project to culture, transform and to optimise expression of a recombinant protein in cells of the enterobacterium *E.coli*. The exact parameters developed for the *E.coli* which overexpress the HSV1 UDGase are given in section **4.2**.

2.2.1 CULTURE AND STORAGE OF *E.COLI*

To propagate a viable bacterial culture, the following techniques are used. First of all, a sterile environment should be used, provided by working close to a Bunsen burner flame (air inlet fully opened), and vessels should only be open for the minimum required time. Any tubes containing nutrient media should be autoclaved before use, and any

antibiotics or other chemicals that are to be added to the media should be 0.22 µm filtered prior to use.

Bacteria may be stored as stab cultures on a slope of nutrient agar at 4°C, 40 % v/v glycerol stocks at -70°C, or on nutrient agar plates at 4°C. Nutrient agar plates should be re-streaked every 1 to 3 months, depending on the presence of antibiotic. For example, ampicillin will degrade and so stocks with conferred plasmid ampicillin resistance should be re-streaked more frequently than other bacteria. Slope/ stab cultures should be replaced every 4 to 6 months, or much more frequently in the case of ampicillin resistant stocks. Glycerol stocks should be regenerated every 6 to 18 months, regardless of the presence of antibiotic, but depending on the number of times the tube has been thawed. Replica plating is discussed in section **2.1.13.**

The type of media used to grow bacteria can be varied for different tasks. Complex media are usually used, such as yeast-tryptone (YT), Luria-Bertani (LB), or Oxoid No.2. Stronger media are used for greater cell density where yield is a major concern, this can be simply an increase in strength of a complex medium, such as 2xYT or may be a special recipe broth such as Terrific Broth. Terrific Broth is buffered against pH fluctuation and is the medium of choice for fermentation scale growth (5 litres or more).

Growth of *E.coli* is normally carried out at 37°C with aeration, provided by a mechanical agitator such as a shaking table or rotary shaker, or by a sterile air supply in a fermenter. Initially, a growth curve of the strain being used should be constructed, such that the average times of onset of the various growth stages can be assessed. This is done by taking absorbance measurements at 450, 600 or 650 nm in a spectrophotometer against a nutrient medium control sample. Knowledge of a growth curve greatly simplifies the culturing of bacteria, and allows an accurate assessment of the effects that an overexpressed protein has on cell growth during the process of optimisation of expression of the protein. Expression of recombinant protein is discussed in section **4.2.**

2.2.2 PREPARATION AND TRANSFORMATION OF COMPETENT *E.COLI*

Cells that are readily able to take up naked DNA are said to be competent; the technique used to do this in this project is adapted from that of Cohen (1972). Briefly, a

5 ml pre-culture of a chosen strain of *E.coli* was grown to saturation overnight in 1.5 x Oxoid No.2 broth at 37°C. A second 5 ml culture was then grown to mid-logarithmic phase from a 250 µl inoculum of the overnight growth in 1.5 x Oxoid No.2 broth at 37°C. One ml of this culture was used to inoculate 50 ml of 1.5 x Oxoid No.2 broth and this was grown to mid-logarithmic phase at 37°C. The cells were harvested by centrifugation at 4 000 *g* for 10 minutes at 4°C, after which time the supernatant was discarded and the pellet resuspended in 25 ml of ice-cold 75 mM calcium chloride/ 15 % glycerol, near a flame. The cells were held on ice for 10 minutes and then recentrifuged. The supernatant was discarded and the cells resuspended in 3 ml of 75 mM calcium chloride/ 15 % glycerol. The cells were held on ice for a minimum of an hour, up to a maximum of 24 hours after which time they were dispensed into 1.5 ml microcentrifuge tubes as 200 µl aliquots. The bulk of the aliquots were immediately transferred to -70°C storage, where they were found to remain competent, albeit less so with time, for over 18 months.

Transformation was achieved by adding DNA, (from nanogram quantities to a microgram) to the thawed bacterial suspension. The tube was tapped gently to mix in the DNA, and then stored on ice for 45 minutes. The tube was then transferred carefully to a 37°C water bath and left for 10 minutes as a heat-shock step. The tube was transferred back onto ice and left for 2 minutes. A volume of 300 µl of 1.5 x Oxoid No.2 broth was then added to the tube. The tube was flicked to mix the contents and stored at 37°C for 90 minutes with occasional mixing of the contents. The cells were then spread onto two nutrient agar plates as 250 µl aliquots containing the antibiotic of choice. The plates were then incubated at 37°C for 12 to 14 hours. Transformation of cells from the same batch of competent *E.coli* in the absence of any added DNA, constituted the negative control. There should be no colonies on this plate. As a positive control, cells from this batch were transformed with a known quantity of highly purified closed circular DNA.

Transformants were characterised for the presence of the DNA of interest by small-scale preparation of plasmid DNA (Section **2.1.5**) or by rapid analysis of colony lysates (Section **2.1.12**). The plates were replicated for long term storage by replica plating, which is discussed in section **2.1.13**, or by the sterile transfer of colonies one at a time using sterile toothpicks or sterile pipette tips to a numbered grid on a master plate.

2.2.3 NUTRIENT MEDIA AND *E.COLI* STRAINS USED IN THIS PROJECT

The following nutrient media were used during the course of the project (values per litre):

Yeast-Tryptone medium (YT)	*Luria-Bertani medium (LB)*	*Terrific Broth (TB)*
8 g bacto-tryptone	10 g bacto-tryptone	12 g bacto-tryptone
5 g bacto-yeast extract	5 g bacto-yeast extract	24 g bacto-yeast extract
5 g sodium chloride	10 g sodium chloride	4 ml glycerol
	NaOH adjustment to pH 7.0	0.017 M KH_2PO_4
		0.072 M K_2HPO_4

In addition, Oxoid No.2 nutrient broth was used. To achieve a 1x concentration, 25 g are dissolved in a litre of distilled water. All media are made up to the final volume with distilled water. Nutrient agar is made by adding 15 g bacto-agar per litre of nutrient medium prior to addition of water. All media are autoclaved prior to use.

The following *E.coli* strains were used in this project. The genotypes are presented alongside, and the uses made of the strains in this project are detailed at the end of the list. Full explanations of the uses of each strain may be found in Sambrook *et al.* (1989).

STRAIN	*GENOTYPE*
BL21(DE3)	*F⁻ ompT hsdS$_B$ (r_B^- m_B^-) dcm gal* (λcIts857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene *1*)(DE3)
DH5	*supE*44 *hsdR*17 *rec*A1 *endA* 1 *gyrA* 96 *thi*-1 *relA* 1
DH5α	*supE*44 Δ*lac*U169 (ϕ80 *lacZ*ΔM15) *hsdR*17 *recA* 1 *endA* 1 *gyrA* 96 *thi*-1 *relA* 1
JM105	*supE endA sbcB*15 *hsdR*4 *rpsL thi* Δ(*lac-proAB*) F'[*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ*M15]
JM107	*supE*44 *endA* 1 *hsdR*17 *gyrA* 96 *relA* 1 *thi* Δ(*lac-proAB*) F'[*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ*M15]
JM109	*recA* 1 *supE*44 *endA* 1 *hsdR*17 *gyrA* 96 *relA* 1 *thi* Δ(*lac-proAB*) F'[*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ*M15]

The strains DH5 and DH5α (Hanahan, 1983) were often used for the initial transformation of recombinant plasmid DNA following ligation. The strains are

particularly efficient in the uptake of DNA during transformation. Experience in the lab is that DH5α is a good strain for obtaining plasmid DNA to be sequenced, the advantage being that the DNA seems to be cleaner following Qiagen® preparation thus giving a strong and easily readable sequence that is not smeared. The strains JM105, JM107, and JM109 (Yanisch-Perron *et al.*, 1985) were also used to transform recombinant plasmids, as well as to attempt expression of recombinant proteins from plasmids containing *lac* or *trc* promoters. Uses of the *E.coli* strains JM105, JM107, JM109, and BL21(DE3) (Studier and Moffatt, 1986), are discussed further in sections **4.1.2**, **4.1.3**, **4.1.4**, **4.2.0**, **4.2.1**, and **4.3.2**.

2.3 PROTEIN PURIFICATION AND ANALYSIS TECHNIQUES

Purification is a vital step in the detailed analysis of a protein. In crystallographic analysis, it is especially important to use highly purified protein. Substantial purification is attainable with recombinant proteins (whether soluble or insoluble (Sections **4.2** and **4.3**)) because they are often produced at very high levels. The rationale behind the methods used in protein purification is introduced in the following sub-sections. This section lists the materials and stocks that were used to purify the UDGase enzymes of *E.coli* and HSV1, both overexpressed in *E.coli*. In addition, the techniques used to analyse the purity of the protein thus obtained are briefly introduced. More in-depth considerations of these techniques are given elsewhere (Freifelder, 1982; McPherson, 1982; Skopes, 1987). The purification protocols developed are presented in detail in section **4.3**.

2.3.1 THE AIMS OF PROTEIN PURIFICATION

It is possible to remove small molecules and ions from a protein of interest by techniques such as dialysis or ultrafiltration. It is also relatively easy to remove very large species such as membranes and nucleic acids by precipitation techniques, in conjunction with high-speed centrifugation. However, other proteins are more difficult to remove as they are of the same order of size. To this end, it is necessary to remove most, or all other protein contaminants by a series of steps that take advantage of the particular properties of the protein of interest with respect to all other proteins. Parameters used are relative size (achievable by gel filtration chromatography, selective ultrafiltration and selective

dialysis), relative charge (achievable by ion-exchange chromatography), stability in solvents (achievable by organic solvent precipitation), buffers (some proteins exhibit limited solubility and/ or stability in some types of buffer), different temperatures (some proteins are able to survive short periods at highly elevated temperatures whilst maintaining solubility and activity), pH (by observing the effects of pH on protein stability and solubility), and of course, the biological specificity of the protein of interest (achievable by affinity chromatography). Sometimes, it is not possible to fully isolate the protein without resorting to unfolding of the tertiary or secondary structure and attempting to reconstitute this at a later point.

2.3.2 COMMON METHODS EMPLOYED IN PROTEIN PURIFICATION

Several common methods can be employed to separate a protein of interest from the thousands of other cellular proteins present following cell lysis. After cells have been harvested, lysed and centrifuged at high speed (Section **2.3.4**), two distinct cellular fractions result. The soluble component, referred to as the supernatant, is the fraction to which the following methods are most easily applied. The insoluble fraction, which is condensed into a pellet, must be pre-treated with detergents and/ or denaturing agents in order to resolubilise the aggregated proteins and separate them from the mass of lipid, carbohydrate, nucleic acid and ribosomal matter before the following methods can be carried out. Before the common methods used in this project are discussed, the resolubilisation of aggregated proteins from the cell pellet will be dealt with.

The method by which insoluble protein components of the cell are resolubilised, and then re-folded is often worked out by trial and error, though a 'standard' method would be along the following lines: the pellet recovered after centrifugation should be washed by resuspending in the original buffer and recentrifuging, discarding the supernatant, to make sure that any residual soluble protein is removed. The pellet can then be treated in a number of ways. Firstly, resuspension of the pellet in a denaturing solution can be carried out immediately. Solutions of urea up to 8 M, or of guanidine hydrochloride up to 6 M are the most commonly used denaturants. Alternatively, non-ionic detergents such as Triton X-100, or N-Lauroyl sarcosine, or ionic surfactants, can be added at low percentage weight for volume to resolubilise only selected proteins. The gentlest alternative is to resuspend the pellet in a solution of 1 or 2 M sodium chloride,

and to analyse the resulting supernatant for the protein of interest. If the protein of interest is not in this soluble fraction, a number of other proteins will be, due to a 'salting-in' effect (Skopes, 1987), thus it is a useful purification step.

The use of the gentler steps first, in the case of a protein that remains insoluble, is an aid in purifying it prior to resolubilisation of all remaining proteins in denaturant solutions. Nucleic acids that are present in the cell pellet are often lost at the initial pellet washing stages, though they will eventually be completely separated from the proteins during the purification stages. The membrane-associated components can usually be removed by ultracentrifugation after the proteins have been resolubilised in denaturant solutions. However the pellets are often not very compact and particulate carry-over is a problem. Thus, recentrifugation is often carried out on the supernatant fraction. Once the proteins have been resolubilised and cleared of all particulate contaminants, they can often be purified in much the same way as ordinarily soluble proteins. Exceptions to this will be mentioned in the relevant sections.

The following is a brief list of commonly utilised techniques in protein purification that were used during this project, presented according to the property of the protein that the method exploits. These methods can be divided into two main groups: chromatographic methods, and batch methods.

1) Separation on the basis of solubility in high salt:

Batch technique:

Ammonium sulphate fractionation:

This technique relies on the fact that some proteins exhibit a 'salting-out' effect (Skopes, 1987) when the ionic strength of a solution, in which they are present as dissolved particles, increases to a certain molarity. At this point, the competition between salt ions and protein for water in which to be dissolved favours the salt ions. Beyond this point, the protein is no longer in solution and forms a precipitate. The more exothermic this transition is, the more likely it is that the precipitated proteins will suffer thermal denaturation. If the precipitation is a gentle and gradual process however, the protein will maintain its conformation and enough water to remain stably folded. With organic solvents, the effect of precipitation is very exothermic, and consequently the precipitation must be done at very cold temperatures; -20°C is typical for a precipitation using acetone for example (Skopes, 1987). Precipitation in organic solvents is not very reproducible due

to the large destabilising influences inherent in this form of precipitation, thus salt-based precipitation is the favoured method.

Ammonium sulphate is the salt of choice for the task of precipitating proteins selectively, because of the very gradual effect of precipitation (Skopes, 1987). The majority of proteins will precipitate in 40 % to 60 % saturated solutions of ammonium sulphate, though some will precipitate at lesser, or greater, % saturation values. The method is particularly good at removing cell debris such as membranes, associated proteins, and ribosomes at below 20 % saturation. The technique is reproducible if addition of the ammonium sulphate is slow, under stirring and at 4°C. The mixture of protein should be of constant concentration each time the method is carried out for best results.

The technique was performed as a series of 'cuts', punctuated by partition of soluble and insoluble proteins at each stage using high-speed centrifugation. When contaminating proteins had precipitated, the pellet was discarded and the supernatant was again stirred at 4°C while ammonium sulphate was added slowly, until the next desired cut-off point is reached. When the desired protein was precipitated, the supernatant was discarded and the protein in the pellet re-dissolved in a reasonable volume of stabilising buffer. Dialysis to remove residual ammonium sulphate was then carried out. The best cut-offs were determined by trial and error, and the points of isolation of the desired protein were determined by visualisation on a stained polyacrylamide gel and/ or an assay for biological activity in the redissolved sample.

2) Separation on the basis of molecular size:

a) Chromatographic techniques:

Gel-filtration:

This technique is accomplished using column chromatography. The chromatographic matrix is composed of porous beads which must be pre-swollen in the buffer to be used. Once swollen, the beads have an average pore size that enables proteins of a range of molecular masses between about 3 000 to 70 000 to pass through them. Proteins larger than the cut-off for this range pass outside of the beads and are the first to emerge from the column. The larger of the proteins that pass through the pores are the first to emerge in the resolved range, the smallest proteins are held up in the matrix the longest and are the last to emerge. Buffer exchange is possible because of this effect. The

longer the column, the better the resolution, because the number of theoretical separation plates down the length of the column is greater (Freifelder, 1982). Gel-filtration beads are available in a number of pore sizes, giving a choice of optimum separation ranges. Most bead types can tolerate denaturing agents, and so can be used with both supernatant and resolubilised pellet fractions. Very fine beads packed at high density enable high pressure/ high speed/ high resolution runs to be performed. The drawback with this method is that proteins within about 5 000 molecular mass of each other are often poorly resolved in the most common molecular mass range of 20 000 to 50 000. Another drawback is that unless a large, expensive, and thus slow, column is used, the sample size in terms of volume must be small for optimal resolution. This makes multiple runs or high protein concentrations unavoidable.

b) Batch techniques:

i) Dialysis:

The separation of proteins on the basis of molecular size can be achieved by selective dialysis. Visking tubing has a molecular mass cut-off of 14 000 to 16 000, thus proteins with smaller masses will be lost over several changes of dialysis buffer. The membrane allows free passage of all solutes smaller than the average pore size, and the contents of the dialysis bag and the exchange buffer tend toward equilibrium. The solutes toward the upper limit for passage are lost far more slowly than very small solutes such as ions. Thus the main use for dialysis is buffer exchange.

ii) Pressure ultrafiltration:

In common with dialysis, the basis for separation is a porous membrane which allows the escape of solutes less than the molecular mass cut-off. The membrane is the only outlet for the solution in the pressure cell, and this method is generally used for the concentration of protein samples. The membranes come in a wide range of molecular mass cut-offs, and thus give more scope for the coarse separation of proteins in terms of molecular size than does dialysis.

3) Separation of proteins on the basis of molecular charge:

a) Chromatographic techniques:

Ion-exchange:

The exploitation of the charge of a protein in a buffer of known pH is a very good means for the separation of that protein from proteins with an opposite charge. Unlike gel-filtration which separates most protein of one particular molecular size from most of other

sizes, ion-exchange can separate *all* protein of one charge from *all* protein of the opposite charge at the pH of interest. Also, depending on the strength of the charged group attached to the chromatographic matrix, the degree of charge of the bound proteins can be exploited to separate them by means of a gradient elution that relies on ionically competing the proteins off the matrix. The more strongly bound proteins will remain bound under higher ionic strengths than the more weakly interacting proteins. For very highly charged proteins at a particular pH, weaker ion-exchange resins can be used to exploit the degree of charge. Likewise, with more weakly charged proteins, stronger ion-exchange resins give very good results. Like gel-filtration matrices, the ion-exchangers can be obtained in very fine high density/ high pressure/ high speed / high resolution designs. There is rarely a problem with overloading a column in most laboratory scale ion-exchange, and loading volume is not critical. The sample must be in a low ionic-strength buffer at the outset however.

b) Batch techniques:

Ion-exchange:

As well as their usefulness in column chromatography, ion-exchange resins can be used to further clarify a soluble protein fraction, for instance a supernatant following cell lysis, by batch mixing. The supernatant can be added to the ion-exchanger, which has been pre-equilibrated in the appropriate buffer and mixed for a reasonable time to allow binding to take place. If the protein of interest is known not to bind the ion-exchanger, then the protein can be removed simply by centrifuging the mixture and recovering the supernatant. If the protein does bind, then the supernatant can be discarded, the ion-exchanger washed with the appropriate ionic-strength buffer to allow elution, and the centrifugation step repeated. This method does not have the resolution of column chromatography, but it is far quicker to perform. If the parameters for ion-exchange purification have been well defined by column chromatography, then the use of a batch method can be extremely useful in speeding up a purification. This can be critical with proteins that are found to rapidly lose activity due to instability.

4) Separation of proteins on the basis of complex molecular interaction:

Chromatographic techniques:

i) Affinity:

Affinity chromatography applied to protein purification, in the truest sense, refers to the exploitation of binding interactions between a protein and a biologically relevant

ligand of the protein. However, the term is commonly used in a much looser manner to describe the binding of the protein of interest to any other complex molecule when that binding is not strictly of a simple surface charge type interaction, such as in ion-exchange chromatography. Due to this definition, the more diverse binding phenomena such as dye-interaction chromatography and immobilised metal affinity chromatography come under this banner. Strictly, such techniques should more appropriately be grouped under pseudo-affinity methods; however, for simplicity, dye-interaction will be mentioned in this section. Immobilised metal affinity chromatography is dealt with in section **2.3.3.**

The binding of a protein to an immobilised ligand by ligand-specific interaction allows the separation of that protein from contaminants that do not share the affinity for the immobilised ligand. As such, this technique is potentially the most powerful method available in protein purification. For the technique to be truly useful however, the interaction should be loose enough to be disrupted without having to resort to denaturing the protein to elute it. Sometimes, this is unavoidable in order to achieve the best purification. The many potential non-specific interactions of other proteins with the immobilised ligand also mean that the purer the preparation with respect to the protein of interest prior to this step, the better the end result in terms of absolute purification. The affinity method can also be performed in batch binding.

ii) Dye-interaction:

The exact mode of binding of a protein to an immobilised dye molecule will differ according to the type of molecular interaction that occurs. Some dyes may be acting as pseudo-ligands for the protein of interest, for example cibacron blue F3GA with nucleotide-binding proteins, whereas in many cases the exact interaction is not known (Skopes, 1987). The interactions may be simply ionic, or may involve the interaction of the dye with a group of surface residues in a manner reminiscent of an active site interaction, even though these residues are not biologically significant in the protein's key biological function. Due to this unpredictable array of potential interactions, the use of dye-interaction methods is very much a trial and error approach, and test-kits containing several types of immobilised dye are available (SIGMA Chemical Company Ltd.). As with ion-exchange and affinity, this method lends itself well to batch binding.

2.3.3 IMMOBILISED METAL AFFINITY CHROMATOGRAPHY

Although many functional groups are found to be involved in metal binding in metalloproteins, the case is far simpler when looking for general affinity to metal ions as a means of separating proteins. The technique of immobilised metal affinity chromatography takes advantage of the dominant binding characteristics to first row transition metal ions by the side chain of the amino acid histidine (Hemdan *et al.*, 1989; Arnold, 1991). It has been shown that histidine is the major factor in retention of protein on chelated metal columns of the Ni^{2+}, Cu^{2+}, Zn^{2+} and Co^{2+} types, and studies with engineered proteins have shown that increasing the number of histidines, and also moving histidines to particular environments within the protein can significantly enhance protein retention on such columns (Arnold, 1991).

The most effective columns are prepared by chelating the metals with iminodiacetate, a "tridentate" chelator, resulting in 4-6 co-ordinate ligands per metal ion. The remaining co-ordination sites are filled by water or buffer and can be displaced by the protein as it binds through histidines interacting with the immobilised metal ions. The protein may be eluted from the column by use of a pH gradient or free metal ions to compete with the column for the protein, or by use of imidazole which competes with the protein for the metal sites on the column. Regeneration of the columns is achieved by replenishing the chelated metal.

Effective purification in a single step is possible because histidine is a relatively sparse component of most globular proteins, accounting for 2 % of total residues on average (~5 % in *E.coli* UDGase and ~4 % in the HSV1 UDGase product of the second candidate UL2 ORF). Thus it might be expected that if a particularly histidine rich region were engineered onto a protein it would be retained on a metal affinity column more strongly than native cellular proteins. Particularly good histidine binding regions are found to be the *His*-X_3-*His* containing α-helices which bring two histidines into close proximity, and linkages of six histidines at N-termini or C-termini of proteins (Arnold, 1991).

2.3.4 METHODS FOR CELL HARVEST, LYSIS, AND SEPARATION OF THE SOLUBLE AND INSOLUBLE FRACTIONS

E.coli cells can be harvested from liquid growth media by centrifugation at 3 000 *g* for 5 to 10 minutes. The cell pellets can then be rapidly frozen to -70°C and stored for months at this temperature, or they may be immediately resuspended in a buffer that is compatible with the biological stability of the protein of interest. This buffer should include a serine protease inhibitor such as phenylmethanesulphonyl fluoride present at levels of 0.1 mM to 1.0 mM to inhibit proteolytic degradation, and may contain other protease inhibitors such as EDTA and EGTA if the protein of interest does not require metal co-factors for its structure or function. Cell lysis is usually by mechanical disruption, for example, in a French press, or by ultrasonication. Where it is compatible with the purification protocol, hen egg white lysozyme may be added to the resuspended cells prior to this to weaken the cells and increase lysis efficiency. Lysozyme/ freeze/ thaw/ ultrasonication is an example of such a method. In the case of proteins that bind nucleic acids, yields of the protein may be significantly increased if the nucleic acids are precipitated by the addition of 0.1 volume of 10 % w/v streptomycin sulphate and incubated at 4°C or on ice for a minimum of 30 minutes.

The first steps in the purification of a protein of interest from whole cells is to determine its solubility. Insoluble proteins are sometimes found to be aggregated due to the choice of buffer for cell lysis or a lack of metal in the case of metalloproteins. These types of protein aggregation are avoidable. However protein insolubility can be due to failures in protein folding and/ or deposition as aggregated inclusion bodies in the cell. This can sometimes be overcome by slowing the growth of the cells that are producing the protein of interest, but it is often the case that the only remedy is to attempt an unfolding/ refolding protocol.

In the case of a soluble protein, the next step is to remove insoluble material, including precipitated nucleic acids and/ or precipitated membranes and carbohydrates (which are in any case usually in the insoluble or centrifuge-deposited fraction) by high-speed centrifugation. This step may follow the further precipitation of some proteins and other cellular components by the use of ammonium sulphate. The clarified supernatant can then be processed further (Section **2.3.2**).

2.3.5 BUFFERS AND MATERIALS USED IN CELL LYSIS AND SEPARATION OF THE SOLUBLE AND INSOLUBLE CELLULAR FRACTIONS OF *E.COLI*

Cell lysis buffers used in the optimised protocol:

1) HSV1 UDGase, native protein from pTrc99A (pTS106·1)
   20 mM Tris·HCl / 10 mM EDTA pH 8.25
   15 % v/v glycerol
   1.5 mM DTT
   1 mM PMSF

2) HSV1 UDGase, fusion protein from pRSET B (pBT1)
   100 mM Tris / 50 mM MES / 50 mM Acetic acid , NaOH to pH 8.0
   500 mM sodium chloride
   15 % v/v glycerol
   1 mM PMSF

3) *E.coli* UDGase, native protein from pTrc99A (pEU234·2)
   20 mM Tris·HCl pH 7.8
   1.5 mM DTT
   1 mM PMSF

Sonication was used to lyse cells (MSE soniprep 150) , buffers 1 and 2 were then modified by addition of 0.1 volume 10 % w/v streptomycin sulphate. Centrifugation was carried out in a Sorvall SS34 rotor/ RC5B centrifuge.

2.3.6 PURIFICATION BUFFERS AND CHROMATOGRAPHIC MATRICES USED IN THIS PROJECT

All column steps were carried out at 4°C using the Econo-system low pressure chromatography apparatus (Bio-Rad Laboratories Inc).

1) HSV1 UDGase, native protein from pTrc99A (pTS106·1)

All columns were pre-equilibrated to a stable base-line using a buffer identical to the cell lysis buffer except that it contained only 0.1 mM PMSF. The bound protein elution was carried out in each case with the cell lysis buffer modified to include 2 M sodium chloride and PMSF at 0.1mM. Dialysis steps between column steps were done against cell lysis buffer modified to contain PMSF at 0.1 mM.

Columns packed:

a) 5 cm (internal diameter) x 10 cm (length) low pressure Econo-column (Bio-Rad Laboratories Inc) casing packed with Whatman DE52 pre-swollen microgranular DEAE-cellulose anion exchanger. The packing material was prepared as follows. A 500 ml beaker was filled to the 250 ml mark with the cellulose, de-ionised water was added which reduced the moist cellulose into a paste, addition of water took the total volume reading in the beaker to 300 ml. The beaker was swirled gently for 1 minute to thoroughly wet the cellulose, and the cellulose was allowed to settle under gravity for 10 minutes. The fluid was drawn off slowly with a pipette. When very little fluid remained above the surface of the wet cellulose, 1 M Tris·HCl pH 8.25 was added and the procedure just outlined was repeated. This was repeated twice more with 0.1 M Tris·HCl pH 8.25, and then once more with the column equilibration buffer. Finally, column equilibration buffer was added to the 300 ml mark on the beaker and the beaker swirled to form a very loose slurry. This was poured slowly into the open column casing and allowed to settle under gravity. The column was then ready for equilibration and use. The column is re-useable, and is cleaned and stored in 0.1 M Tris·HCl pH 8.0, 2 M sodium chloride at 4°C.

b) 2.5 cm (internal diameter) x 20 cm (length) low pressure Econo-column (Bio-Rad Laboratories Inc) casing packed with carboxy-methyl cellulose pre-swollen microgranular cation exchanger (Sigma Chemical Company Ltd). The packing material was prepared in an identical manner to the DE52 anion exchanger. The column is re-useable and is cleaned and stored in an identical manner to the DE52 anion exchanger.

c) 5 x Affigel-Blue Econo-Pac cartridges ((5 ml matrix) Bio-Rad Laboratories Inc) daisy-chained together via Luer-lock connections. The columns are prepared for use, and

cleaned and stored as outlined in the literature accompanying them. Briefly, A pre-wash in an isopropanol rich buffer is done to elute residual dye that has dissociated from the column. This is followed by a wash through with equilibration buffer/ 2M sodium chloride, then a wash through in equilibration buffer. The sample is loaded after a stable base-line is attained, and is eluted as required. The column is then washed again in equilibration buffer/ 2M sodium chloride, followed by a 2 M guanidine hydrochloride wash, and is finally washed with 0.05 M sodium phosphate buffer pH 8.0/ 0.02 % w/v sodium azide. The columns are then stored at 4°C.

d) 1.5 cm (internal diameter) x 30 cm (length) low pressure Luer-lock system chromatography column (Sigma-Chemical Company Ltd) casing packed with Polyuridylic acid Sepharose 4B affinity matrix (Sigma Chemical Company Ltd). The matrix needs to be swollen for 24 hours at 4°C in low salt buffer, such as the equilibration buffer. The column is then packed with the slurry and several bed volumes are passed through to remove saccharide-based stabilisers. Once this is done, the column is ready for use and is cleaned and stored in the same manner as the DE52 anion exchanger.

2) HSV1 UDGase, fusion protein from pRSET B (pBT1)

a) A 5 ml bed volume PROBOND® (Invitrogen Corporation) immobilised nickel affinity column was poured as a slurry (50 % v/v) into a 0.9 cm (internal diameter) x15 cm (length) K 9/15 jacketed column casing (Pharmacia LKB Biotechnology), and equilibrated with the cell lysis buffer to achieve a stable base line.

b) Elution was achieved using the cell lysis buffer modified to include imidazole at the desired concentration, and re-adjusted to pH 8.0.

3) *E.coli* UDGase, native protein from pTrc99A (pEU234-2)

The protein is deposited as insoluble inclusion bodies, thus solubilisation and re-folding buffers are required prior to the single column step.

a) Pellet wash buffer was prepared as for the cell lysis buffer, except that it contained

no PMSF, and contains 2 M sodium chloride.

b) Protein solubilisation was carried out using deionised 8 M urea. The deionisation was achieved by stirring the urea for 30 minutes at room temperature in the presence of analytical-grade mixed-bed resin AG®501-X8 20-50 mesh (Bio-Rad Laboratories Inc).

c) Protein refolding was carried out in the cell lysis buffer, modified to include 10 % v/v glycerol and 50 mM sodium chloride, with PMSF at 0.1 mM.

d) The polyuridylic acid sepharose 4B column described in (1d) of section **2.3.6** was equilibrated with the cell lysis buffer modified to include 10 % v/v glycerol, and PMSF at 0.1 mM. The refolded protein was diluted 1 to 1 with this column equilibration buffer prior to loading.

e) The bound protein was eluted with the column equilibration buffer modified to include 2 M sodium chloride. Dialysis was against the column equilibration buffer without glycerol.

2.3.7 CONCENTRATION OF PROTEIN SAMPLES

Samples of protein with volumes greater than 2 ml were concentrated in high-pressure stirred cells (Amicon corp) with 50-65 p.s.i pressure supplied by a nitrogen cylinder. The molecular mass cut off of the membrane (YM-10 (Amicon corp)) in the cell being used was ~10 000. For small volume (0.2 to 2 ml) concentrations, the samples were centrifuged at 7 350 *g* in centricon SR3 microconcentrators (Amicon corp). The molecular mass cut off of the membrane in these units is ~3 000. An example of the use of these devices is given in section **5.1.1.**

2.3.8 ANALYSIS OF PROTEIN BY DISCONTINUOUS SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS

This technique, which can be abbreviated to SDS-PAGE, is an adaptation of the method of Laemmli (1970). This technique exploits the fact that proteins denatured with

SDS will migrate in an electric field through a polyacrylamide gel in proportion to their molecular masses (Shapiro *et al.*, 1967; Weber and Osborn, 1969).

A discontinuous gel system is employed, consisting of two portions, the stacking gel and the resolving (or separating) gel. The stacking gel consists of a large pore-size polyacrylamide matrix at a low pH relative to the resolving gel and running buffer. The proteins are not markedly restricted as they travel through the stacking gel, and their relative mobility in the electric field is intermediate between the buffer ions in the stacking gel below and the running buffer ions above. As proteins and running buffer begin to enter the stacking gel the proteins begin to stack up into a thin band that travels between the stacking gel buffer below and the advancing running buffer. At the interface between the stacking and resolving gels the proteins begin migration on the basis of molecular mass from the thin stack they have formed. The degree of separation of proteins of differing molecular masses is determined by the percentage of polyacrylamide in the resolving gel matrix. The higher the percentage, the smaller the pore size and the more difficult it is for large proteins to migrate at an appreciable rate through the matrix. Thus small proteins are well resolved in higher (15-20 % polyacrylamide) percentage gels, and vice-versa for larger proteins (7-12 % polyacrylamide).

In this project, gels were run with a constant current (SE600 PAGE system, Hoeffer Scientific; Mini-Protean II PAGE system, Bio-Rad Laboratories Inc.; Power Pac 300, Bio-Rad Laboratories Inc.). Gels were stained following the run with the protein binding dye Coomassie Brilliant Blue R (SIGMA Chemical Company Ltd.), which is sensitive to approximately 1µg of protein. Gels were destained to remove dye in the gel matrix that was not bound to protein. This was achieved with a 40 % v/v methanol/ 15 % v/v acetic acid mixture. Alternatively, gels were silver stained, this method is sensitive to just 10 ng of protein. Kits were used to achieve this (Bio-Rad Laboratories Inc). Gels were stored submerged in 5 % acetic acid or were dried onto Whatman 3MM paper under vacuum.

The stocks used were SDS running buffer, which was stored at 10x concentration and used at 1x concentration where Tris·HCl is 25 mM and pH 8.3, glycine is 250 mM, and SDS is 0.1% w/v. Also used were SDS sample loading buffer, stored at 2x concentration and used at 1x concentration where Tris·HCl is 50 mM and pH 6.8, β-mercaptoethanol is 100 mM, SDS is 2% w/v, bromophenol blue is less than 0.25% w/v, and glycerol is 10 %. A stock solution of 30 % acrylamide (30 % acrylamide

: 0.8 % bis-acrylamide) was used, the catalyst for polymerisation of acrylamide/ bisacrylamide was *N, N, N', N'*- tetramethylethylenediamine (TEMED) in addition to a 10 % w/v solution of ammonium persulphate. Buffers for making up the gels were 1.5 M Tris·HCl pH 8.8 for the resolving gel, and 1 M Tris·HCl pH 6.8 for the stacking gel. A 10 % w/v solution of SDS was also used.

2.3.9 METHODS USED TO DETERMINE PROTEIN CONCENTRATION

The concentration of protein throughout the purification of HSV1 UDGase was determined by the use of the Bio-Rad protein assay, which is based on the method of Bradford (1976). Pure UDGase was found to be quantifiable by spectrophotometric methods, where 1.57 absorbance units at 280 nm were due to approximately 1 mg/ml of UDGase (following dialysis after the final poly-U sepharose column step in the purification) as judged by the Bio-Rad protein assay.

Briefly, 800 µl of the protein sample (which may require prior dilution in the buffer that the protein is dissolved in) was added to 200 µl of the neat Bradford reagent. This was then incubated at room temperature for a minimum of 15 minutes, and not more than 50 minutes. The absorbance at 595 nm was then recorded for each sample and the concentration determined from comparison with a standard curve constructed using bovine serum albumin. The assay is sensitive for concentrations of protein up to 20 µg/ml. A variation of the assay is sensitive to up to 200 µg/ml. This is achieved by diluting the Bradford reagent 1 in 5 prior to addition to the protein sample.

2.3.10 AN ASSAY TO DETERMINE THE PRESENCE AND EXTENT OF URACIL-DNA GLYCOSYLASE ACTIVITY

Using the substrate prepared as described in section **2.1.16**, the presence and relative extent of UDGase activity was determined using the following assay method.

The assay was carried out in a total volume of 50 µl per measurement, of which 47 µl was the buffer, at 50 mM and of known pH at the assay temperature (default buffer was Tris·HCl pH 8.0 at 37°C), 1 µl was 100 mM DTT, 1 µl was the substrate DNA, and 1 µl was the enzyme suitably diluted. In order to stably dilute the enzyme to very low

concentrations, the following dilution mixture was used: 0.3 M NaCl, 50 mM Tris·HCl pH 8.0, 1 mM Na·EDTA, 1 mM DTT, 0.01 % w/v bovine serum albumin (BSA) (Fraction V, SIGMA Chemical company).

The assay was started by adding the ice-cold enzyme to the assay mixture, pre-incubated at the assay temperature for two minutes, and mixing thoroughly with a pipette for 10 seconds. If the enzyme required dilution, this was carried out immediately before the addition of the enzyme to the reaction mixture. The timer was started immediately before the addition of the enzyme to the reaction mixture. The assay was run either for 10 minutes, when looking for enzyme activity during a purification, or as a time course when determining relative activity. In such a time course, two independent assays were run under the same conditions, and two samples per time point were withdrawn (550 µl initial reaction volume) and quenched. The time points are 2, 4.5, 7, 10, and 20 minutes. The linear part of the reaction was difficult to track, but could be approximated using time points less than 4.5 to 7 minutes (Section **7.2.1**). The control reaction was to set up one tube with a 50 µl reaction volume utilising the control radiolabelled thymine containing DNA, and to quench it at 20 minutes.

The reactions were stopped by removing 50 µl aliquots from the reaction and adding them to 50 µl of the quenching mixture, 0.5 mg/ml bovine serum albumin (Fraction V, SIGMA Chemical Company), 0.5 mg/ml denatured and sonicated DNA from salmon testis (SIGMA Chemical Company), 0.5 M sodium acetate pH 5.2, in a 1.5 ml microcentrifuge tube. To this was added 200 µl of ethanol pre-chilled to -20°C, and the tubes were agitated thoroughly to mix and immediately placed at -70°C for 20 minutes. The tubes were then removed and placed at -20°C until all tubes had been collected. The tubes were then spun at >12 000 *g* for 25 minutes at 4°C. Then, to 3ml aliquots of 'Ecoscint A' scintillation fluid in scintillation vial inserts, 200 µl of each supernatant was added, the were vials capped, shaken, and then counted in the tritium window of a scintillation counter (Packard Tri-Carb™ 1600-TR Liquid Scintillation Analyzer).

2.4 PROTEIN CRYSTALLISATION METHODS

The crystallisation of a protein can enable investigations into the three-dimensional structure of that protein at the atomic level to commence. The methods that were used to achieve the crystallisation of the HSV1 UDGase are explained in detail in section **5.1**, and

the optimisation of the crystals grown is presented in section **5.2.** Other techniques of use in protein crystal growth are presented in section **5.3.** The following sections detail the theory behind screening and optimisation experiments, and isomorphous replacement using heavy atoms.

2.4.1 THE RATIONALISATION OF PROTEIN CRYSTALLISATION : SCREENING METHODOLOGIES

A number of approaches can be taken when seeking to crystallise a new protein for the first time. All such approaches rely on the minimal solubility region, or regions, for that protein being identified (McPherson, 1990; Weber, 1991). A range of molecules have been identified as the most common effective crystallising agents or 'precipitants' (PEGs and related molecules, ammonium sulphate, 2-methyl-2,4-pentanediol, sodium chloride) and these can be utilised in any approach designed to locate crystallisation conditions (Gilliland, 1988; Jancarik and Kim, 1991). If, after a week or month, the protein does not appear to be crystalline, then an expanded search can be undertaken which will involve small adjustments to the pH, ionic type and concentration, and possibly temperature. If no crystals are observed in the expanded search then co-factors, substrates and inhibitors of the enzyme can be added. Again, if there are no crystals, some additives may be used which can be any type of small molecule, organic or inorganic. Typical additives are non-ionic detergents (β-octyl gluco[pyrano]side, Triton X-100, N-lauroylsarcosine), di-amines, organic solvents (ethanol, acetone, phenol, dioxane, toluene etc.) (McPherson *et al.*, 1986b; Timasheff and Arakawa, 1988; Derewenda *et al.*, 1989).

If no crystals are observed at this stage, then it may be that the crystals require an extended time to nucleate and begin growth; time lags of several months before crystallogenesis occurs are not uncommon (McPherson, 1990). Often, the chosen type of crystallisation set-up will be wrong for the protein in question. The most commonly used set-ups (McPherson, 1990; Weber, 1991; DuCruix and Giegé, 1992) (Figure **2.4**) are vapour diffusion by hanging or sitting drop in a sealed chamber, microdialysis in a dialysis button, liquid/ liquid diffusion in capillaries, and microbatch under oil. Initial screening is often best carried out by microbatch. In this method the protein and precipitant are immediately mixed to give a set concentration. This will usually show some sign of crystalline material if the mixture is at a metastable zone concentration.

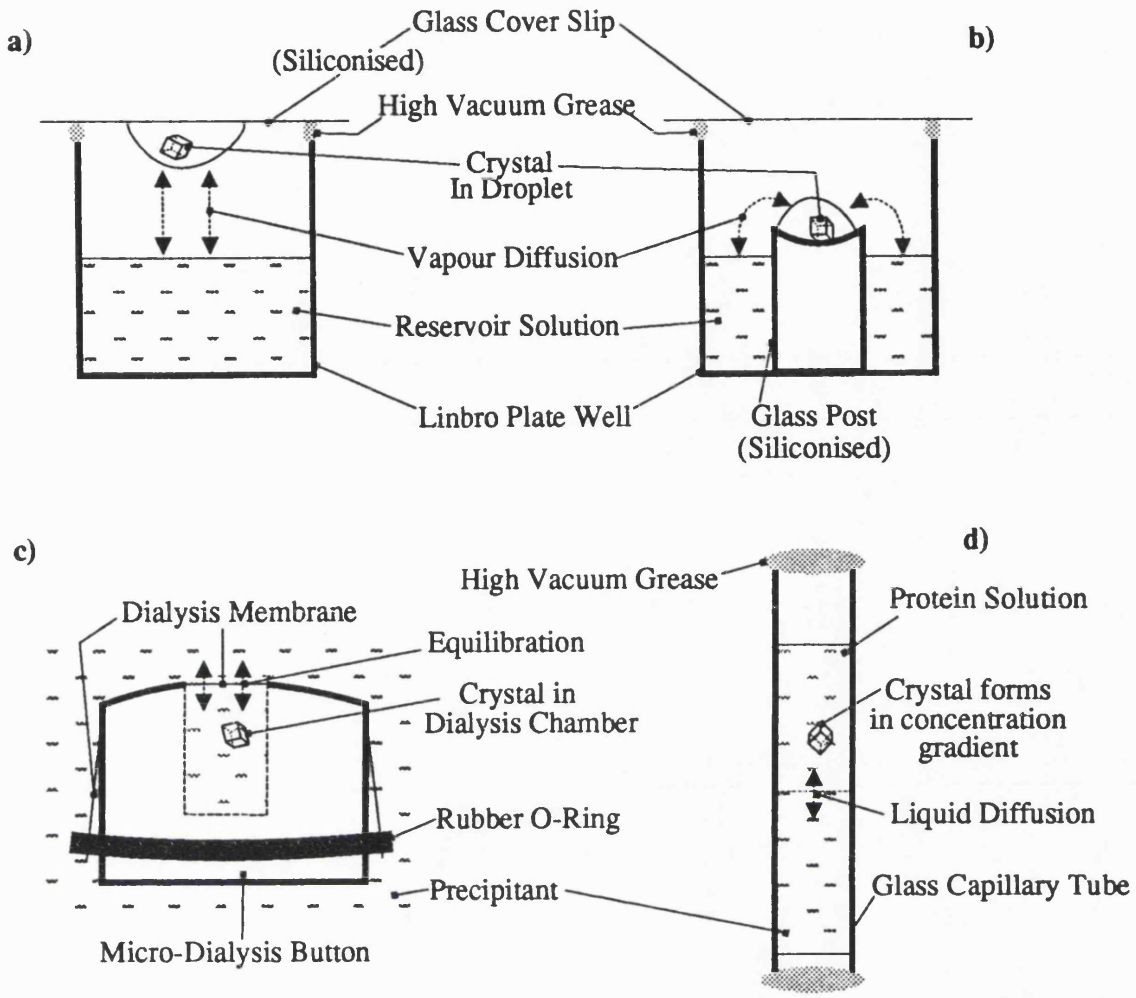


Figure 2.4

Commonly used crystallisation set-ups:

Vapour diffusion in, (a) hanging drop and (b) sitting drop, formats.

c) Microdialysis.

d)Liquid/liquid diffusion.

e)Microbatch.

Crystals of diffraction quality may grow under such conditions, but it is just as likely that a slower approach to the metastable region may be required, by dialysis or the diffusion techniques, to obtain such crystals.

Screening techniques have to effectively and efficiently sample the n-dimensional space that somewhere may contain conditions for the crystallisation of the protein in question. Useful starting points then are: a) at many pH values, using multiple combinations of known crystallising agents, additives and counter-ions , or b) at a pH within 0.5 pH units from the known or predicted isoelectric point of the protein utilising the common crystallising agents.

It may appear that scheme a) would be far too complicated to carry out, but in fact it has been shown to be the most effective method to date for locating crystallisation conditions. The approach taken by scheme a) is known as a factorial search and careful design of a factorial experiment can give a great deal of information. The most successfully implemented version of this scheme in crystallography is the incomplete factorial experiment (Carter [Jr.] *et al.*, 1988; Carter [Jr.], 1992) which uses only a fraction of the total possible number of mixtures which are carefully selected to give the maximum amount of information on multiple interactions while effectively reducing the total amount of protein required. The fact that many proteins have crystal forms far from the pI shows that many regions of minimum solubility exist for a protein molecule depending on the surrounding environment and that it is worthwhile expanding the search to cover a range of pH values. An even more 'incomplete' approach has been used to generate the 'sparse-matrix' selections of precipitant combinations which have been extremely successful in identifying crystallisation conditions for many proteins (Jancarik and Kim, 1991). The approach used in defining this set of solutions, which has spawned many similarly derived sets since, is simply to explore only those areas of n-dimensional space which have been most generally successful in producing crystalline protein in the past. That is, the mixtures that have been documented as having given rise to crystals when added to a protein in the past, are used either in original or slightly modified form (Section **5.1.1**).

Method b) can be carried out as an incomplete factorial (Section **2.4.2**) varying only precipitants and additives or counterions, or may be done by gradient sampling. This is where one precipitant at a time is screened from 0 % to n %, where n is a reasonable maximum concentration, in an attempt to find the region where solubility and precipitation cross over. At this point, additives and counterions may be used to bring on

crystallisation, though it has been observed that crystals may grow solely in the presence of precipitant at the cross-over point which should in theory be the metastable zone. Method b) can be very sparing of protein if a series of 'binary chops' (Shaw Stewart and Khimasia, 1994) are carried out (Figure **2.4.1**). This method can also be a good enough guide to enable the plotting of a rough solubility diagram.

2.4.2 A BRIEF DISCUSSION OF FACTORIAL EXPERIMENT DESIGN

The factorial experiment is one which is useful in identifying the optimum conditions for a process by revealing the important factors within it (Carter [Jr.], 1992). If the effects of different species upon the process and upon each other are known, then it should be possible to identify the optimum conditions for the process. This is readily applicable to crystallography, where a number of different conditions are important in producing a protein crystal.

A contributing factor to a process is known as a main effect. In growing crystals of a protein the main effects are usually all or some of the following: temperature, pH, protein concentration, ionic strength of medium, presence or absence of additive, presence or absence of cosolvent, presence or absence of organic precipitant (Section **5.2**). A combination of main effects is known as an n-factor interaction, where n is the number of factors involved. A full factorial experiment would reveal many n-factor interactions but is not sparing of protein. In experiments with more than four factors, protein can be saved by introducing a new method of factorial search, either a fractional factorial or an incomplete factorial. These two approaches differ in the way that multi-factor space is sampled.

In the full factorial, since all possible combinations of factors are assessed there is no confusion as to which factor or factors is responsible for the most dramatic effects. This type of confusion is known as confounding, and it is used on purpose in the fractional factorial design. The fractional factorial in effect squashes multi-factor space into a smaller grid and as a result, some factors simultaneously increase or decrease in level together. That is, temperature and pH for instance, are always in the same column for many experiments in a particular design. This linking together of factors is called aliasing and the result is that a large volume of n-dimensional space can be sampled in only a few experiments, though a large main effect cannot be immediately assigned to any

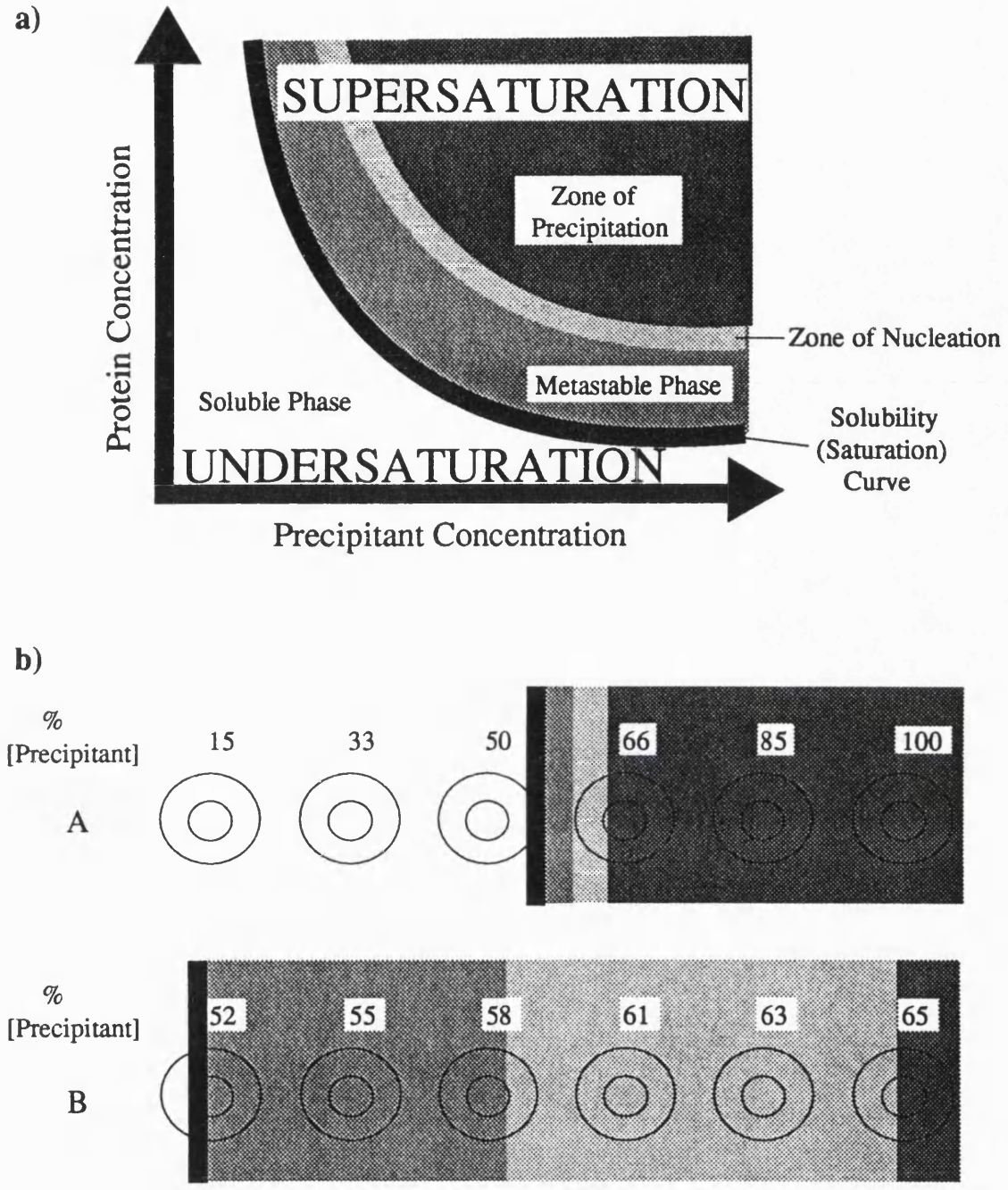


Figure 2.4.1

a) A typical solubility diagram. Crystals may appear in the zone of nucleation, and grow in the metastable zone. This region is often very narrow and difficult to locate.

b) The gradient sampling method. This is a good method for locating the metastable zone. This can be achieved with minimal protein by carrying out a series of 'binary chops'. This involves putting protein in the central well of the concentration range (A) as a start point. Depending on whether the drop precipitates after an arbitrary wait, the next protein addition will be the first or last drop in the range. The next addition will be the midpoint between the clear and precipitated drops in the range. This is continued on finer ranges (B) until the crossover between solution and precipitant is found. This is likely to be the metastable zone.

one factor. Thus a second grid which splits the aliased factors up will be required to assess individual contributions.

The incomplete factorial experiment, though designed with shrinking multi-factor space in mind, achieves the aim in a different way to the fractional factorial experiment. Rather than squashing the larger space onto a smaller grid as in the fractional factorial design, the incomplete factorial shrinks the space by a selection process. This is akin to removing some boxes from the grid and stitching it back together again so that it occupies a smaller area. The experimental combinations that are used are those that convey the maximum amount of information about n-factor interactions in the space remaining without introducing large amounts of confounding. Thus there are holes in the grid which means that although very large amounts of space can be efficiently sampled, some inferences must be made. It is not surprising that the incomplete factorial has been successful in identifying crystallisation conditions for many proteins, as the number of factors which can be included in just one set of experiments is large, without being exhausting on the amounts of protein required. A promising condition can then be sampled more finely on a separate grid.

2.4.3 MATERIALS USED IN CRYSTALLISATION EXPERIMENTS IN THIS PROJECT

A light microscope, with a cross-polarising stage, was used for crystal manipulation. Air displacement pipettes accurate from 0.1 to 10 µl were used to set up droplets for crystallisation. Terazaki plates with 72 wells and liquid paraffin were used in the microbatch procedure (Figure **2.4**). Linbro cell culture plates, siliconised glass cover slips and silicone vacuum grease were required for the hanging drop vapour diffusion method (Figure **2.4**). Liquid/ liquid interface diffusion (Figure **2.4**) required glass capillaries which need not be siliconised. Microscope slides, siliconised glass cover slips and siliconised glass petri dishes were required for manipulation of crystals. Single animal-hairs mounted in glass capillaries with wax were used to manipulate crystals that were stuck to the surfaces of wells in which they have grown. These hairs were also used for microseeding (Section **5.3.3**).

2.4.4 GENERATING HEAVY-ATOM SUBSTITUTED ISOMORPHOUS CRYSTALS FOR SOLUTION OF THE PHASE PROBLEM

The visualisation of the three-dimensional atomic structure of a large macromolecule, such as a protein, by X-ray crystallography requires that the diffraction data be 'focused' into a final image. This is what happens in an optical system such as a microscope; however to date, there is no known method by which X-rays can be focused to reconstitute an image of the magnified object (Figure **2.4.4**). Thus the diffraction pattern must be stored, and the data must somehow be reconstituted into an image by some other means. This data, though containing information about the magnitude of the diffracted radiation, does not contain any relative phase information. Herein lies the 'Phase Problem' of macromolecular crystallography.

The Phase Problem can be overcome if perfect or near-perfect isomorphs of the crystals could be found that contain intense point-scatterers for which phase information can be assigned. These scatterers should be regularly arrayed in the crystal lattice, such that they affect each reflection by constructively or destructively contributing to its intensity. Then, if the real space positions of these scatterers can be determined, approximate phases can be calculated (Section **3.1.2**).

Ideally, the contribution to each reflection by the intense scatterer should be large and there should be only one or two well accounted for sites, which makes subsequent phase calculation more straightforward. The approach taken to solving the phase problem in proteins has been to incorporate heavy atoms into the molecules, which must be large in terms of electrons available for diffraction (Blake, 1968). These heavy atoms may be introduced by reacting the protein in solution, or more commonly by soaking a crystal with a reactive heavy atom containing compound. Such compounds exchange the metal with reactive groups found in some amino acids, either to form a stable ionic pairing, a co-ordination complex, or a covalent bond. The most successfully used of these are the platinum, mercury, gold and uranyl compounds. The compound di-potassium tetrachloroplatinate(IV) is the most successfully used to date (Blundell and Johnson, 1976; Petsko, 1985; McPherson 1982), and it forms co-ordination complexes, predominantly with methionine and histidine residues in a protein. The most widely attempted reaction is that of mercury compounds with free thiol groups located on cysteine residues. This reaction forms a covalent bond between the metal and the thiol, and often there are very

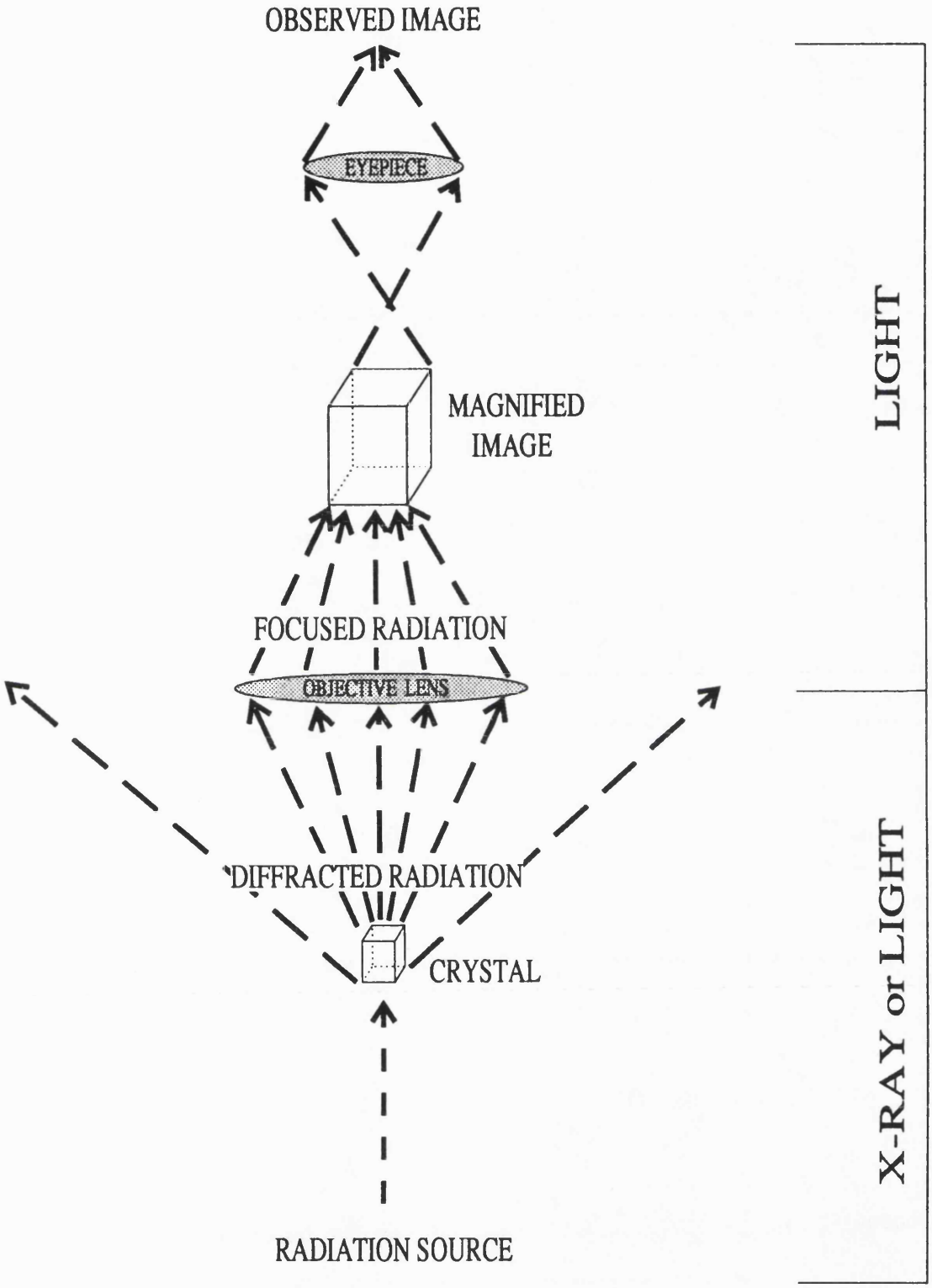


Figure 2.4.4

Magnification of an image. With a light microscope, the diffracted rays can be reconstituted into a magnified image by focusing lenses. With X-rays this is not possible, and the diffracted rays must be captured and the image reconstituted by calculation of phases.

few such amino acids in a protein, or many such groups are inaccessible, making this type of reaction most favourable in terms of the possibilities of solving the phase problem from only a limited number of sites.

Gold compounds such as potassium tetrachloroaurate(III) tend to form co-ordination complexes in much the same way as platinum, and have also been highly successful. Uranyl derivatives are often poor in terms of having multiple sites of reaction, some very weak, making subsequent calculations difficult, but have the advantage that uranium is an immense scatterer and provides excellent intensity changes, making it perfect for obtaining phase information. Many other compounds have been used successfully, particularly silver containing compounds. Silver is particularly effective at reacting with histidine residues. Many compounds react to form loose ionic associations that are, nonetheless, of good enough specificity to be useful. Metals such as cadmium fall into this category. Lanthanide compounds, such as samarium(III) chloride, have been shown to have good reactivity with phosphate groups in protein/ nucleic acid complexes, and many other properties are contributed to various metal compounds. Fuller reviews of observed reactivity and effectiveness of heavy atom compounds may be found elsewhere (Blake, 1968; Blundell and Johnson, 1976; Petsko, 1985; McPherson, 1982).

Other approaches are to introduce the amino acid methionine labelled with a point scatterer in the form of selenium into proteins. This is achieved by growing a methionine auxotrophic variant of the organism which produces the protein in media that contains seleno-L-methionine as the sole source of methionine (Doublié and Carter [Jr.], 1992). Selenium is not a very heavy element in comparison to most of the atoms that can be used successfully for isomorphous replacement, but it does provide a useful anomalous signal (Section **3.1.2**) when used with a tuned radiation source (Hendrickson *et al.*, 1989; Hendrickson *et al.*, 1990; Hädener *et al.*, 1993).

Finally, inhibitors, substrates, and ligands have been labelled with heavy atoms and bound to the relevant protein in order to solve the Phase Problem. Examples of this include DNA binding proteins where iodinated DNA is bound to the protein. The main problem so far has been that although certain potentially reactive residues may be found in a protein, they do not always react with the heavy atom compounds. This may be because they are buried and sterically inaccessible to the compound, or that the local environment of neighbouring amino acids modifies the degree of ionisation on the target amino acid, or that the amino acid is unavailable for reaction because it forms a key

interface in the crystal lattice. It has been observed that some metals react well with a protein in free solution, but not in the crystalline form. If the pre-reacted protein/ metal complex is subsequently crystallised, the crystal may no longer be isomorphous, or the protein may no longer crystallise at all. Thus, the preparation of isomorphous heavy atom derivatives is still very much a trial and error approach (Section **2.4.5**), though some approaches do exist to modify amino acids in the protein, either in free solution or in the crystalline state in an attempt to make them reactive targets for heavy metal compounds (Benesch and Benesch, 1958; Benisek and Richards, 1968; Sigler, 1970; Riley and Perham, 1973; Gallwitz *et al.*, 1974; Mowbray and Petsko, 1983). This approach is not universally applicable, and not consistently successful to date, the main problems again being lack of isomorphism or inhibition of crystallisation.

2.4.5 PREPARING CRYSTALS FOR ISOMORPHOUS REPLACEMENT BY SOAKING

Large crystals (>200 µm on the a/ b faces) were transferred into 1 ml of stabilisation solution in a 3 ml capacity petri dish. The dish was then covered with Nescofilm™ and left overnight. This has the effect of exchanging all the solution in the solvent channels of the crystal for the stabilisation solution, which is free of DTT. The crystals were then stored in Terazaki plate wells under oil, in 10 µl droplets of stabilisation solution. The removal of DTT is important if isomorphous replacement (Section **2.4.4**) is to be attempted with mercury compounds which specifically react with cysteine residues in the protein. Cysteine residues are protected from oxidation by the presence in solution of DTT which keeps them in the reduced state. The removal of DTT from crystals is not as deleterious as the removal of this reductant from protein in free solution, as cysteine residues in crystals have been found to be reactive after extended periods in the absence of a reducing agent when a protein is in the crystalline state.

To attempt isomorphous replacement, a known volume of a solution of a heavy-atom compound dissolved in stabilisation solution was added to the droplet, and the crystal was allowed to soak in this solution in the dark for a set time. The concentration of heavy atom compound and the duration of the soak are a matter of trial and error, but a usual first attempt might be of 1 mM of the compound for 12 hours or so. The problems are that the compounds are usually very reactive and may disrupt the crystal to the extent

that it no longer diffracts X-rays. If crystals are in abundance, a high and low concentration may be attempted concurrently, just in case one of the crystals is destroyed by the soak. For very reactive compounds soaks of a few minutes duration may suffice, as it has been shown that the solvent channels of many crystals are completely infiltrated by a new solution in this time (Wyckoff *et al.*, 1967; Hajdu *et al.*, 1987).

Some compounds react to form covalently bound metal-protein products. These can be 'back-soaked' in stabilisation solution for 30 minutes or more, in order to remove residual unreacted compound and reduce the general background signal. In the case of compounds that react non-covalently this is not recommended. Once the soak had been completed, the crystals were mounted for diffraction as detailed in section **3.2.2.**

3 TECHNICAL DISCUSSION, METHODS, AND MATERIALS PART 2: X-RAY CRYSTALLOGRAPHY OF MACROMOLECULES

This chapter deals with the remainder of the techniques used in this project. These are X-ray diffraction data collection from crystals of a macromolecule, and the subsequent processing of that data. The sections are arranged, as far as possible, in the order in which the techniques would be carried out. To introduce the chapter, a description of the theory behind X-ray diffraction analysis and its application to macromolecules is included. More detailed descriptions of this have been published elsewhere (Blundell and Johnson, 1976; Glusker and Trueblood, 1985; Ladd and Palmer, 1993).

3.1 X-RAY CRYSTALLOGRAPHY IN THE ANALYSIS OF MACROMOLECULAR STRUCTURE

In order to resolve the complex three-dimensional structure of a macromolecule, it is necessary to locate the spatial co-ordinates of its constituent atoms. As the resolution of detail which is possible is limited by the wavelength of the radiation used, X-ray radiation, with a wavelength of the order of interatomic distances, is required. Due to the very small size of even very large macromolecules, a single macromolecule will not suffice to act as the object in an X-ray structure study. A crystal of the macromolecule is ideal however, because it is a regularly repeating three-dimensional lattice comprised of hundreds of millions of copies of the constituent macromolecule. A crystal of a macromolecule acts as a diffraction grating for radiation that has a wavelength close to interatomic distances, and also serves to amplify the diffracting power of the constituent macromolecule by several orders of magnitude. As yet, there is no method of focusing an X-ray beam which has been diffracted, and so the diffraction pattern must be captured and deconvoluted by means of Fourier calculations.

3.1.1 SYMMETRY, THE CRYSTAL LATTICE, RECIPROCAL SPACE, AND X-RAY DIFFRACTION

Crystals of macromolecules are regular repetitions of a basic structural pattern. For the purposes of understanding how these can be used to determine the structure of the constituent macromolecule, the nature of the inherent symmetry of crystals must be understood.

If the crystal lattice is imagined to be a three-dimensional array of regularly spaced points, the view from any one of these points in a particular direction being identical to the view in the same direction from any other point in the same lattice, then the simplest division that can be immediately made out is the unit cell. This is the smallest three-dimensional parallelepiped building block which can be repeated infinitely to describe the whole crystal. In the simplest instance, the corners of the unit cell are the points that constitute the lattice. The unit cell can have boundaries which intersect the constituent macromolecules of the crystal, but it always contains all the atoms and bonds that can be found in a whole molecule. The sides of the unit cell are called a, b, and c, and these are related by the angles α (a/b), β (a/c), and γ (b/c). The crystal has indices which may be related to the unit cell axes a, b, and c. If a plane with indices h, k, and l, makes an intercept with the unit cell, it does so at a/h, b/k, and c/l. These are known as the Miller indices.

Symmetry elements may exist within a unit cell, relating the smallest units from which the molecular structure can be described using the space group symmetry operations. This smallest unit is known as the asymmetric unit. The total contents of the asymmetric unit are a minimum of one molecule, and always a whole number of molecules.

There are nine possible simple unit cell types. Addition of face and body centring possibilities brings this to 14, which define the 14 possible Bravais lattices. Application of various symmetry elements to these results in a total of 230 permitted space groups. Determination of the space group of a crystal is important because once it is known, calculations can be done with respect to a very small region of the crystal, the asymmetric unit of the unit cell. Thus the process of diffraction analysis is greatly simplified.

The symmetry operations are as follows. First of all, point symmetry can be

described, that is n-fold rotation (n= 1, 2, 3, 4, or 6), or mirror reflection, or n-fold rotatory inversion through a point ($\bar{n}$= 1, 2, 3, 4, or 6). This leaves at least one point within the object fixed. These symmetry elements can be combined in just 32 ways in three-dimensions to form the 32 crystallographic point groups. In terms of their simplest symmetry, there are 11 types called Laue groups which have the same centrosymmetric point groups.

When the point symmetry operations are combined with translations, space symmetry operations are produced. These are n-fold screw axes, and glide planes (½ or ¼ along a, b, or c or a face diagonal). The possible combinations of pure rotations, rotary inversions, screw axes, and glide planes results in the 230 possible space groups, or ways in which molecules can pack to ensure that the contents of adjacent unit cells are identical. Of these, only 65 can be assumed by asymmetric molecules such as all proteins, which are L-amino acid based.

Diffraction by a crystal may be likened to diffraction of light by slits, in that the electrons which scatter X-rays in the atoms act as secondary X-ray sources just as slits diffracting light act as secondary light sources. X-ray diffraction by crystals can be simulated by light diffraction through arrays of holes, where the wavelength of light with respect to the spacings of the holes is in proportion to the wavelength of X-rays with respect to the spacing of atoms. The diffraction effects are seen to be almost identical (Glusker and Trueblood, 1985).

With light diffraction through slits, it is found that a single slit determines the size and shape of the intense region (envelope) of the diffraction pattern, a narrow slit giving a wide envelope, and *vice versa*. The diffraction pattern arises due to interference effects of waves that are in and out of phase. With more than one slit, the frequency and number of sampling regions, or points where maxima occur, are determined by spacing of the slits, and a larger number of slits give rise to sharper and sharper patterns where the minima between the maxima become increasingly faint. A crystal is made up of many unit cells, thus sharp diffraction maxima are seen. This may be thought of as a scale up of the sampling of diffraction of a single unit cell.

Light diffraction from a grating of a line of holes appears as lines perpendicular to the direction of the original grating due to interference effects reducing scattered light intensity in other regions to zero. The lines will appear horizontal if the holes are running

top to bottom of the grating, and vertical if the holes are running left to right. If a grating with a two dimensional array of holes is used, the resulting diffraction pattern is of spots with maxima where the lines would intersect, thus the lattice of the diffraction pattern is found to be the reciprocal of the grating holes. Similarly, the lattice points on an X-ray diffraction pattern are found to be the reciprocal of the crystal lattice. The spots that constitute the diffraction pattern are part of a reciprocal lattice. The positions of the spots in the diffraction pattern give information about the unit cell, with spacings of spots between adjacent layers of the lattice being the reciprocal of the unit cell dimensions. The reciprocal cell axes are a*, b*, and c*, and if d(*hkl*) is the spacing between points in the crystal lattice then generally d*(*hkl*)=K/d(*hkl*). The constant, K is usually taken as the wavelength of the X-rays used. The other information from the spots comes from their intensities which is information regarding the arrangement of atoms in the unit cell.

The spots in a diffraction pattern are often referred to as reflections. This is due to the interpretation of diffraction by crystals made by W. L. Bragg. The Bragg equation considers diffraction as though diffracted beams were reflected from planes passing through the crystal lattice. The reflections are analogous to those from light on a plane mirror where the angle of incidence is equal to the angle of reflection. Waves which are scattered from adjacent planes will be in phase by multiples of a wavelength only from certain angles. The equation is written $n\lambda=2d\ \sin\theta$, where λ is the wavelength of the radiation used, n is an integer analogous to the order of diffraction, d is the spacing between lattice planes in the crystal, and θ is the complement of the angle of incidence of the X-ray beam and thus the complement of the angle of scatter (reflection). It follows that $n\lambda$ is the path difference between waves scattered from adjacent lattice planes which have equivalent indices.

3.1.2 PHASE DETERMINATION AND STRUCTURE SOLUTION

The diffraction of X-rays by a crystal is seen as a pattern of spots, referred to as reflections (Section **3.1.1**). These reflections, with indices *h*, *k*, *l* are found at various angular positions in the pattern. These positions are related by 2θ to the dimensions of the crystal lattice. In addition, the reflections have intensity, I, which is related to the types and relative positions of all atoms in the unit cell. In fact, each reflection contains contributions from the entire structure of the crystal. This is because the scattered intensity

at any angle is the summation of waves scattered from all angles.

Any reflection with indices *h*, *k*, *l* on the diffraction pattern has both amplitude, $|\mathbf{F}|$, and a phase angle, α. This information is called the structure factor, $\mathbf{F}$. The intensity of the scattered radiation, I, is proportional to the square of the amplitude, $|\mathbf{F}|^2$. $\mathbf{F}$ and $|\mathbf{F}|$ are related by the following equation:

$$\mathbf{F}\,(hkl) = |\mathbf{F}\,(hkl)|\, e^{i\,\alpha\,(hkl)} = A\,(hkl) + i\,B\,(hkl)$$

where i is an "imaginary number" (Figure **3.1.2 a**). Since $\mathbf{F}$ has a contribution from all atoms in the unit cell (Figure **3.1.2 a**), and I is a simple function of $\mathbf{F}$, it follows that the intensity of each reflection contains information from all these atoms. The amplitude of scattering for any atom is dependent on $\sin\theta/\lambda$, and is called the atomic scattering factor, *f*. An approximation of the scattering by a group of atoms can be made by summation of their independent amplitude and phase contributions (Figure **3.1.2 a**). In the case of a real crystal, the summation of the scattering by atoms in many unit cells produces an average picture of the atoms in the 'average' unit cell of the crystal. General atomic vibration, and slight imperfections in the alignment of adjacent unit cells contribute to a smearing effect of the calculated electron density. The temperature factor, B_{iso} , is related to the mean square amplitude of displacement of the atom from its equilibrium position, and is thus a measure of the degree of smearing. This increases with increasing temperature.

The phase information is lost in the capture of the diffraction pattern. This is the 'Phase Problem' of X-ray crystallography, and so methods by which approximate phases are calculated must be used to obtain a starting point for the calculation of the electron density map. With proteins, the methods of multiple heavy atom isomorphous replacement, and multi-wavelength anomalous dispersion (MAD) (also known as multi-wavelength anomalous scattering) have proven successful in solving the Phase Problem.

The calculation of phases by using isomorphous heavy atom derivatives relies on two main criteria. Firstly, the crystals must be isomorphous in their unit cell dimensions to within ~ 0.5 %, and in their internal atomic arrangements. However, one or more of the atoms in the constituent molecules must be modified with respect to the nature of its scattering properties. In addition, to be useful the derivative thus prepared must modify the average intensity value appreciably. That is the reflection intensities of the derivative crystal must have a measurable percentage difference to those of the native crystal. It follows then, that the useful lower limit of the atomic size of a 'heavy' atom, in terms of

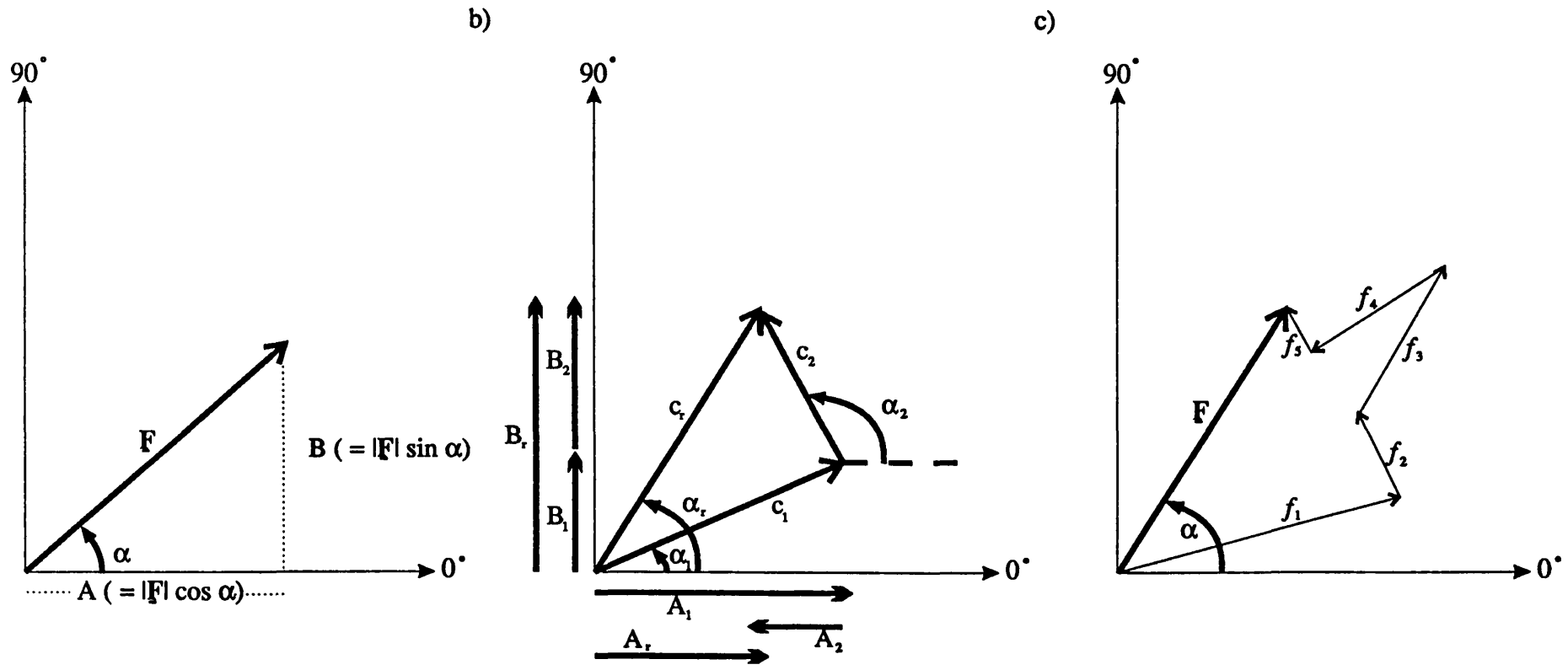


Figure 3.1.2 a

a) The relationship between the phase angle, α, the structure factor, F (whose length, or amplitude, is |F|), and A and B. One of these axes (A) is referred to as the real axis, and the other (B) as the imaginary axis.

b) The result of the addition of two waves, c_1 and c_2, with phases α_1 and α_2, is the same as the result of adding two vectors. Namely the formation of vector c_r, with phase α_r.

c) Vector addition of the contribution of each atom, namely the atomic scattering factors, *f*, in their respective directions (*f* is scalar, but the scattering from the atoms has a phase, taken into consideration here), gives rise to the structure factor, **F**.

usefulness in isomorphous replacement, depends on the molecular mass of the protein to which the atom is bound. In other words, though iodine may give a measurable difference in intensity values with a protein of 20 000 molecular mass, it will not be useful with an 80 000 molecular mass protein. On the other hand, uranium will be useful in both cases.

The method used to calculate approximate phases from heavy atoms must first reveal their real space co-ordinates. The method commonly used to do this is the calculation of an Isomorphous Difference Patterson map. This map is constructed in a three-dimensional grid with the same shape and size as the unit cell. The axes of the grid are u, v, w, and the Patterson function, P (u, v, w), is calculated for each point in the grid. The Patterson function depends only on the values of $|\mathbf{F}|^2$, which are derived from the intensities without the need for phase information. Vectors from the origin of a Patterson map are in close agreement with the magnitude and direction of interatomic vectors in the crystal (Figure **3.1.2 b**).

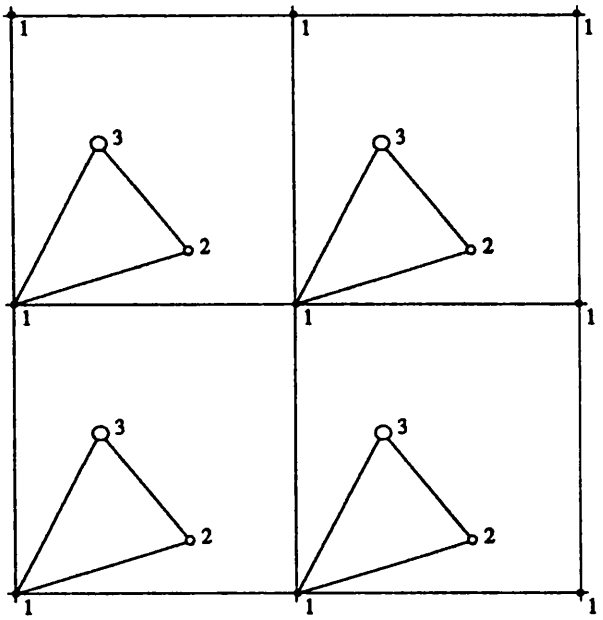
The peaks in the map correspond to the sums of the vectors. Peak height is related to the atomic numbers of the atoms at each end of each vector. Vectors with heavy atoms at one or both ends tend to dominate a Patterson map. If a Patterson map is constructed from the differences in the measured structure factor amplitudes between the native data and the derivative data, then the heavy atom positions can be located. Certain space-group symmetry dependent portions of the Patterson map, known as Harker sections, contain all the cross vectors of space-group symmetry related atoms. Thus, inspection of the Harker sections in an Isomorphous Difference Patterson map will show up the important vectors due to heavy atoms. The phases are derived from the structure factor measurement, where:

$\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$ (P refers to the native protein, H refers to the heavy atom)

This initial estimate of phases is likely to be inexact for a non-centrosymmetric structure such as a protein. To overcome this, the calculated phases from one heavy atom derivative can be compared with another derivative. This derivative can be at the same position, but substituted far more, or less, than the original, or it may be a compound which has reacted at a completely different site in the molecule. If anomalous dispersion (also called anomalous scattering) information is available, then this can also be used to refine the phase estimation.

Anomalous dispersion effects are due to discontinuities in the absorption properties of an atom with respect to wavelength. These discontinuities are referred to as absorption

a)



b)

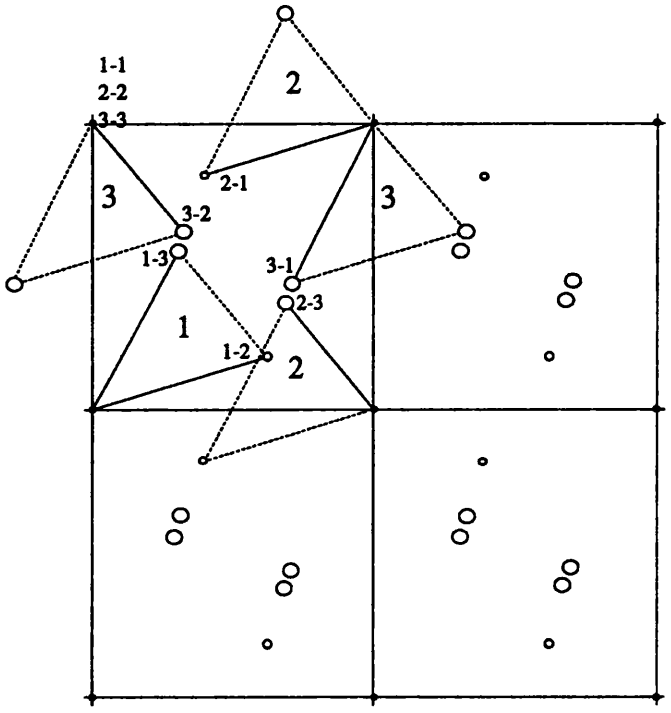


Figure 3.1.2 b

Patterson map construction from known atomic co-ordinates.

a)Four adjacent unit cells in a 2-d plane are displayed for a structure composed of 3 atoms. Each atom may be related to the next by a vector. Because atom 3 is large, all vectors which it makes have large peaks.

b) The Patterson map for the structure in (a). The map is constructed in an area of space the same size as the unit cell. Four maps adjacent to each other are shown to emphasise that all the peaks in the map are due to all possible vectors. Large numbers within the lines indicate which atom must be placed at the origin to obtain the peaks at those positions in the map.The pairs of numbers connected by hyphens refer to the atoms which make the vectors shown by solid lines.

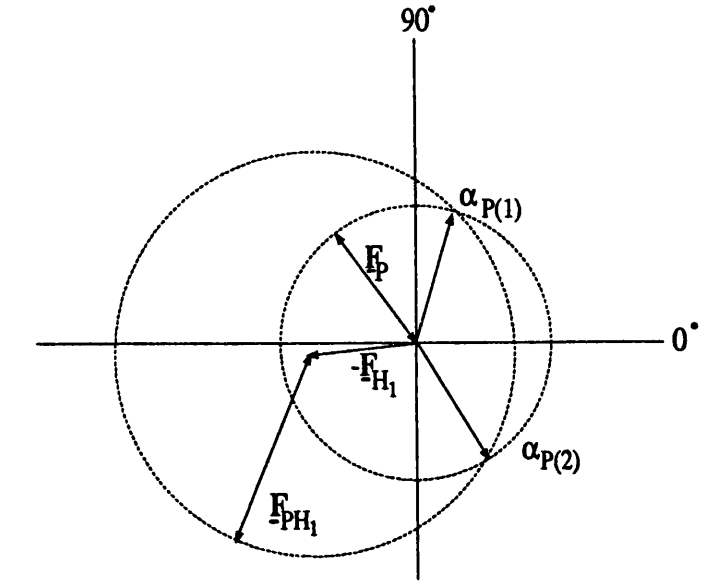
edges, and at wavelengths at, and very slightly below, these edges, the X-rays can excite an electron in the absorbing atom to a higher quantum state, or eject it completely. This excitation effect alters the phase change on scattering (absorption being akin to an effective change in the path length). The resulting phase change from the absorbing atom will be different from that of the non-absorbing atoms in the rest of the structure, and so an anomalous difference in intensity is observed in the Friedel pairs of reflections. The Friedel pairs are related *h*, *k*, *l*, and *-h*, *-k*, *-l*, reflections on opposite sides of the diffraction pattern. Normally, the intensities for these pairs of reflections is equal (Friedel's Law).

To be used more effectively in heavy atom isomorphous replacement, the anomalous scattering information can be combined with a situation where the anomalous scattering effects for the atom are different or non-existent. Anomalous scattering due to the heavy atom is one way of improving the estimate of phases. This property is exploited in the MAD phasing technique, where data is collected at two or three (or more) different wavelengths at which the absorption effects of the heavy atom are different. The use of atoms which would otherwise be below the useful atomic number cut-off is possible if they have significant anomalous scattering at the wavelengths chosen.

According to the law of cosines, there will be two solutions for any phase calculated from the position of a single heavy atom derivative. This is shown graphically in figure **3.1.2 c**. The collection of derivatives must continue until the estimated error in the calculated phases (lack of closure of the circles) is low enough. This is indicated by the value of the figure of merit, m. A value of m above 0.6 usually results in an interpretable electron density map. A figure of merit of 0.8 indicates an average error in calculated phases of ± 40°, a value of 1.0 indicates that there is no error.

The electron density can be calculated from the combination of the estimated phases with the observed amplitude information, since the electron density at any point in a unit cell is related to both of these. If the estimated phases produce a structure that is not too much in error, then refinement can be done to get a better agreement between the calculated and observed amplitude values. The map is interpreted using the primary structure information and the known geometry of the constituent amino acid groups. A discrepancy index, R factor, is used to judge what constitutes a good fit. A random centrosymmetric structure has an R value of 0.67. A good macromolecule structure is generally in the range of 0.1 to 0.2, but caution is advised since very precise, yet

a)



b)

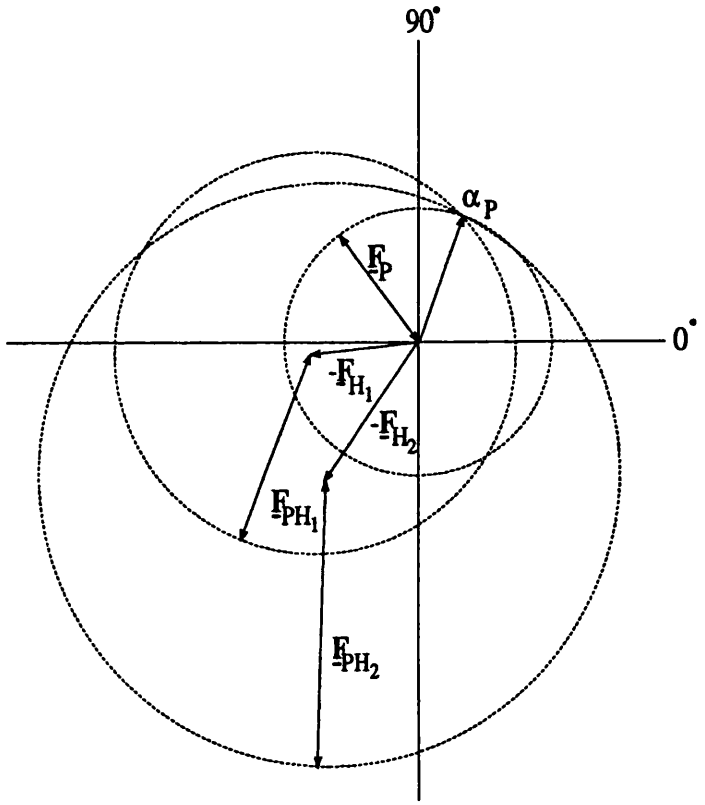


Figure 3.1.2 c

Obtaining phase information from isomorphous derivatives.

a) In the absence of non-crystallographic symmetry, for a non-centrosymmetric structure such as a protein, the phases may have any value from 0° to 360°. As shown in this example, a single heavy atom derivative is likely to be insufficient to give an unambiguous estimate of the true phase. If the origin is the centre of a circle which has radius $|\underline{F}_P|$, then a second circle may be drawn with its origin at radius $|\underline{F}_{H_1}|$ and angle $\alpha_{H_1}+180°$, from the origin of the first circle. The two circles intersect at points which lie close to the true phase, α_P. The intersections may be too far apart to give a reliable estimate of the phases. A second derivative, indicated in (b), shows that the true phase lies close to $\alpha_{P(1)}$, when its circle is superimposed.

c)

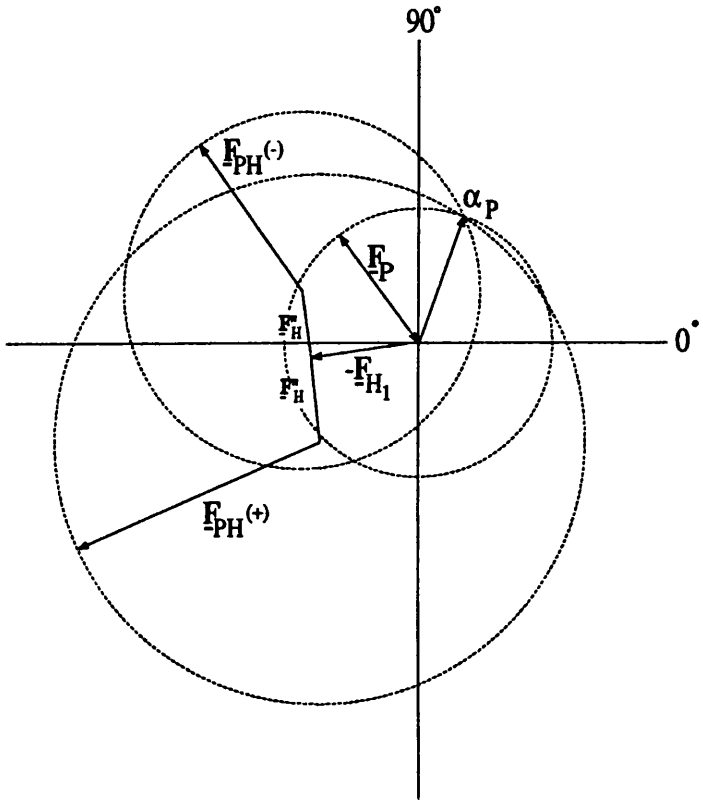


Figure 3.1.2 c (Continued . .)
Obtaining phase information from the anomalous scattering contribution of a heavy atom. c) The contribution to the phasing from anomalous scattering can improve the phase estimate. Anomalous scattering, indicated by $\underline{F}''_H$ breaks Friedel's law, resulting in two different values of $\underline{F}_{PH}$ designated + and -. Thus two phase estimates can be made from a single derivative. Large anomalous scattering contributions have been used to solve the Phase Problem by optimised multi-wavelength data collection.

inaccurate, measurements will yield figures such as these.

3.2 THE COLLECTION OF X-RAY DIFFRACTION DATA

The equipment and techniques used to collect the X-ray diffraction data from crystals of the HSV1 UDGase are presented in this section.

3.2.1 HARDWARE REQUIRED FOR DATA COLLECTION AND DATA PROCESSING

Diffraction (Section **3.2.3**, Section **3.2.4**) was achieved using a laboratory CuKα rotating anode X-ray source, and data collected using a MAR research image plate system. Some data was collected using a Rigaku R-axis/ image-plate system, and data was also collected on a MAR research image plate using tuned synchrotron generated X-rays (Section **3.2.3**). In addition to this, some data collection was also performed using an Enraf-Nonius FAST area detector system (Section **3.2.4**). Data was processed using the CCP4 (and CCP4 associated) suite of programs (CCP4, 1979) on a UNIX® based computer system (Section **3.3**).

3.2.2 MOUNTING CRYSTALS FOR DATA COLLECTION

To carry out X-ray diffraction studies, the crystal and some stabilisation solution were transferred directly into a lead-free glass capillary, which was attached to a 1 ml syringe. The syringe was then used to draw the crystal and some of the solution further into the capillary. At this point, the crystal was allowed to settle onto the wall of the capillary and the bulk of the solution was rapidly drawn away from the crystal. Filter paper wicks were used to remove excess solution from the crystal, and bee's wax was used to seal the open end of the capillary. A layer of bee's wax was then applied to the far end of the capillary, which was then carefully crushed to free it from the syringe. More wax was then applied to form a good seal. A small volume of the solution was kept inside the capillary to maintain a vapour pressure which prevents drying out of the crystal. One end of the capillary was then carefully wrapped in a small strip of plasticine. The plasticine was used to secure the closed capillary to a goniometer. This was connected to the main

spindle assembly of the diffraction apparatus.

3.2.3 SETTING UP A MOUNTED CRYSTAL FOR DIFFRACTION: PART 1 - MAR IMAGE-PLATE AND RIGAKU R-AXIS/ IMAGE-PLATE DETECTORS

An image-plate is a thin layer of photosensitive material (very fine $BaFBr:Eu^{2+}$ crystals) which is oxidised to an excited state ($Eu^{2+} \rightarrow Eu^{3+}$) when incident X-ray photons strike it. The latent images thus formed have a half-life measurable in hours. However, the decay in the excited regions has been shown to be measurable after about 5 minutes After exposure to X-rays, the plate is automatically illuminated with visible laser radiation which causes an electronic transition ($Eu^{3+} \rightarrow Eu^{2+}$) yielding luminescence from the stimulated regions. This luminescence is proportional to the intensity of the X-ray photons that caused the stimulation. The luminescence is measured by a scanning photomultiplier and stored as a digital signal. The image is then erased by a hydrogen erase-lamp illuminating the surface of the image-plate and causing the entire surface to assume the unexposed (Eu^{2+}) state.

The ϕ-oscillation geometry/ image-plate detector system is most useful for straightforward collection of data. That is, the crystal is placed in the X-ray beam path, and a series of angular oscillation images about ϕ are collected by the detector, scanned, and written to a computer file. The images can then be processed (Section **3.3**). For collections of particular desired sets of reflections (spots), the FAST area-detector system is of particular use (Section **3.2.4**).

The mounted crystal/ goniometer assembly was connected to the spindle, which ensured that the crystal in the capillary was close to the X-ray beam path, and could be easily manoeuvred into this path. The centring of the crystal in the beam path was achieved by viewing the crystal through a microscope, or on a TV monitor. The view was perpendicular to the beam path (Figure **3.2.3**). Firstly, the crystal was brought into view by translating the capillary lengthwise. Once in view, a cross-hair scale was used to first align the crystal to a common point at ϕ=0° and ϕ=180°, and then at ϕ=90° and ϕ=270°. A further check was made at this point that the crystal was still in the beam at the first two positions, and any adjustments were then be made by repeating the procedure for all four positions as described. The final check was that the crystal was suitably in the beam

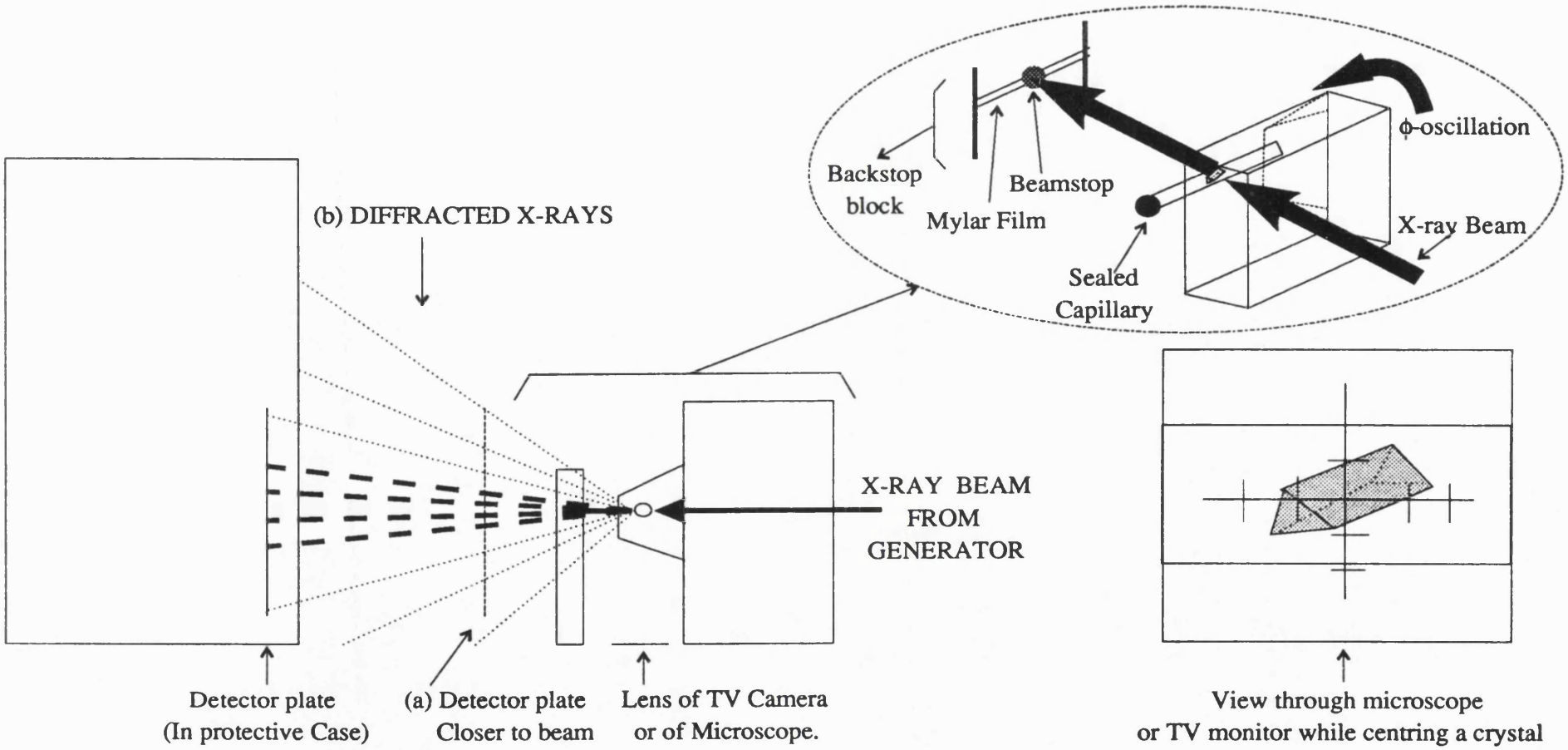


Figure 3.2.3

Typical set-up for X-ray diffraction. The system illustrated is a MAR-Image Plate type system, other systems are principally similar.

a) The detector captures more high-angle reflections at closer detector to beam distances (Section **3.1.2**, Section **3.2.3**).

b) The nature of the atomic scattering factor, *f*, dictates that higher angle reflections are progressively weaker (Section **3.1.2**).

through the whole 360°of ϕ.

Once the crystal was centred, the data collection hardware was set up. This usually took the form of a computer windows display which allowed remote handling of the various parts inside the X-ray enclosure. The enclosure is a lead impregnated perspex unit which contains the X-ray diffraction hardware, no X-rays can enter the enclosure until it is properly closed. The spindle was turned by defining values on the ϕ axis, and an arbitrary ϕ=0° setting was set at the start of the data collection.

Likewise, the distance between the detector and the X-ray beam source could be varied. This is useful for various reasons: firstly the strength of diffracted X-rays that meet the detector is greater, the closer in to the beam the detector is. This is useful if a weakly diffracting crystal is being studied. Secondly, the larger the unit cell of the crystal being studied, the closer the spots recorded on the detector will tend to be (Section **3.1.1**) . This may present a spot overlap problem during processing (Section **3.3**), thus greater distances may be required to solve this problem. Thirdly, the resolution of the reflections recorded toward the edge of the detector is higher, the closer in the detector is to the beam. This is because the higher resolution reflections are diffracted at higher angles, and will only impinge on the detector if it is close enough to be in their path. These reflections are also much weaker, thus the closer in the detector is, the more likely they are to be significantly recorded above the background (Figure **3.2.3**).

Initially, a 1° test oscillation was usually taken to check that the crystal was in the beam, and to give an indication of how strong the diffraction of the crystal was. The response of the image-plate detector to incoming X-rays is linear to a certain point, after which the pixels are registered as overloaded. The software reported back on the maximum pixel intensity in the image. The strength of diffraction was gauged by the relative intensity of pixels in spots to pixels in the background, and a reasonable exposure time per degree was estimated from this. Utility programs which allow profiles of pixel intensity values to be sampled across a spot gave a good indication on how far above the general background the peaks of the spots were.

Other considerations were how long the crystal will be exposed to the X-rays since radiation damage occurs at a detectable rate. Some crystals are very resistant to damage, others very prone to damage. The final consideration was in determining the number of degrees per image taken. The unit cell and symmetry will dictate how crowded the spots

will appear to be, a very sparse distribution of spots means that more than one degree of data may be recorded in one image. The advantage of this is that the overall time for data collection is reduced because a plate-scan will happen less than once per degree.

Finally, the collection strategy was programmed. The number of degrees of data required was programmed, with the number of degrees per oscillation image and an exposure time per degree. More than one pass of the spindle over the same oscillation image was programmed. That is, a 15 minute pass between 1° and 2° was performed as three 5 minute passes over the same angle. This should have resulted in more consistent spot decay on the image plate.

When satisfied that the data would be of suitable quality, the collection was left to proceed. Occasional checks on the images and to see if the hardware was operating normally were carried out. Processing of image-plate data is described in section **3.3**.

3.2.4 SETTING UP A MOUNTED CRYSTAL FOR DIFFRACTION: PART 2 - ENRAF-NONIUS FAST AREA-DETECTORS

A Fast scanning Area, Sensitive Television detector is the technology used in the FAST area-detector system. A silicon intensifier target (SIT) camera tube is connected to an amplifier and then to an analogue to digital converter. X-ray photons are converted to visible light by a photosensitive layer (Gd_2O_2S:Tb) which coats a flat fibre-optic plate connected to a DEP image intensifier that is directly fitted to the SIT camera tube. The multi-stage amplification is required because the number of visible photons produced in the first stage is below the threshold sensitivity of the camera. The data is read 25 times a second from a bank of diodes used to store the electronic image. When these diodes discharge, a video signal is produced. This fast-scanning method accumulates statistically significant numbers of X-ray photons by adding these digitally converted fast frames to an image storage bank.

The FAST area-detector system is particularly useful for collecting desired sets of reflections in order to determine the unit cell and space-group (Section **3.1.1**) of a new crystal type, or to compare a certain set of reflections from a putative isomorphous heavy-atom derivative with the analogous set of reflections from a native crystal. This is possible because the spindle assembly is manoeuvrable in three planes (known as kappa geometry),

κ (kappa), ω (omega), and ϕ (phi), and the data is collected in steps of 0.1° allowing a precise wedge of data to be processed in order to accurately determine the orientation of the crystal. The crystal is then manoeuvred, using any combination of the three motors, into such a position that reflections of the desired region can be collected. The system is also capable of collecting data in a similar manner to the ϕ-oscillation/ image-plate systems (Section **3.2.3**), though the nature of the detector is such that only half as much data is recorded per image compared to image-plate systems.

This is because the detector is small, and it follows that for reasonable distances away from the crystal (Section **3.2.3**), the higher angle reflections would be missed if it were centrally positioned at all times. Thus the detector is mounted on a movable θ (theta) block, and needs to be swung out to various values of θ in order to record the higher angle reflections. However, because it is swung out at an angle with respect to the crystal, the image produced represents only a half of the total diffraction pattern. This means that symmetry related equivalent reflections (Freidel pairs) cannot be compared directly with their partners. In the collection of heavy-atom isomorphous derivatives, this means that one of the reflections constituting a Friedel pair of reflections (Section **3.1.2**), which contain the anomalous scattering information, is not picked up and any anomalous scattering contribution to the phasing is lost.

An important step in the collection of data on the FAST was to take a 'dark image', this is an image of the detector before exposure to X-rays, and was required to be used as a background reference during data processing. The FAST detector is sensitive both to general cosmic radiation, and to noise generated by other electrical appliances, and thus the background may fluctuate according to prevailing conditions. Dark images were also taken at further points in the data collection so as to keep a reliable background estimation for all the collected data, which enabled the processing to be more accurate.

3.3 A SHORT DESCRIPTION OF X-RAY DIFFRACTION-DATA PROCESSING USING THE CCP4 PROGRAM SUITE AND ASSOCIATED PROGRAMS

The following is a list of the programs used in this project to process diffraction data from crystals of the HSV1 UDGase. These programs are distributed as part of the CCP4 Program Suite (CCP4, 1979). A brief description of the tasks performed by each program is given in this section.

1) The MOSFLM suite: A modular package that determines spot co-ordinates and intensities, performs autoindexing to solve the crystal orientation, refines the unit cell, scales and integrates the data (determines absolute intensities) from the images collected, and applies appropriate corrections. The suite (current version 5.20) consists of the following programs:

a) The IMSTILLS program: Creates a list of the co-ordinates and intensities of significant spots which it has located on an image. This list is written to a file which is read by the program 'REFIX'.

The significance of spots was rated by a user-variable threshold setting. This setting refers to the number of standard deviations above the general background that a pixel must be in order that it, and adjoining pixels, is called a spot. In order for IMSTILLS to determine a reliable reference background, the useful area of the image-plate to be scanned was input. This means giving the co-ordinates for the screen size, the area of the screen taken up by the image, and the area of the image to ignore as being occluded by the beamstop and its shadow. Most importantly of all, the accurate position of the beam centre was input. This was especially important, because the next program 'REFIX' relies on the accuracy of this value. The midpoint, in degrees, of each oscillation image was also user defined. The program then divides the image up into sectors, known as 'bins', and the strongest spots in each bin have their intensities and co-ordinates written to the output file.

In order to get a reliable value for the beam centre, a powder ring image is required. Such an image was produced well by the bee's wax used to seal the capillary containing the crystal. This image consisted of two rings which are of fixed radius for any fixed detector to beam (at the capillary) distance, thus the accurate distance of the detector

from the capillary was revealed by their radii. The accurate centre of the rings was found by using an interactive graphics display, where the cursor was positioned and entered at several points on the image of the ring. The program performed a least squares fit of the points which determined the centre of the ring.

b) The REFIX program: This program performs autoindexing to solve the orientation of the crystal, and outputs this in the form of a matrix.

This is achieved by the program reading the output from IMSTILLS to obtain the spot co-ordinates. The wavelength of the X-rays used, the detector to beam distance, and the unit cell dimensions were all input. These values need only be accurate to a few percent, as the program refines them to an extent. In addition, the spacegroup may be input, but if the unit cell dimensions and spacegroup are unknown the program will attempt to solve these. However this is not an altogether reliable means of doing so. The FAST area detector is the instrument of choice for this particular task. In addition, constants and expected errors regarding the hardware were input, as was a threshold setting regarding which spots could be used by the program (Kabsch, 1988; Kabsch, 1993).

c) The MOSFLM program: This program processes the diffraction images to the raw data contained in them, namely the reflection indices, intensities, and standard deviations. These data are also suitably corrected to account for background contributions, hardware imperfections, and geometric effects caused by the way the data is collected.

This program takes an orientation matrix, which is the output from REFIX, creates a list of reflections and post-refines it, until the shifts between images are very small, using partial reflections (reflections which are partly on one image and partly on the next) before integrating the reflections according to a spot 'mask', and applying detector-independent geometric and absorption corrections. The detector parameters are also refined. The mask is refined slightly from a manually defined box. This box is measured in pixels, and must include space not only for the spot, whose shape is defined within the box, but also for background pixels. The program then attempts to fit the spots to the mask. Spots whose boxes overlap are deemed too close, and are rejected as overlapping. Spots which do not fit the box closely enough are also rejected. The spots which fit the boxes well are then mathematically corrected to subtract the local background (in the box)

from them. This is done for each spot in turn, taking account of local background gradients. That is, a spot may have more background to subtract from one side than the other. Comparisons of spot intensities and profiles from several areas of the detector are made, integration is performed for each region, and the results are written to a file.

The geometric corrections performed are the Lorentz and polarisation corrections, and absorption corrections may also be performed. The Lorentz factor corrects for the relative times that reflections spend in the plane of the detector, due to the geometry of the diffraction hardware. Reflections close to the axis of rotation will have more time in the plane than those at the edge, and this must be corrected. An X-ray beam (other than a synchrotron beam) is not polarised until it passes through a crystal, whereupon it loses a small amount of energy in polarisation. This can also be easily corrected by the program. Absorption effects can also be corrected, though this is more difficult to carry out without direct measurement of the degree of absorption by the crystal from which the data was collected. Estimated absorption correction is possible. Once the corrections have been applied, the data is scaled to the asymmetric unit as *hkl*s by finding the symmetry equivalents of measured indices, and written as an MTZ file that is read by the program SORTMTZ. R-factors for symmetry related reflections are calculated and written to a summary file which contains other statistics, to be viewed by the user.

2) The SORTMTZ program: Re-arranges the data so that equivalent sets of *hkl*s are adjacent and outputs a file that is read by the next program.

3) The ROTAVATA program: This program accepts data that has processed to the *hkl* level, from image-plate or FAST area detectors. The program calculates scale factors for overlapping batches of data, and can also calculate temperature factors (which are resolution dependent).

4) The AGROVATA program: Reads in the intensities of *hkl*s, applies the scale factors calculated by ROTAVATA, adds the partially recorded reflections, averages good agreements between symmetry equivalent reflections, and rejects bad ones. Adjustment of the bias of fully, and partially recorded reflections is carried out, such that the mean is close to zero, and the standard deviation is close to unity. The data is reduced to the asymmetric unit at this stage. The output is a file containing the mean intensities, and a

statistical analysis for the user.

5) The CAD program: combines assorted data, that is the data from two or more data sets, such that they are aligned by *hkl*. This is the output file.

6 i) The LOCAL program: Adjusts the scaling of the native and derivative data sets by first adjusting the overall scaling by temperature factor, then by locally scaling this adjusted data. The local scaling is of small areas of reciprocal space. This is thought to be more accurate than scaling across the whole of reciprocal space, since there may be a genuine inconsistency between different areas of reciprocal space.

6 ii) The FHSCAL program: This program can be used instead of LOCAL, to adjust the scaling between the native and derivative data sets in shells of equal volume segments of reciprocal space. This achieved using the formula of Kraut (1962).

7) The SCALEIT program: Can be used to globally scale reciprocal space, or used to analyse the isomorphous differences between a derivative and native data set.

8) FFT: This calculates the difference Patterson map between the native and derivative data sets by fast fourier transform. The symmetry used is the Laue group symmetry. The output can be graphically generated to allow manual inspection to find the and solve the peaks, or the next programs may be used to do this automatically.

9) PEAKMAX: This locates the significant peaks in the Patterson difference map.

10) VECSUM: This is a vector verification program, and it looks for the symmetry related positions of anything identified as a peak in the previous program. Peaks, which are the results of vectors between atoms, will solve for symmetry related positions, whereas spurious noise will not. Once this has been done, the vector map is put back into PEAKMAX, and the real space co-ordinates of the real peaks are calculated and displayed in the output.

11) MLPHARE: This program refines all the heavy atom positions from derivative data sets using a maximum likelihood method, and calculates the phases.

4 CLONING, EXPRESSION, AND PURIFICATION OF HSV1, AND *E.COLI*, URACIL-DNA GLYCOSYLASES

To conduct crystallographic trials, optimise crystal size (Chapter **5**), and collect data from suitably sized crystals with a protein that has never before been crystallised, milligram quantities of the protein are desirable. In addition, the protein should preferably be as pure as possible (Giegé *et al.*, 1986; Wood, 1990; Lorber and Giegé, 1992). In order to produce the protein in large quantities on a routine and easily reproducible basis, expression under a strong promoter in a recombinant expression system is the method of choice. This also has the advantage that the levels of other proteins as a percentage of total protein is lower, thus making a high degree of purification more easily attainable.

Considering that the HSV1 UDGase is normally produced in eukaryotic cells, it was anticipated that some problems might be encountered if an attempt was made to express the protein in a prokaryotic system. This might be due, for instance, to differential codon usage which might cause poor translation (Adhya and Gottesman, 1978; Robinson *et al.*, 1984). Also, problems due to cytotoxicity might arise. This was not expected to be the case considering the high degree of conservation of the UDGases, although it is possible that highly elevated levels of any protein that is normally present in much smaller amounts may produce some unexpectedly adverse effect. However, considering the large range of expression vectors for use in *E.coli* and the ease of culture and lysis of these cells, it was decided that this prokaryote would be the first choice host for large-scale expression of the protein.

4.1 CLONING STRATEGIES FOR HSV1 URACIL-DNA GLYCOSYLASE

A putative ORF encoding a UDGase is located in the UL2 region of the HSV1 genome (Caradonna *et al.*, 1987). There are three potential candidate ATG start codons, located at positions 9886, 10156, and 10261 in HSV1. The ORF runs to position 10890 (Figure **1.3.2**). The resulting polypeptides would contain 334, 244, and 219 amino acids respectively, all of which would be in the known molecular mass range of UDGases

(Section **1.2.2**).

The plasmid pGX23 (Mullaney *et al.*, 1989), is pAT153 (Twigg and Sheratt, 1980) containing a *Bam*H I fragment which corresponds to position 2907 to 11825 in HSV1. The first strategy was to use PCR (Section **2.1.1**) to obtain the gene encoding UDGase, making the assumption that the start codon was at position 9886 in HSV1. Primers for the reaction had already been designed and synthesised at the start of this project, and were designed to introduce *Bam*H I restriction enzyme sites at the 5' and 3' ends of the gene (Figure **2.1.1**).

4.1.1 AMPLIFICATION OF THE UL2 ORF BY PCR

Using 0.2 µg of each primer with incorporated *Bam*H I sites, and 20 ng of intact pGX23 plasmid as template DNA, PCR was carried out using 5 units of *Taq* DNA polymerase, with a final concentration of 0.8 mM dNTPs, in a total volume of 50 µl. As a control reaction the gene *tag*1, encoding the *E.coli* Tag protein, was amplified using known ratios of primers designed for this gene, and *E.coli* genomic DNA as template. In addition the *E.coli* UDGase gene, *ung* was amplified by PCR as above, using 0.2 µg of *E.coli* genomic DNA as template, with 0.2 µg of each primer designed to introduce an *Nco* I site at the 5' end of the gene and a *Hin*d III site at the 3' end (Figure **2.1.1**).

The program used was one which had been previously optimised for the amplification of *tag*1, after much refinement of the recommended parameters outlined in section **2.1.1**. This program was:

30 cycles of

2:00 minutes at 93°C denaturing temperature

2:00 minutes at 55°C annealing temperature

5:00 minutes at 73°C extension temperature

Under these conditions the UL2 ORF appeared to amplify weakly judging by the size of the expected band, observed on an ethidium bromide stained 1 % agarose gel (Section **2.1.2**). However, there were other bands visible in the same lane indicating the need for a higher annealing temperature, or perhaps the addition of less template in the reaction, as these bands could be due to an excess of pGX23 template DNA. The *tag*1 control amplified as expected, verifying the activity of the polymerase batch used. A

negative control, which included all reagents used in the positive control except a DNA template, showed some non-specific amplification. The *E.coli ung* gene did not amplify under the conditions used.

The UL2 PCR was adjusted slightly by lowering the quantity of template DNA to 4 ng, and adding glycerol to 10 % v/v which has proved useful in increasing yields of PCR products (Lu and Nègre, 1993). The original program parameters were retained, and an ~1 kb product (Figure **4.1.1**), the expected size of the UL2 ORF, was strongly amplified with no non-specific products visible. Control lanes showed nothing in the negative control and amplification in the positive control.

The *E.coli ung* amplification was successful when the program parameters were changed thus:

30 cycles of

2:00 minutes at 93°C denaturing temperature

2:00 minutes at 59°C annealing temperature

5:00 minutes at 73°C extension temperature

In addition the level of total *E.coli* genomic DNA template was decreased to half the original quantity. There were no visible bands apart from a strongly amplified ~700 bp product, the expected size of the *E.coli* UDGase gene, *ung* (Figure **4.1.1**).

4.1.2 ATTEMPTED CLONING OF THE UL2 ORF (9886 -10890) AMPLIFIED BY PCR

The expression vector chosen was pET3a, which is further discussed in section **4.2.1**. The aim was to clone the amplified ORF corresponding to 9886 - 10890 of UL2 into this plasmid. Following transformation of the *E.coli* strain, JM107, transformants were to be selected by use of the ampicillin antibiotic resistance conferred by the plasmid. Recombinant plasmids would then be transformed into the *E.coli* strain, BL21(DE3) for expression of recombinant protein.

A fraction of the PCR mixture was run on a 1 % agarose gel stained with ethidium bromide, and viewed on an ultra-violet lightbox of long wavelength to minimise DNA damage. A scalpel was used to remove the ~1 000 bp band from the gel, and the DNA was recovered from the agarose using the Geneclean® procedure (Section **2.1.9**).

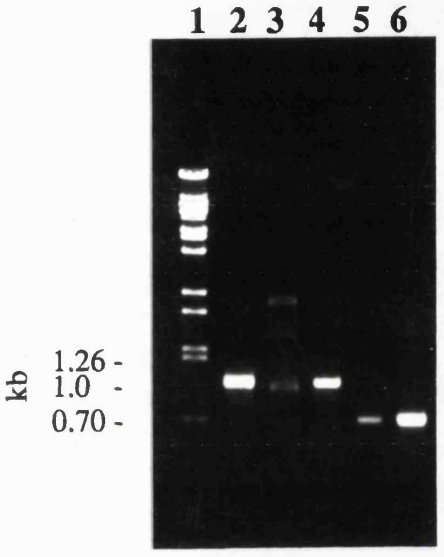


Figure 4.1.1

Amplification by PCR of the HSV1 UL2 (9886-10890 ORF), and *E.coli ung*, genes. Ethidium bromide stained 1% agarose gel showing bands migrating at the expected sizes relative to a λ-DNA/ *Bst*E II digest (Lane **1**). Lanes **2**, **3**, and **4** show the expected 1009 bp product of the HSV1 UL2 (9886-10890 ORF) PCR. Lanes **5**, and **6** show the expected 704 bp product of the *E.coli ung* PCR.

Differences between otherwise identically prepared samples (2, 3, 4 for HSV1 UL2, and 5, 6 for *E.coli ung*) indicate that minor inconsistencies in the pipetting of constituents of the reaction can have a dramatic effect on the specificity (cf. 3, with 2 and 4), or final yield (cf. 2 with 4, and 5 with 6) of PCR.

Restriction digestion (Section **2.1.3**) of this DNA fragment with *Bam*H I was then carried out at 37°C.

The plasmid was also digested with *Bam*H I at 37°C, and in addition 10 units of calf intestinal phosphatase (CIP) were present in the reaction mixture to prevent plasmid self-recircularisation during ligation. The plasmid was Geneclean® treated to remove enzymes, and 0.1 µg of plasmid was incubated with insert DNA at various molar ratios overnight at 15°C in the presence of 1 unit of T4 DNA ligase. *E.coli* bacteria were then transformed with a portion of the ligation mix and cultured on ampicillin/ nutrient agar plates for 12 hours at 37°C. Control plates were: a positive control of bacteria transformed with a closed circular plasmid to prove cells were suitably competent; bacteria transformed with digested plasmid alone, both CIP treated and non-CIP treated, that had been incubated with T4 DNA ligase to test the extent of recircularisation without the presence of the UL2 insert; a negative control was bacteria that had undergone the transformation protocol without the addition of DNA to show that the competent cells were not resistant to antibiotic without the presence of a resistance conferring plasmid.

Plasmids containing the inserted UL2 fragment were identified by restriction digestion with *Bam*H I following isolation of plasmid DNA from ampicillin resistant colonies using the small-scale plasmid DNA preparation method outlined in section **2.1.5.** Positive clones were expected to yield bands corresponding to ~4.6 kb for the plasmid and ~1 000 bp for the inserted fragment when run out on an ethidium bromide stained 1 % agarose gel.

The orientation of the inserted fragment was to be determined by a further restriction digest of any plasmid found to contain an insert. A unique *Pst* I site exists in the pET3a plasmid approximately 1.2 kb upstream of the unique *Bam*H I cloning site and a unique *Pst* I site also exists at 9897 of HSV1. A *Pst* I restriction digest of the plasmid would reveal the linearised 4.6 kb band of the plasmid in the absence of an insert. In the case of a plasmid with an inserted UL2 fragment, this digest would reveal double banding. In the correct counterclockwise orientation, the *Pst* I site in UL2 lies ~2.2 kb from the plasmid *Pst* I site. However in the incorrect clockwise orientation it lies only ~1.2 kb from the plasmid *Pst* I site. So in the correct case, expected band sizes are ~2.2 kb and ~3.4 kb. In the incorrect case these bands will be ~1.2 kb and ~4.4 kb (Figure **4.1.2 a**).

After several attempts using this strategy, the UL2 fragment was still not found to

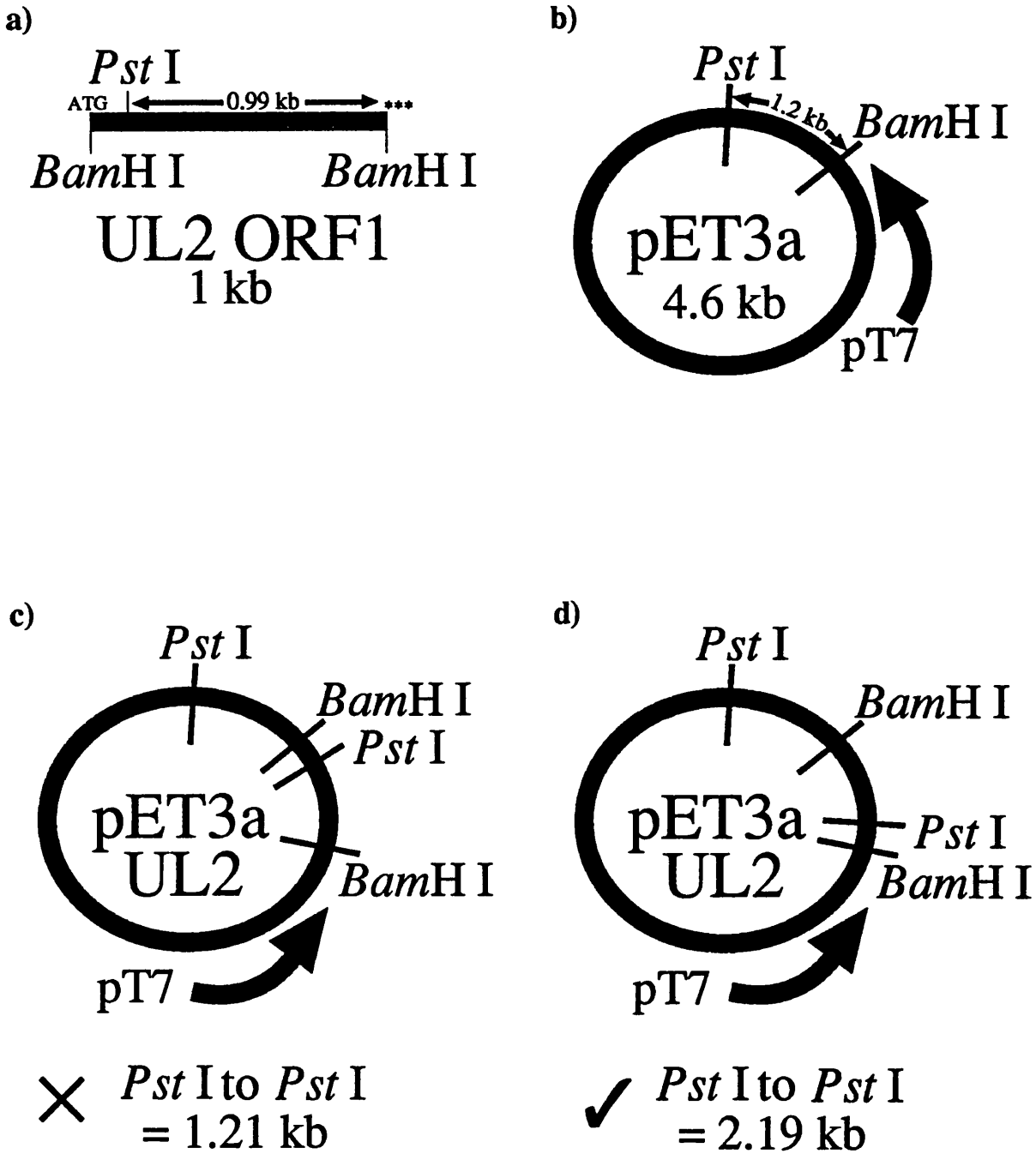


Figure 4.1.2a

Strategy to determine the orientation of an inserted DNA which can assume either of two directions in a plasmid. The example is of the strategy worked out for the cloning of the first candidate open reading frame of UL2 into the plasmid pET3a.

be present in a pET3a plasmid. Several misleading false positive colonies were seen, and it was clear from the control plates that either some plasmids were still recircularising despite being treated with CIP or that the restriction endonuclease was not cleaving all plasmid molecules. A further precaution of extracting the *Bam*H I restriction digested/ calf intestinal phosphatase treated plasmids from 1 % agarose gels, as a linear ~4.6 kb band, was taken prior to incubation with the UL2 fragment and T4 DNA ligase to ensure that only restriction digested plasmids were present. Some recircularisation was still seen to take place, as judged from the control plates, and so it appears that the CIP was not successfully removing 5'-phosphate groups from all plasmid molecules.

The large numbers of ampicillin resistant colonies obtained after transformation with the ligated plasmids, made thorough screening of such colonies by small-scale plasmid preparations and restriction enzyme analysis impractical. To overcome this, a rapid relative sizing protocol for closed circular plasmids from crude cell lysates was used (Section **2.1.12**). Each lane of the gel corresponds to a different colony on the nutrient agar plate, and if any lane differs in migration from a control lane of closed circular plasmid without an insert, the corresponding colony can be selected for small-scale plasmid preparation and restriction enzyme analysis.

In practice, many colonies appeared to differ in migration on a 1 % agarose gel and intriguingly none of the colonies appeared to harbour the UL2 fragment as judged by restriction enzyme analysis (Figure **4.1.2 b**). Most commonly, spuriously migrating pET3a plasmids, lacking any inserted DNA, were found to be the cause. Some more minor observations of non-identifiable inserts, as well as contaminating ampicillin resistant bacteria, were recovered on different occasions. However, the rapid sizing method is not entirely unsuccessful in aiding the identification of clones (Section **4.1.4**, section **4.1.5**).

The next methods attempted utilised a plasmid vector PCR1000 (Invitrogen Corporation), as well as pET3a, in conjunction with DNA hybridisation methods to identify colonies harbouring recombinant plasmids. The PCR1000 vector (~3.0 kb) makes use of the terminal transferase activity of *Taq* DNA polymerase (Clark, 1988), which adds adenylic acid moieties to the 3' ends of DNA molecules which have been fully extended in PCR. The vector contains unpaired 5' thymine bases and so in theory, ligation of PCR fragments will occur in the presence of this vector, thus negating the requirement of restriction enzymes. The PCR1000 plasmid contains DNA which encodes the α-peptide complement of the *E.coli* IPTG inducible β-galactosidase gene *Lac*Z. This enables X-gal

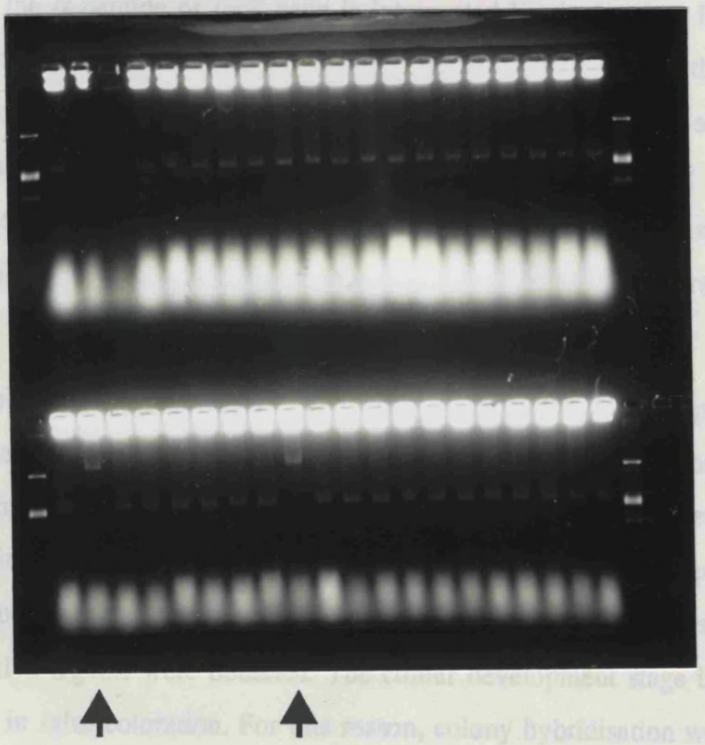


Figure 4.1.2 b

The typical appearance of a colony screening by rapid lysis run out on a 1% agarose gel and stained with ethidium bromide (Section **2.1.12**). Two lanes appear to have migrated differently from the others in the lower row. This result was found to be a false positive, but the appearance is very similar for plasmids which have been successfully ligated to the required DNA fragment.

degradation to take place in *E.coli* strains deficient in this portion of *LacZ*, resulting in blue colonies on ampicillin/ nutrient agar plates impregnated with IPTG and X-gal, when plasmids lack inserted DNA sequences. In the case of a recombinant plasmid, the region encoding the α-peptide of *lacZ* gene is interrupted by the inserted DNA, and no X-gal degradation can take place, resulting in normal white colonies on these plates; thus all white colonies were regarded as possible recombinants. From small-scale preparations of the plasmids, it was clear that they harboured no inserts. Owing to the numbers of transformants obtained however (hundreds on average), it was not feasible even to prepare the colonies for rapid sizing from crude lysates and so a new screening method was required.

Colony hybridisation, using the UL2 PCR product as a probe, with a non-radioactive detection system was used (Section **2.1.13**). Often, positive signals were obtained from blue colonies as well as white ones, even under stringent wash conditions, and even in the case of such positive signals, the restriction digested plasmids showed no evidence of inserted DNA sequences. Even when the procedure was used with pET3a, false positive signals were obtained. The colour development stage is time critical, and can result in false coloration. For this reason, colony hybridisation was carried out with ^{32}P end-labelled UL2 mid-sequence primer (15 bp) (Section **2.1.14**), and the results viewed on autoradiographs; again they proved to be littered with false positive signals.

An end-labelling experiment was then carried out with the UL2 PCR product. The ends of the PCR product were ^{32}P-labelled using T4 polynucleotide kinase and then the fragment was restriction digested with a large excess of *Bam*H I for 20 hours at 37°C. The digested and undigested samples of the UL2 PCR product were run side by side on a 12 % acrylamide/ 6 M urea gel and the autoradiograph examined. It was clear that although much of the radioactivity had migrated as very low molecular weight bands, signifying restriction digestion from the main DNA fragment, some radioactivity remained in the main high molecular weight DNA region. This could either be due to incomplete digestion of the DNA or it could signify that one of the two ends of the DNA is not being restriction digested (Figure **4.1.3 c**).

It was concluded that for some reason, the UL2 PCR product was resistant to cloning in *E.coli*, and that this might be due to either a) a cytotoxic effect, thought to be unlikely owing to the lack of a T7 RNA polymerase in *E.coli* JM107, or b) a bad sequence in the restriction enzyme site built into one or both primers, which would be

tested by the next set of experiments, or c) secondary structure features in the DNA sequence of UL2 which might prevent efficient plasmid replication, which would also be tested in the next set of experiments.

4.1.3 AN ALTERNATIVE APPROACH TO THE CLONING OF THE UL2 ORF

The UL2 ORF (9886-10890) can be directly removed from pGX23 by restriction digestion if the first twelve bp are lost, as a *Pst* I site exists at position 9897 in HSV1. The second potential start point for the ORF is at position 10156 in HSV1 and can be cut at the ATG with *Nco* I leaving the entire ORF intact. The third potential start codon has no convenient restriction site close by, and this length of ORF would have to be isolated by PCR. A number of restriction enzyme sites lie downstream of the ORF.

The plasmid pUC19 (Norrander *et al.*, 1983) contains a polylinker sequence (also known as a multiple cloning site, this sequence is specially built to allow recognition by several different restriction endonucleases along its length) into which the gene of interest may be cloned as a fusion with a methionine codon (and a number of other codons depending on the restriction enzyme site used) which, when transcribed acts as the start point for translation. The methionine codon lies upstream of the polylinker and downstream of a *lac* promoter which directs transcription. It was determined that if the UL2 ORF was cut at the 9897 *Pst* I site and ligated to the *Pst* I site in pUC19, then an in-frame fusion would result. An appropriate downstream restriction enzyme site for the pUC19 polylinker was deemed to be a blunt cutting restriction enzyme site hybrid between *Sma* I in pUC19 and *Eco*R V at position 11084 in HSV1. This arrangement would result in a partial fusion sequence for the UL2 ORF, with the first four codons of the UL2 ORF replaced by 10 codons from pUC19. The fourth codon of the UL2 ORF and the tenth codon of pUC19 are both for alanine (Figure **4.1.3 a**).

The pUC19 plasmid also enables blue/ white colony colour selection due to disruption of *lac*Z α-complementation by inserted DNA. The UL2 fragment for insertion was restriction digested with *Pst* I/ *Eco*R V and run on a 1 % agarose gel, migrating as an ~1.2 kb band , and pUC19 was restriction digested with *Pst* I/ *Sma* I and migrated as an ~2.7 kb band when run on a 1 % agarose gel. Both the UL2 fragment and pUC19 were extracted from the gel by the Geneclean® procedure. Overnight incubation with T4 DNA ligase was carried out at 15°C, and the *E.coli* strain JM109 was transformed with the

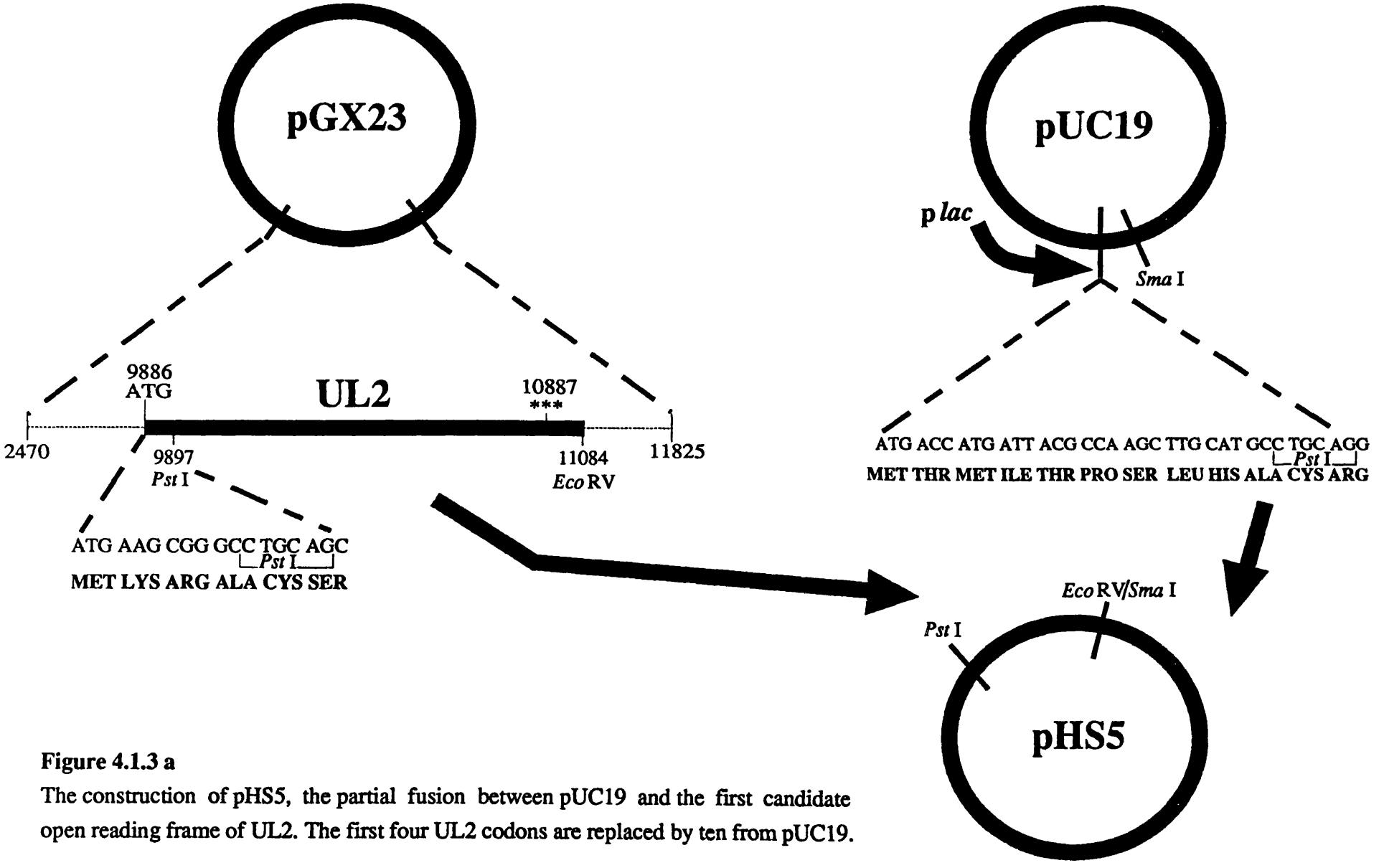


Figure 4.1.3 a

The construction of pHS5, the partial fusion between pUC19 and the first candidate open reading frame of UL2. The first four UL2 codons are replaced by ten from pUC19.

DNA and plated out on ampicillin/ nutrient agar impregnated with IPTG and X-gal. Small-scale plasmid preparations from white colonies were restriction digested with *Eco*R I/ *Hin*d III and run on a 1 % agarose gel. Colonies harbouring recombinant plasmids were identified by the presence of two bands of ~1.2 kb and ~2.7 kb. The first such result was colony number 5 and was called pHS5. Further restriction enzyme analysis on a *Pst* I/ *Sac* I fragment from pHS5 to check for the presence of the internal UL2 sites, *Nco* I, and *Xba* I was also carried out to verify the result (Figure **4.1.3 b**). This result showed that the UL2 fragment did not contain inhibitory secondary structure, neither did it appear to be cytotoxic, and also suggested that the cloning of the PCR amplified UL2 ORF had not been successful due to errors in the restriction site sequence built into the primers. This was further tested by the experiment outlined below.

The UL2 PCR amplified fragment was restriction digested with *Nco* I/ *Bam*H I and run on a 1 % agarose gel. Two fragments of ~270 bp and ~730 bp were seen, and the smaller of the two, corresponding to the 5' end of the ORF was isolated by the Geneclean® procedure. This segment was ligated to an *Nco* I/ *Eco*R I fragment of the recombinant pHS5. This ligation mix was cleaned by running on a 1 % agarose gel and cutting out the ~1.2 kb band that corresponded to the ligation product, and this was isolated by the Geneclean® method. This was ligated to pUC19, previously restriction digested with *Bam*H I/ *Eco*R I and Geneclean® isolated from a 1 % agarose gel. *E.coli* JM109 were transformed and white colonies from ampicillin/ nutrient agar plates impregnated with IPTG and X-gal were picked for small-scale plasmid preparation. Recombinant colonies were identified by restriction digestion with *Bam*H I/ *Eco*R I, and those which yielded two bands of ~1.2 kb and ~2.7 kb were taken to be recombinants. Many such colonies were identified, the first of which was colony number 1; this was named pLig 1. This result suggested that the 5' PCR primer (coding strand primer) did indeed specify a *Bam*H I site, and this was confirmed by sequencing the 5' ~260 bp of pLig 1 (Figure **4.1.3 c**). With this, and the evidence from the end-labelling experiment, it was inferred that the 3' PCR primer (non-coding strand primer) was at fault.

4.1.4 CLONING OF THE SECOND CANDIDATE ORF OF UL2 (10156-10890)

The recombinant plasmid pHS5 was restriction digested with *Nco* I/ *Sac* I, and the ~950 bp band on a 1 % agarose gel was isolated by the Geneclean® procedure and ligated

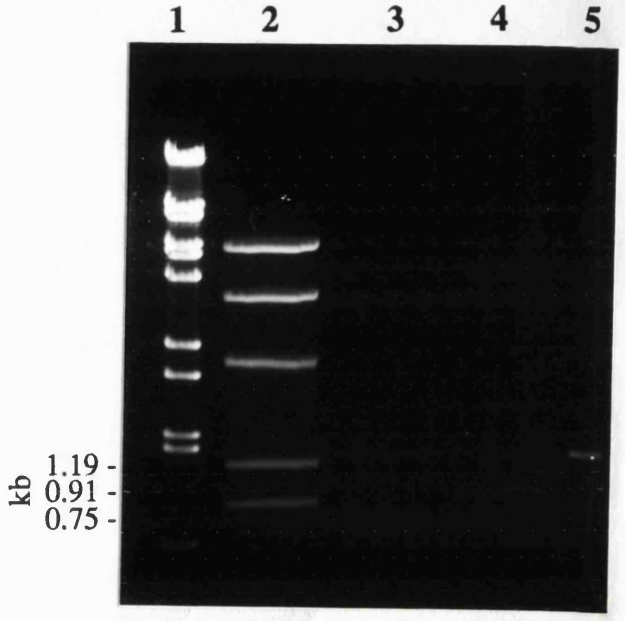


Figure 4.1.3 b

1% ethidium bromide stained agarose gel, showing that the UL2 ORF was correctly isolated from pGX23 and cloned into pUC19 to form pHS5.

Lane **1** shows the relative size marker, λ/ *Bst*E II digest.

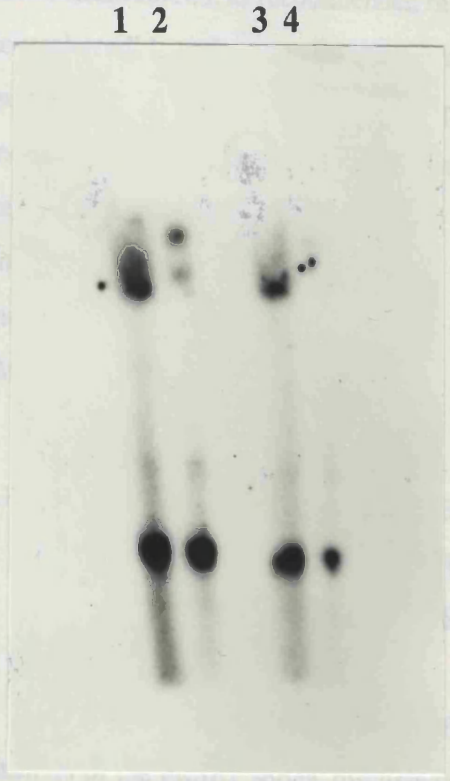
Lane **2** shows the pGX23 *Eco*R V/ *Pst* I double digest. The 1.19 kb band corresponds to UL2 (9897-11084). This was successfully cloned into pUC19 to form pHS5.

Lane 3 and lane 4 are digests confirming that unique sites present in UL2 are also present in an excised *Pst* I/ *Sac* I fragment from pHS5 (Lane 5).

Lane **3** shows that the excised fragment (Approx. 1.19 kb) is cut with *Xba* I, resulting in 0.75 kb and 0.44 kb (latter not visible) bands.

Lane **4** shows that the excised fragment (Approx. 1.19 kb) is cut with *Nco* I, resulting in 0.91 kb and 0.28 kb (latter not visible) bands.

Lane **5** shows that the *Pst* I/ *Sac* I fragment from pHS5 migrates at approximately 1.19 kb, the size of the UL2 fragment from pGX23 used to construct pHS5.



Autoradiograph showing the results of a 5' end-labelling experiment to determine the extent of digestion of the HSV1 UL2 ORF PCR product by *Bam*H I.

Lane **1** (1.2 µg) and lane 3 (0.45 µg) are undigested control samples.

Lane **2** (4.5 µg) and lane 4 (1.8 µg) are samples digested with 20 units of *Bam*H I for 24 hours at 37 C.

Autoradiograph showing the nucleotide sequence of the 5' end of the HSV1 UL2 ORF PCR product coding strand (Approx. 260 bp) present in pLig1. The second loading (**L2**) shows the sequence of the *Bam*H I site, thus the coding strand primer is not in error.

Figure 4.1.3 c

Digestion of the HSV1 UL2 ORF (9886-10890) PCR product is not complete, but the coding strand PCR primer is conclusively shown to specify the correct restriction site.

to an *Nco* I/ *Sac* I restriction digested pTrc99A plasmid isolated from a 1 % agarose gel using the Geneclean® procedure. This plasmid contains a polylinker sequence, but is optimised for expression of sequences cloned into an *Nco* I restriction site downstream of a *trc* promoter. There is no colour selection with this plasmid and so a rapid plasmid sizing screen by colony lysis was carried out. Transformation was carried out using the *E.coli* strain JM105, as recommended in the literature accompanying pTrc99A. There were several candidates for recombinants, and *Nco* I/ *Hin*d III restriction digestion of plasmids isolated by small-scale plasmid preparation from these colonies revealed several recombinants, identified from double banding of ~950 bp and ~4.2 kb on a 1 % agarose gel (Figure **4.1.4**). The first such colony identified was number 1 from a plate dated 10.6, and the plasmid was named pTS106·1. This plasmid was subsequently transformed into the *E.coli* strain JM109 for the purpose of comparing levels of UDGase activity with the JM109/ pHS5 clone.

4.1.5 CLONING OF THE FULL LENGTH (9886-10890) AND SECOND CANDIDATE (10156-10890) UL2 ORFs INTO A T7 RNA POLYMERASE/ FUSION PEPTIDE EXPRESSION SYSTEM

Both candidate ORFs, 9886-10890 and 10156-10890, were also cloned into plasmids of the pRSET series, discussed further in section **4.2.2.** These plasmids contain a polylinker sequence for insertion of genes, and three vectors exist to provide each of the three reading frames for each restriction enzyme site, namely pRSET A, pRSET B, and pRSET C.

The full length UL2 ORF (9886-10890) was cloned into pRSET A as a *Bam*H I/ *Eco*R I fragment from pLig 1. The UL2 inserted fragment from pLig 1 migrated as an ~1.2 kb band on a 1 % agarose gel following restriction digestion with *Bam*H I/ *Eco*R I, and was isolated using the Geneclean® method. The pRSET A plasmid was similarly restriction digested with *Bam*H I/ *Eco*R I, and isolated as an ~2.9 kb band from a 1 % agarose gel by the Geneclean® method. These were ligated with T4 DNA ligase overnight at 15°C and transformation of *E.coli* strain JM109 was carried out. Ampicillin resistant colonies which contained recombinant plasmids were characterised by the rapid sizing method from crude lysates, as no blue/ white colour selection is available in the pRSET plasmids. Likely colonies were then picked for small-scale plasmid preparation, and the

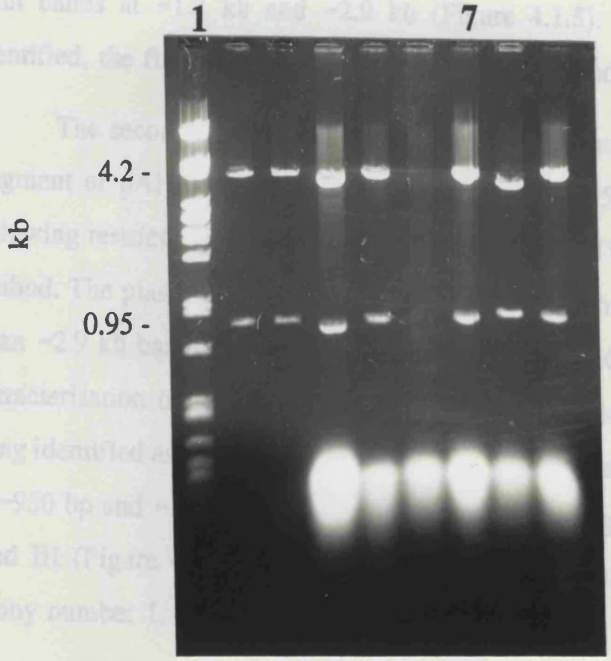


Figure 4.1.4

Identification of positive clones of the second candidate UL2 ORF in pTrc99A, prepared by plasmid miniprep and digestion with *Nco* I/ *Hin*d III.

Lane **1** shows the relative size markers, λ/ *Pst* I digest.

All other lanes show the digests, with lane **7** corresponding to pTS106.1

The plasmid pTrc99A migrates close to 4.2 kb, the UL2 fragment migrates at 0.95 kb

DNA was restriction digested with *Bam*H I/ *Eco*R I. Colonies were taken to be recombinants when the restriction digest revealed double banding on a 1 % agarose gel, with bands at ~1.2 kb and ~2.9 kb (Figure **4.1.5**). Several such recombinants were identified, the first of which was colony number 1 and this was called pAH1.

The second candidate UL2 ORF was cloned into pRSET B as an *Nco* I/ *Hin*d III fragment of pAH1. This fragment migrated as an ~950 bp band on a 1 % agarose gel following restriction digestion with *Nco* I/ *Hin*d III and was isolated using the Geneclean® method. The plasmid was also restriction digested with *Nco* I/ *Hin*d III and was isolated as an ~2.9 kb band from a 1 % agarose gel using the Geneclean® method. Cloning, and characterisation of recombinants was carried out as detailed above, with recombinants being identified as those colonies which gave rise to plasmids that yielded double banding of ~950 bp and ~2.9 kb on a 1 % agarose gel following restriction digestion with *Nco* I/ *Hin*d III (Figure **4.1.5**). Several such colonies were identified, the first of which was colony number 1, and the plasmid was named pBT1.

4.1.6 CLONING OF SEQUENCES OTHER THAN UL2

In addition to the UL2 ORFs, the *E.coli* gene *ung* was also cloned. The error rate of *Taq* DNA polymerase is high due to the lack of a 3'→5' proofreading function (Dunning *et al.*, 1988), and so the original PCR product made using *Taq* DNA polymerase (Section **4.1.1**) was not cloned. Instead, the PCR was repeated using $Vent_R$® DNA polymerase which contains a 3'→5' proofreading function, to be more sure of fidelity of the gene sequence. The PCR conditions were identical to the conditions that were successful when *Taq* DNA polymerase had been used, except that the buffer was different for $Vent_R$® DNA polymerase, and 6 mM magnesium sulphate was supplemental to this buffer for the successful PCR. Once the correctly sized ~700 bp PCR product had been obtained, which had restriction enzyme sites introduced both 5' and 3' by the primers, it was extracted by the isolation of the band from a 1 % agarose gel by the Geneclean® method. Restriction enzyme targets are *Nco* I at the 5' end and *Hin*d III at the 3' end. Restriction digestion of both the *ung* fragment and the plasmid pTrc99A with these restriction enzymes was performed. This was followed by the removal of the DNA from a 1 % agarose gel, *ung* as an ~700 bp band, and pTrc99A as an ~4.2 kb band and both were isolated by the Geneclean® procedure. Overnight ligation of a mixture of the plasmid and the *ung*

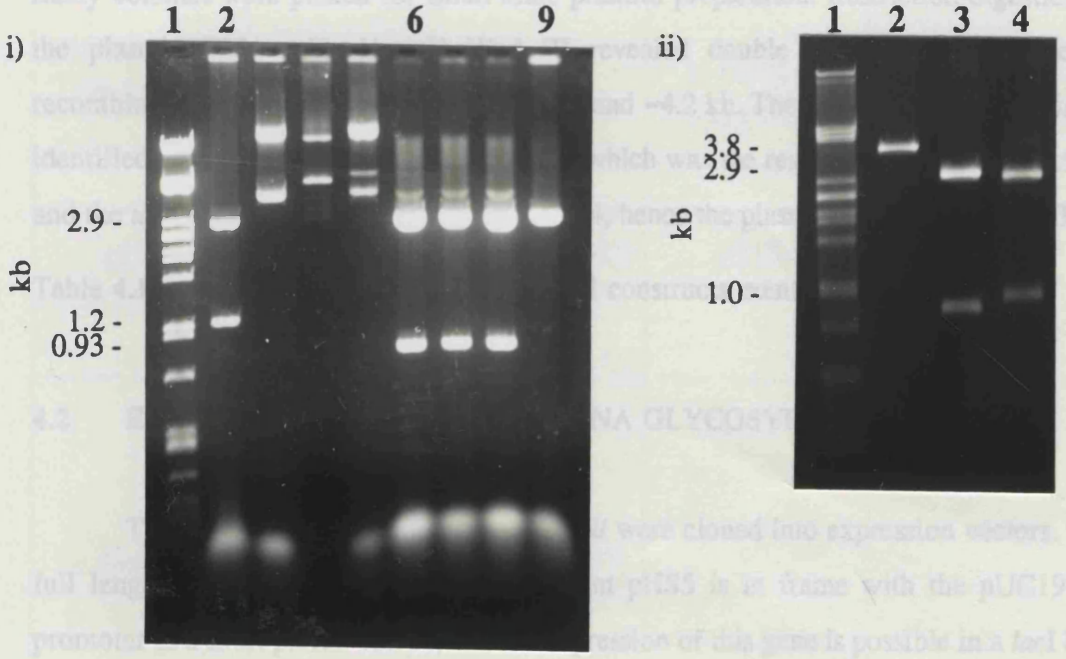


Figure 4.1.5

Identification of positive clones of the first and second candidate UL2 ORFs inserted into pRSET A and pRSET B respectively, forming pAH1 and pBT1 respectively.

i) Lane **1** shows the relative size marker, λ/ *Pst* I.

Lane **2** shows pAH1, digested with *Bam*H I/ *Eco*R I. The pRSET A plamid migrates close to 2.9 kb, and the UL2 insert migrates close to 1.2 kb.

Lane **6** shows pBT1, digested with *Nco* I/ *Hin*d III. The pRSET B plasmid migrates close to 2.9 kb, and the UL2 insert migrates close to 0.93 kb.

Lane **9** shows pRSET B plasmid linearised with *Hin*d III, close to 2.9 kb.

ii) Lane **1** shows the relative size marker, λ/ *Pst* I.

Lane **2** shows pBT1 digested with *Hin*d III, close to 3.8 kb.

Lane **3** shows pBT1, digested with *Nco* I/ *Hin*d III.

Lane **4** shows pBT1 digested with *Nde* I/ *Hin*d III. The poly His/ T7 gene 10/ porcine enterokinase/ UL2 fragment is excised, and migates close to 1 kb.

fragment was carried out in the presence of T4 DNA ligase at 15°C. The *E.coli* strain JM109 was transformed with the DNA and recombinants were selected on ampicillin/ nutrient agar plates. Rapid plasmid sizing from crude cell lysates was carried out, and likely colonies were picked for small-scale plasmid preparation. Restriction digestion of the plasmid DNA with *Nco* I/ *Hin*d III revealed double banding in the case of recombinant plasmids, with bands of ~700 bp and ~4.2 kb. The first positive recombinant identified, was colony number 2 from a plate which was the result of ligation of pTrc99A and the *ung* fragments of PCR tubes 2, 3, and 4, hence the plasmid was named pEU234·2.

Table **4.1** and figure **4.1** are a summary of all constructs mentioned previously.

4.2 EXPRESSION OF HSV1 URACIL-DNA GLYCOSYLASE IN *E.COLI*

The UDGase genes of HSV1 and *E.coli* were cloned into expression vectors. The full length UL2 ORF part of the recombinant pHS5 is in frame with the pUC19 *lac* promoter as a short partial-fusion, and the expression of this gene is possible in a *lac*I host such as *E.coli* JM109 when IPTG is added to 1 mM. The second candidate UL2 ORF and the *E.coli ung* gene are both clones in the pTrc99A expression vector as pTS106·1 and pEU234·2 respectively. These are both optimally placed to express from the *trc* promoter of the plasmid which also contains a copy of the *lac*I^q gene. Expression in a host such as *E.coli* JM109, is effected on addition of 1 mM IPTG. Both the full length and second candidate UL2 ORFs were also cloned in the pRSET vectors, and expression of fusions of the genes with a plasmid specified sequence is possible upon addition of 1 mM IPTG, in an *E.coli* host such as BL21(DE3), which can encode T7 RNA polymerase, required for expression from the T7 Φ10 promoter of the plasmid.

4.2.1 HIGH LEVEL EXPRESSION SYSTEM IN *E.COLI* USING T7 RNA POLYMERASE

The expression vector pET3a (Rosenberg *et al.*, 1987) was the initial choice of cloning/ expression vehicle. This plasmid, one of a series of plasmids derived from pBR322 (Bolivar *et al.*, 1977) with the name pET (plasmid for Expression by T7 RNA polymerase), utilises a strong T7 promoter, Φ10, to drive expression of a gene which is inserted in the counterclockwise orientation at a unique *Bam*H I site in the vector. The

Table 4.1 Summary of constructs made in this project

RECOMBINANT PLASMID Gene expression	HOST PLASMID	INSERTED DNA	SOURCE OF INSERT DNA	RESTRICTION ENZYME SITES
pHS5 No expression	pUC19	HSV1 UL2 ORF 9897-11084	pGX23	*Pst* I → *Eco*R V/ *Sma* I
pLig1 out of frame produced for nucleotide sequencing	pUC19	HSV1 UL2 ORF 9886-11084⁺	UL2 PCR 9886-10156 + pHS5 10156-11084⁺	*Bam*H I → *Nco* I + *Nco* I → *Eco*R I
pTS106·1 Overexpressed soluble crystallised	pTrc99A	HSV1 UL2 ORF 10156-11084⁺	pHS5	*Nco* I → *Sac* I
pAH1 No expression	pRSET A	HSV1 UL2 ORF 9886-11084⁺	plig1	*Bam*H I → *Eco*R I
pBT1 Overexpressed Soluble Not crystallised	pRSET B	HSV1 UL2 ORF 10156-11084⁺	pAH1	*Nco* I → *Hind* III
pEU234·2 Overexpressed Insoluble Not crystallised	pTrc99A	*E.coli ung* gene	*E.coli* JM109 genome PCR	*Nco* I → *Hind* III

NB ⁺ indicates that some additional plasmid derived sequence is appended to the 3' end of the inserted DNA.

All of the above constructed recombinant plasmids confer resistance to 100 µg/ml of the antibiotic ampicillin.

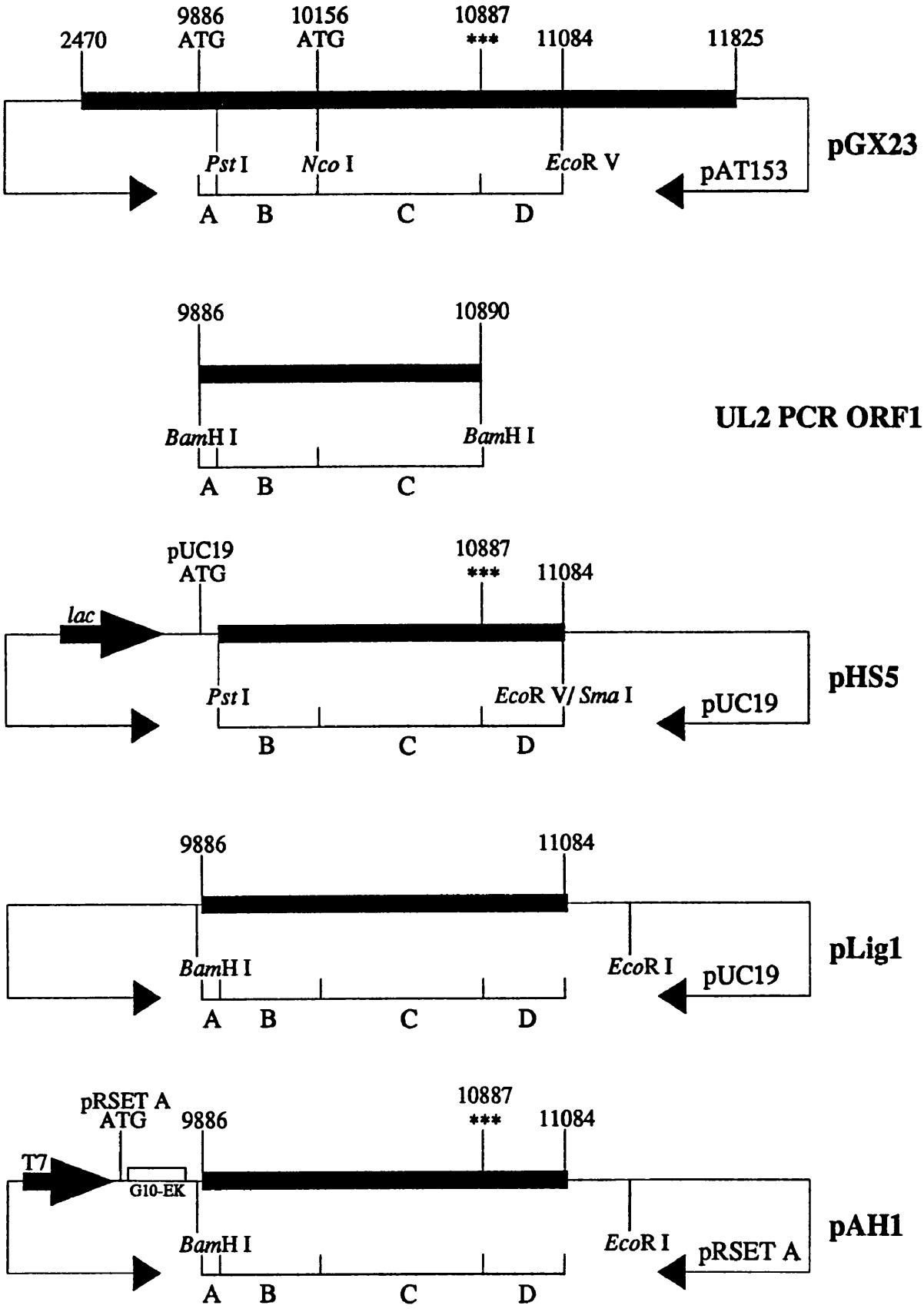


Figure 4.1

Diagrammatic representation of the recombinant plasmids constructed during this project. The plasmid pGX23 was kindly donated (Mullaney *et al*,1989) at the outset. A PCR was carried out to produce the *Bam* HI site flanked UL2 open reading frame.

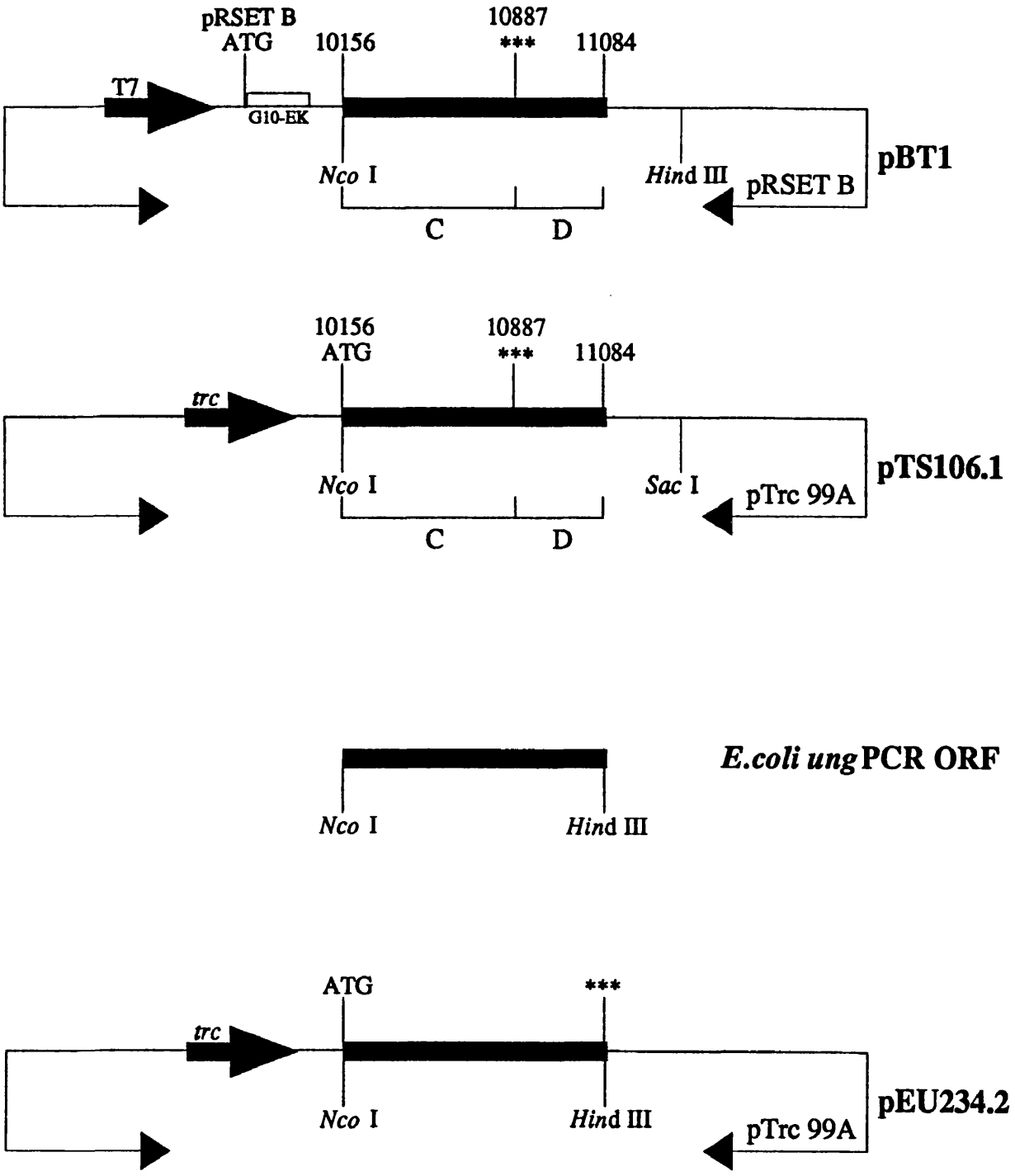


Figure 4.1 (Continued . .)

Diagrammatic representation of the recombinant plasmids constructed in this project. Promoters, translation initiation and stop codons, restriction enzyme sites, and parent plasmids are indicated. Co-ordinates refer to base pairs of the HSV1 strain 17 genome. G10EK = Sequence derived from bacteriophage T7 gene 10 + Porcine enterokinase cleavage site.

orientation is such that the gene is not transcribed by the *tet* promoter of the plasmid. Expression of recombinant protein is achieved by an IPTG inducible T7 RNA polymerase transcribed only under the control of the *lac*UV5 promoter (Maizels, 1973). Both *lac*UV5 and the T7 RNA polymerase gene are present in a DNA fragment which also contains the *lac*I gene. This DNA fragment is present in the chromosome of *E.coli* strain BL21 as part of an integrated λ-derived phage, DE3. The insertion is in the phage *int* gene, and so a stable lysogen is formed which can only be excised from the chromosome by a helper phage. The strain thus modified is referred to as BL21(DE3) (Studier and Moffatt, 1986).

This strain can be used with a chloramphenicol resistance-conferring pACYC184 plasmid (Chang and Cohen, 1978), which contains the T7 lysozyme gene inserted in the unique *Bam*H I site. T7 lysozyme, besides specifically cleaving the bond between *N*-acetylmuramic acid and L-alanine in the *E.coli* peptidoglycan (Inouye *et al.*, 1973), also inhibits transcription by binding to T7 RNA polymerase (Moffatt and Studier, 1987) which is present at low levels even before induction of cells with IPTG. The T7 lysozyme is produced under the control of a weak T7 promoter, Φ3.8 (McAllister *et al.*, 1981), thus ensuring the system is tightly regulated. The degree of control over expression by T7 RNA polymerase can be selected by choosing the pLysS or pLysE version of the pACYC184 plasmid (Studier *et al.*, 1990). In the case of pLysE, the T7 lysozyme gene is oriented such that it can be transcribed by the plasmid *tet* promoter, and so large amounts of T7 lysozyme accumulate in the cells. In addition to providing rigorous control over the expression of recombinant protein, this plasmid also weakens the host cells substantially such that growth is limited. In the case of pLysS the opposite orientation is taken by the T7 lysozyme gene, placing it under the control of Φ3.8, and so very small quantities of T7 lysozyme are produced by any T7 RNA polymerase molecules that are present prior to induction and also by *E.coli* RNA polymerase. This also has the effect of preventing further transcription by T7 RNA polymerase molecules (Studier *et al.*, 1990).

Upon induction, the level of T7 RNA polymerase far exceeds the levels of T7 lysozyme, which is under the control of the weaker Φ3.8 promoter, and the recombinant protein is transcribed vigorously after a short lag period in which existing T7 lysozyme is titrated out by the T7 RNA polymerase. The T7 RNA polymerase is able to elongate RNA chains five times faster than the *E.coli* RNA polymerase. In addition, the Φ10 promoter, unlike the Φ3.8 promoter, is very poorly recognised by *E.coli* RNA polymerase, so the expression of the protein of interest is very much limited to high level expression

following induction of cells with 1 mM IPTG. This is extremely useful in the case of recombinant proteins that are rapidly turned over by the *E.coli* host or are cytotoxic. The T7 RNA polymerase has no strong terminator sequence, however a natural terminator, TΦ, that occurs in the phage, is incorporated in the plasmid to increase transcription efficiency. As a further contingency there is an RNAse III site for cleavage of runaway transcripts.

Recombinant protein expressed in this system can constitute up to 50 % of the total cell protein within 1 to 3 hours if the mRNA is efficiently translated, and target RNA can accumulate to levels similar to that of ribosomal RNA (Studier *et al.*, 1990). A further advantage of this system is that *E.coli* BL21 is deficient in both the *lon* protease (Gottesman, 1990) and the *ompT* outer membrane protease (Grodberg and Dunn, 1988), which helps to reduce proteolytic turnover of recombinant protein. The pLysE or the pLysS plasmids also simplify the process of cell lysis to no more than a few freeze-thaw cycles.

For extremely cytotoxic gene products, the control by the T7 promoter can be harnessed by having no T7 RNA polymerase present in the host cells. The T7 RNA polymerase is delivered to the cell by a λ-derived bacteriophage CE6 in which the T7 RNA polymerase gene is under the control of the phage p_L and p_I promoters. The phage contains a thermolabile repressor and lysis mutations. It has been observed that when this phage is used to infect *E.coli* strain HMS174, the target DNA is transcribed with such vigour that the phage is not able to replicate. In the case of the UL2 gene, the inducible system with pLysS was chosen, as the UDGase was not anticipated to be cytotoxic to *E.coli* considering the high degree of homology of this protein with that produced by the host. The requirements for *E.coli* BL21(DE3) cells to maintain the plasmids are 25 µg/ml chloramphenicol for pLysS and 50 to 100 µg/ml ampicillin for pET3a. Induction of cells would normally be carried out at an OD_{600} of ~0.5 to 1.0 with 1 mM IPTG, with cell harvest 1 to 5 hours later. The pET3a plasmid was not used in the end, due to the error in the non-coding strand UL2 PCR primer. However a T7 expression system was used to express the ORF products as fusions.

4.2.2 T7 RNA POLYMERASE EXPRESSION SYSTEM FOR FUSION PROTEIN PRODUCTS THAT CAN BE PURIFIED USING IMMOBILISED METAL AFFINITY CHROMATOGRAPHY

The pRSET protein fusion expression vectors enable cloning of the gene of interest in a polylinker sequence, which is downstream of a plasmid-borne T7 gene 10 start methionine. In between these is a DNA sequence coding for a run of six histidine residues followed by a leader sequence which affords stability to RNA transcripts. This is followed by a linking DNA sequence containing a coding region for the amino acid recognition sequence of porcine enterokinase, *Asp-Asp-Asp-Asp-Lys↓Asp*. In addition, there is a transcription terminator loop at the 3' end of the polylinker sequence. The plasmid start methionine is optimally placed for transcription from a phage T7 Φ10 promoter, and like the pET3a plasmid is optimised for expression in the *E.coli* host strain BL21(DE3). The gene of interest is ligated in frame with this methionine at one of the restriction enzyme sites in the polylinker sequence. To this end there are three pRSET plasmids, pRSET A, pRSET B, and pRSET C, which cover each reading frame for any 5' restriction enzyme site chosen; the resulting recombinant protein will thus be a fusion with the residues encoded by the upstream regions. The run of six N-terminal histidine residues in the translated recombinant fusion protein enable substantial purification from a single step IMAC purification protocol, and the porcine enterokinase sequence gives the option of cleaving much of the fusion peptide away following the IMAC step. This is discussed in more detail in section **2.3.3.**

4.2.3 ATTEMPTED EXPRESSION OF THE FULL LENGTH UL2 ORF CONTAINED IN THE RECOMBINANT PLASMID pHS5

The pHS5 recombinant plasmid is a partial fusion between 9897-11084 of HSV1 and the pUC19 expression vector (Figure **4.1.3 a**). The pUC19 plasmid contains a start methionine codon, optimally placed to transcribe from a *lac* promoter, so inserted sequences must be in frame with this methionine. This *lac* promoter contains a mutation which removes catabolite repression of the system (Silverstone *et al.*, 1970; Maizels, 1973), enabling the induction process to function in rich growth media, even in the presence of glucose. The induction of the system is with 1 mM IPTG, causing expression

of the inserted gene by transcription of the sequence under the control of the plasmid *lac* promoter.

The expression of the recombinant UDGase in this system was expected to yield a protein with a molecular weight slightly greater than the ~36 000 predicted for the full length UL2 ORF due to the fact that the first four codons of the UL2 ORF have been replaced by 10 codons from pUC19.

The *E.coli* JM109 cells harbouring the plasmid were grown at 37°C in Oxoid No.2 broth, with ampicillin present at 100 µg/ml, until the cellular A_{600} was ~0.7 to 0.8 and IPTG added to 1 mM. The cells were harvested four hours later by centrifugation for 10 minutes at 4 000 *g*. Samples of cells were taken at 2 hours post inoculation, at induction with IPTG and at cell harvest, as 1 ml aliquots. These cells were pelleted and resuspended in distilled water at an A_{600} equivalent of 4.0. For example if the cell A_{600} of a 1 ml sample was 1.0, then after pelleting the cells would be resuspended in 250 µl of distilled water. From these resuspended cell samples a set volume was withdrawn for discontinuous SDS-PAGE analysis (Section **2.3.8**), the volume withdrawn depending on the size of the gel. The samples thus withdrawn were added to an equal volume of 2x sample loading buffer and frozen at -20°C. The surplus of resuspended cells was also frozen. On discontinuous SDS-PAGE gels, the 20 000 to 45 000 molecular weight area is most clearly resolved in resolving gels which have a polyacrylamide content of 12 to 12.5 %. When all samples were collected, the samples in loading buffer were boiled for 10 minutes and then loaded on the gel. The gel was stained with Coomassie brilliant blue. Following destaining of the gel, it was viewed on a light-box. Control lanes on the gel were cells grown and collected as described, but a) without plasmid, and b) with pUC19 plasmid lacking an inserted DNA fragment.

From the gel, it was clear that no new band was visible in the ~40 000 molecular weight region of the gel in the post induction lane for pHS5. This does not in itself indicate that no recombinant protein has been produced, but it does indicate that it has not been produced at high levels. To determine whether there was any excess UDGase at all, clarified supernatants from harvested post-induction pHS5 samples were assayed for UDGase activity relative to similarly treated post-induction harvest samples from controls a and b. No detectable UDGase activity in excess to that seen in the controls was detected. It was concluded, therefore, that no detectable production of the protein encoded by the full length ORF of UL2 is possible in *E.coli* JM109.

4.2.4 EXPRESSION OF THE SECOND CANDIDATE UL2 ORF CONTAINED IN THE RECOMBINANT PLASMID pTS106·1

The pTS106·1 recombinant plasmid is composed of an *Nco* I/ *Sac* I restriction enzyme digested DNA molecule from the pHS5 recombinant plasmid UL2 ORF region, ligated into an *Nco* I/ *Sac* I restriction digested pTrc99A plasmid expression vector (Table **4.1**). The inserted DNA begins at position 10156 of HSV1, which is the start methionine codon of the potential second ORF of UL2, and this is optimally positioned for expression with respect to the plasmid *trc* promoter and ribosome binding site sequences. The *trc* promoter is a *lac*/ *trp* promoter hybrid, which gives greater target sequence expression than the less efficient *lac* promoter, but still allows control by the *lac* repressor. The plasmid also directs production of the *lac*I^q repressor which serves to keep expression of the recombinant protein at very low levels in the absence of IPTG. As with pUC19, the plasmid-borne promoter contains a mutation which removes catabolite repression of the system, enabling the induction process to function in rich growth media. The induction of the system is with 1 mM IPTG, causing expression of the inserted gene by transcription of the sequence under the control of the plasmid *trc* promoter.

The *E.coli* JM109 cells harbouring the plasmid were grown at 37°C in Oxoid No.2 nutrient broth, with ampicillin present at 100 µg/ml, until the A_{600} was ~0.7 to 0.8. IPTG was then added to 1 mM. The cells were harvested four hours later by centrifugation for 10 minutes at 4 000 *g*. Samples of cells were taken at 2 hours post inoculation, at induction with IPTG and at cell harvest, as 1 ml aliquots. These cells were pelleted and resuspended for SDS-PAGE analysis as outlined in section **4.2.3**. Following destaining of a Coomassie brilliant-blue stained 12 % SDS-PAGE gel, it was clear that the ~27 300 mass protein product predicted from the second candidate UL2 ORF had been expressed to high levels. An intense band had appeared in the post-induction lane at a point which was free of such a band in the pre-induction and induction point lanes.

A clarified supernatant sample from 1 mM IPTG induced *E.coli* JM109 cells without any plasmid had no detectable UDGase activity at concentrations less than 0.1 mg/ ml. However, significant UDGase activity could still be detected in a 1 in 100 dilution of a 0.1 mg/ ml clarified supernatant from 1 mM IPTG induced *E.coli* JM109 cells harbouring pTS106·1. The result suggested that the product of the second candidate ORF of UL2 was a functional UDGase that could be produced to high levels in *E.coli*

(Figure **4.2.4**).

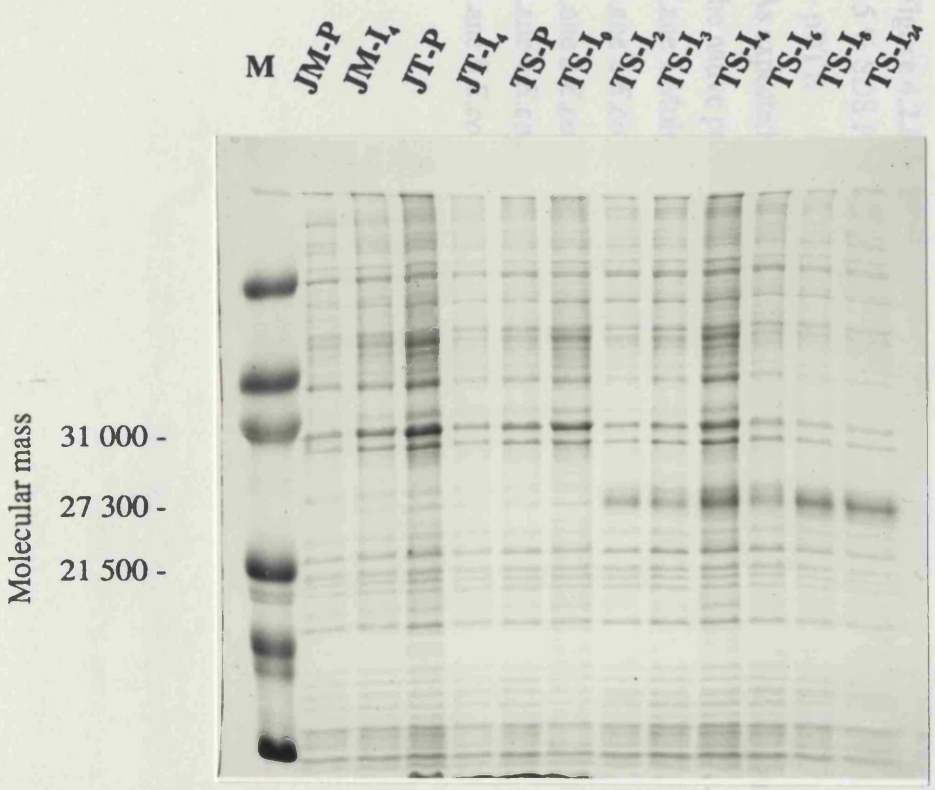
An experiment to determine whether the HSV1 UDGase was fully soluble in *E.coli* was carried out. The whole-cell post-induction sample was subjected to ultra-sonication in order to fully lyse all cells. The lysate was then subjected to centrifugation at 15 000 *g* for 30 minutes and the supernatant and pellet fractions were separately loaded on a SDS-PAGE gel alongside lysed uncentrifuged cells and molecular weight markers. Following destaining of the Coomassie brilliant blue stained gel, it was clear to see that the HSV1 UDGase was apparently all present in the supernatant fraction of *E.coli* (Figure **4.2.4**).

4.2.5 ATTEMPTED EXPRESSION OF THE FIRST CANDIDATE UL2 ORF CONTAINED IN THE RECOMBINANT PLASMID pAH1, AND EXPRESSION OF THE SECOND CANDIDATE UL2 ORF CONTAINED IN THE RECOMBINANT PLASMID pBT1

The recombinant plasmids pAH1 and pBT1 contain the first and second candidate ORFs of UL2 respectively. Both ORFs were initially retrieved from the recombinant plasmid pHS5 and cloned into pRSET A and pRSET B respectively. Detailed discussion of the pRSET plasmids may be found in section **4.2.2**, and the T7 expression system is fully discussed in section **4.2.1**.

The results of attempted expression of the first candidate ORF of UL2 contained in pAH1 are similar to those for the attempted expression of this ORF in pHS5. The plasmid was not found to express a UDGase that could be detected on Coomassie brilliant blue stained SDS-PAGE gels, or by assay of crude cell lysates against *E.coli* native cell controls using the procedures detailed in section **4.2.3**.

The recombinant plasmid pBT1 was found to express high levels of UDGase activity and a new band was visible on a Coomassie brilliant blue stained SDS-PAGE gel in the post-induction sample lane, corresponding to a molecular mass near the expected ~ 31 800 of the fusion product of T7 gene 10/ poly-His/ porcine enterokinase peptide (~4 500) and the product of the second candidate UL2 ORF (~27 300)(Figure **4.2.5**). The recombinant protein was found to be in the soluble fraction upon sonication/ centrifugation



12 % SDS PAGE analysis showing that a new protein is expressed to high levels in *E.coli* JM109 upon induction with 1 mM IPTG. The lanes are identified by **JM** corresponding to unmodified *E.coli* JM109, **JT** corresponding to *E.coli* JM109+pTrc99A, and **TS** corresponding to *E.coli* JM109+pTS106.1. Codes **P**= 2 hours prior to induction, $\mathbf{I_0}$= At Induction, $\mathbf{I_n}$= n hours post induction. **M**= Markers.

The new protein produced by *E.coli* JM109+ pTS106.1 (Lanes **2** to **7**) is in the soluble fraction (**S**). No new band is visible in *E.coli* JM109+ pHS5 (Lanes **8** to **10**). **P**= Pellet, **W**= Whole cells.

Figure 4.2.4

12 % SDS PAGE analysis reveals that a new soluble protein is expressed in *E.coli* cells containing pTS106.1, but not in *E.coli* cells containing pHS5.

Figure 4.2.5

15 % SDS PAGE analysis showing that a new protein is produced in *E.coli* BL21 (DE3) + pBT1.

As expected, the protein has an increase in molecular mass of approximately 4 500 over the native protein produced by *E.coli* strains containing pTS106.1.

Lane **1** Molecular mass markers.

Lane **2** *E.coli* BL21 (DE3)+ pBT1 at induction.

Lane **3** *E.coli* BL21 (DE3)+ pBT1 4 hours post induction.

Lane **4** *E.coli* JM109+ pTS106.1 4 hours post induction.

Lane **5** *E.coli* JM109+ pTS106.1 at induction.

of the post-induction sample. The protein exhibited UDGase activity in this fused state. Activity was still detectable in a 1 in 10 dilution of a 0.1 mg/ ml clarified supernatant, whereas a control sample gave no measurable activity at concentrations of less than 0.1 mg/ ml.

4.3 PURIFICATION OF HSV1 URACIL-DNA GLYCOSYLASE FROM CELLS OF *E.COLI*

This study is aimed at obtaining protein for the determination of its tertiary structure, and this would be relevant to an entire class of highly conserved proteins of hitherto unknown structure. Consequently it would be desirable to avoid unfolding/ refolding the protein if it is avoidable. This is because the formation of diffraction quality crystals is often reliant on the protein being in a homogeneous state (Giegé *et al.*, 1986; Wood, 1990; Lorber and Giegé, 1992). This is not guaranteed in a re-folded protein, as many partially folded, yet soluble, forms of the protein may exist alongside the correctly folded protein. Though it is not always necessary, when seeking to crystallise a protein it is often beneficial to have purified it to at least 95 % of total protein. As with kinetic characterisation, it is not always possible to predict which contaminating species will interfere with the effects being studied, in this case, the specific interactions that influence crystal formation. Furthermore the purer the sample, the easier it is to follow the effect of added substances.

The purification of the highly expressed HSV1 UDGase from *E.coli* cells was attempted both by conventional approaches (Section **2.3.2**), and by a specific affinity technique relying on the presence of an amino acid sequence fused to the N-terminus of the recombinant protein (Section **2.3.3**).

4.3.1 IMMOBILISED METAL AFFINITY CHROMATOGRAPHY AS A SINGLE STEP PURIFICATION PROCEDURE OF THE PROTEIN OVEREXPRESSED BY THE RECOMBINANT BL21(DE3)/ pBT1

The pRSET fusion peptide/ expression vectors are designed to engineer a sequence of six histidine residues onto the N-terminus of the recombinant protein being expressed. This region can subsequently be enzymatically removed following purification. The

purification is carried out on a column matrix called PROBOND® resin which is a chelated Ni^{2+} based resin (Section **2.3.3**).

Following cell growth, induction with IPTG, cell lysis by freeze-thaw/ ultrasonication, removal of nucleic acids with streptomycin sulphate and cell debris/ aggregates by high-speed centrifugation, the resulting clarified supernatant was loaded onto a PROBOND® column pre-equilibrated with the recommended buffer, 40 mM sodium phosphate buffer, 0.5 M sodium chloride pH 8.0. Cell resuspension and lysis had been carried out in this buffer beforehand. Upon stepwise elution at pH 7.0, 6.0, 5.0, and 4.0, it was clear that no peak corresponding to the overexpressed protein was visible. The Coomassie brilliant blue stained SDS-PAGE gel of the crude cell lysate, supernatant, pellet, and all elutions from the column showed that the overexpressed protein was in the cell pellet after cell lysate separation. The conclusion drawn from this was that the protein was insoluble in sodium phosphate buffer. A solubility experiment revealed that the protein was soluble when cells were lysed in Tris·HCl buffer at pH 8.0, and crude cell resuspensions in distilled water also showed the protein to be in the soluble fraction.

It is not possible to achieve a satisfactory pH based elution down to pH 4.0 using Tris·HCl buffer, which buffers between pH 7.0 and pH 9.1 at 25°C, and so a buffer which allowed the protein to remain in solution and which offered a broad pH range with minimal ionic strength fluctuation was required. Such broad range, constant ionic strength buffers have been described previously (Ellis and Morrison, 1982). The buffer selected was Tris/ MES/ acetate buffer, which maintains a constant ionic strength equivalent to 0.1 M at 25°C, in the pH range 9.1 to 3.7. This buffer was modified by the addition of 15 % glycerol, 0.5 M sodium chloride, and dilution to a final buffer ionic strength equivalent to 0.05 M. The overexpressed protein was found to be soluble in this buffer at pH 8.0 and so the purification procedure was attempted once more. This time, the Coomassie brilliant blue stained SDS-PAGE gel showed that the overexpressed protein eluted at pH 4.0 (Figure **4.3.1**). Several solubility problems were encountered however, one of which was that the protein displayed limited solubility at pH 4.0. Attempting to correct this by dialysis into buffer at pH 8.0 resulted in the irreversible precipitation of the protein, as did immediately changing the pH back to pH 8.0 by adding concentrated Tris·HCl buffer pH 8.0 or by adding sodium hydroxide solution. It was concluded that the pH changes required were detrimental to the purification of the protein, and so a different elution strategy was used.

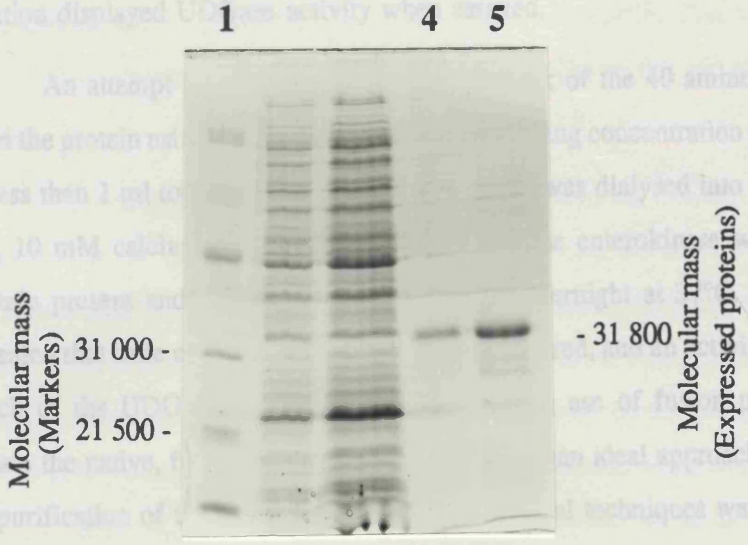


Figure 4.3.1

15 % SDS PAGE analysis of the extent of purification in a single column IMAC step of the recombinant protein from*E.coli* BL21 (DE3)+ pBT1.

Lane **1** Molecular mass markers.

Lane **4** Elution at pH 4.0 of soluble fraction of lysed cells bound to IMAC column after the column was previously washed at pH 7.0, 6.0, and 5.0. Precipitation of the eluted protein results in a much diminished yield compared to that shown in lane 5.

Lane **5** Elution at less than 250 mM, from a gradient of imidazole. The extent of loss of protein by precipitation after dialysis is far less than the pH eluted sample (lane 4).

The modified Tris/ MES/ acetate buffer at pH 8.0 was used for cell lysis and column equilibration. This time however, the elution was carried out by a gradient elution against the same buffer which contained 1 M imidazole, also at pH 8.0. The major protein peak was eluted below 250 mM imidazole and on subsequent SDS-PAGE analysis was revealed to be the overexpressed protein, substantially purified (Figure **4.3.1**). Dialysis of the eluted protein into the modified Tris/ MES/ acetate buffer pH 8.0 without imidazole resulted in some loss of protein due to insolubility, but the sample which remained in solution displayed UDGase activity when assayed.

An attempt was made at removing the bulk of the 40 amino acid fusion peptide from the protein using porcine enterokinase. Following concentration of the protein sample to less than 1 ml total volume, the protein sample was dialysed into 10 mM Tris·HCl pH 8.0, 10 mM calcium chloride, then 1.0 µg porcine enterokinase was added per mg of protein present and the mixture was incubated overnight at 37°C. SDS-PAGE analysis revealed that little cleavage of the protein had occurred, and an activity assay revealed that much of the UDGase activity had been lost. The use of fusion peptide techniques to obtain the native, fully active protein is clearly not an ideal approach in this instance and so purification of the native protein by conventional techniques was attempted.

4.3.2 PURIFICATION STRATEGIES FOR HSV1 URACIL-DNA GLYCOSYLASE FROM CELLS OF *E.COLI*

The initial strategy for purification was based upon the homology of this protein to other UDGases. Purification protocols for the UDGases from *E.coli*, human, *Saccharomyces cerevisiae*, and other organisms (Section **1.2**) tended to have in common the techniques of fractional precipitation by ammonium sulphate addition, size fractionation by gel filtration, ionic charge based separation and DNA affinity. The buffers used tended to include glycerol at 10-20 % v/v in order to allow freezing of samples and to improve stability of pure fractions. These buffers also included DTT at 1-2 mM to prevent oxidation of free sulphydryl groups on cysteine residues. Purification of protein is generally carried out at 4°C to maintain stability and minimise proteolysis.

To obtain overexpressed HSV1 UDGase from *E.coli* cells harbouring the pTS106·1 plasmid, a pre-culture of 5 ml of 1.5 x Oxoid No.2 broth containing 100 µg/ml ampicillin was set up using 20 µl *E.coli*+pTS106·1 glycerol stock. This was incubated on a shaking

table to mid-logarithmic phase at 37°C, then 1 ml aliquots were added to two 5 litre flasks containing 500 ml each of 1.5 x Oxoid No.2 broth containing 100 µg/ml ampicillin and 1 mM IPTG (Initial experiments in the optimisation of the purification protocol were conducted using 100 ml *E.coli* cell cultures). These were then incubated with shaking at 37°C for 16 to 18 hours. The cells were then harvested by centrifugation at 4 000 *g* for 10 minutes, and the pellets washed gently with de-ionised water. Lysis was carried out by several 45 second pulses with 30 second intervals using ultrasonication in 30 ml of purification buffer (Section **2.3.5**) containing 1 mM PMSF. This was done in a chilled water-ice environment to prevent excessive heating of the sample by the ultrasound energy.

The UDGase overexpressed by the *E.coli* harbouring pTS106·1 contains 244 amino acids, and is the protein that would be expressed from the second potential ORF in the UL2 gene of HSV1. The protein has a predicted isoelectric point of 9.52 as determined by the PEPTIDESORT program in the GCG computer program package (Devereux *et al.*, 1984). Measurement of the isoelectric point was attempted using Immobiline dry-strip pH 3.5-10 gels (Pharmacia LKB Biotechnology). The protein was found to be very basic, but the calibration at the basic end of the gel was not very good and it is not possible to assign an exact value, however the measured isoelectric point is greater than 9.

Using this data, it can be predicted that the net surface charge will be positive at pH values below half a pH unit under the isoelectric point. This would mean that the protein would not bind to an anion exchange column such as DEAE-cellulose at pH 8. This is a very useful state of affairs since the majority of alkali soluble *E.coli* proteins will bind to such a column at pH 8. It was decided that following cell lysis, nucleic acid removal and sedimentation of insoluble matter, the cell supernatant fraction would be passed through a DEAE cellulose column. The initial buffer pH for the purification was 7.65, a good deal lower than the isoelectric point. The purification buffer consisted of 15 mM Tris·HCl pH 7.65, 15 % v/v glycerol, 1.5 mM DTT.

Upon passing the supernatant fraction through a column packed with Whatman DE52 cellulose pre-equilibrated with the purification buffer, the unbound protein eluate was collected. This fraction was very dilute, but an aliquot was vacuum concentrated and run out on a discontinuous SDS-PAGE gel of 12 % acrylamide in the resolving gel. The Coomassie staining revealed a heavy band below the 31 500 molecular mass marker, and only a very few other bands which were much fainter. Furthermore, the unbound fraction

of this step exhibited UDGase activity when assayed (Table 4.3.2).

The next set of experiments concentrated on trying to fractionate the unbound fraction with ammonium sulphate. An initial attempt at this was to use 35 % ammonium sulphate and discard the pellet, followed by 65 % ammonium sulphate and discard the supernatant. Dialysis was then carried out to return the solution to the original purification buffer conditions. This step when viewed on a gel showed that the heavy band at just below 31 500 was still present and all other bands, which were mainly of lower molecular weight, were hardly visible. The gel was scanned on a 'Joyce Loebl Chromoscan 3' optical gel scanner and the area under peaks output showed the heavy band to account for 96 % of protein in the gel lane. The major band was analysed by N-terminal sequence analysis (By Dr. N. F. Tottie at the Ludwig Institute for Cancer Research, London W1) and was shown to be identical in the first 10 residues with the polypeptide product predicted for this ORF of UL2 (Figure 4.3.2 a).

The purification protocol was repeated a number of times, but the result could not be repeated (Figure 4.3.2 a). The purity as assessed by optical scan was anywhere between 29 and 85 %. The main difficulties associated with the ammonium sulphate fractionation were that the initial supernatant protein concentration, prior to fractionation, was critical and the protein would often precipitate on dialysis back into the purification buffer.

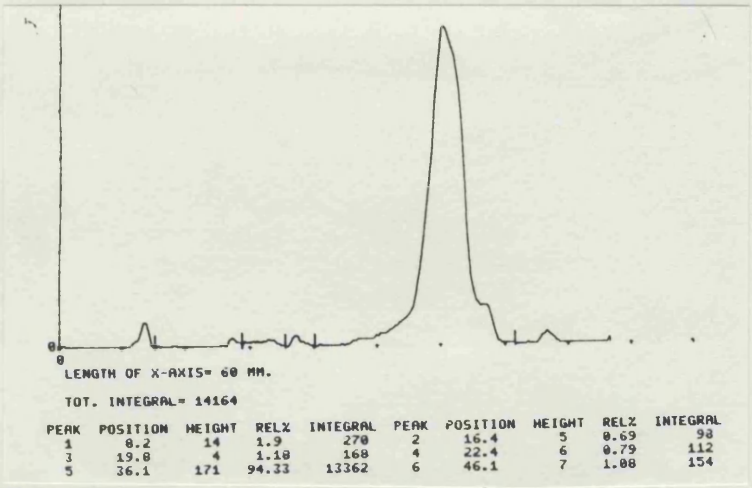
A new approach was constructed involving the affinity of the protein to various chromatography resins. The first such resin used was a dye resin, Affigel Blue which exhibits binding properties with proteins that recognise and bind to nucleotides (Section 2.3.2). The interaction of a protein with this resin can be purely charge based however, because the resin is negatively charged. UDGase was found to bind to Affigel Blue, and was competed off with an increasing gradient of sodium chloride from 0 to 2 M, prior to dialysis back into salt-free purification buffer. This step resulted in a modest purification, but it was felt that an additional step would be required. The rationale behind the next step was that the charge of the protein should be exploited again. The protein was passed through a Pharmacia Mono-S cation exchange column, to which it bound, and was competed off with an increasing gradient of sodium chloride from 0 to 2 M, prior to dialysis into salt-free purification buffer. The Mono-S column provided a good purification in the fractions between 1.6 to 2 M sodium chloride, but the UDGase eluted between about 0.8 to 2 M sodium chloride. In the earlier fractions it came off with other proteins, yet it was still eluting at very strong salt concentrations. It was felt that the affinity for

i) M D L T N G G V S ? A

21 9 18 14.5 9 17 21 16 8 ? 12

ii) 1 2 3 4 5 6 7

iii)



PEAK	POSITION	HEIGHT	RELX	INTEGRAL	PEAK	POSITION	HEIGHT	RELX	INTEGRAL
1	0.2	14	1.9	270	2	16.4	5	0.69	98
3	19.8	4	1.18	168	4	22.4	6	0.79	112
5	36.1	171	94.33	13362	6	46.1	7	1.08	154

Figure 4.3.2 a

i) N-terminal sequence analysis of the protein produced by *E.coli* JM109+ pTS106.1 on induction with 1 mM IPTG, following substantial purification, shown in (ii). Sequencing was achieved by automated Edman degradation, with on-line detection of phenylthiodantoin derivatives at 214 nm.

The sequencing was carried out on approximately 25 to 30 picomole of a sample blot on a PVDF membrane, from a 12 % SDS polyacrylamide gel. The figures under the single letter abbreviations for amino acids are values in picomoles detected.

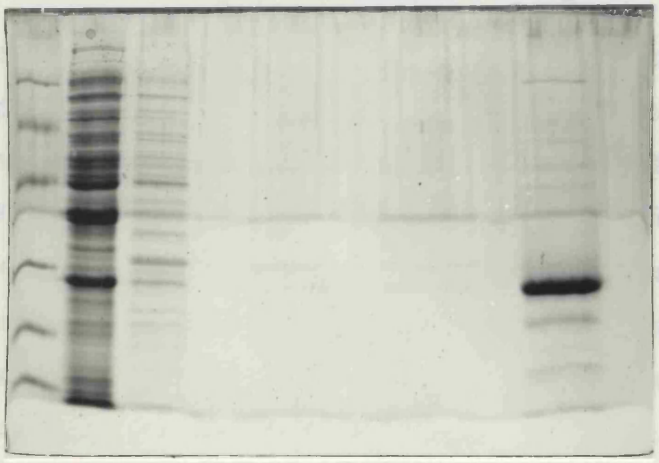
ii) 12 % SDS PAGE analysis shows that the initial attempt at purification of HSV1 UDGase from *E.coli* JM109 +pTS106.1 resulted in a highly purified preparation.

Lane **1** shows molecular mass markers, lanes **2** and **3** show induced whole cells.

Lane **4** shows the soluble fraction of these cells, lane **5** shows the unbound fraction after DE52 column chromatography. Lanes **6** and **7** show the resuspended proteins after 35 % and 65 % ammonium sulphate fractionation respectively.

iii) The scanning densitometer trace for the gel shown in (ii), showing that the major band accounts for over 94 % of the protein in lane 7.

iv) 1 5 6



v) 2 3 5

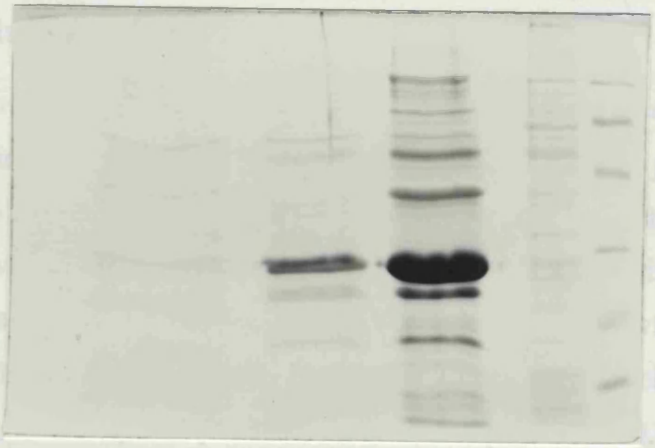


Figure 4.3.2 a (Continued . .)

The high degree of purification achieved in (ii) was not very reproducable.

iv) A further fractionation at 55 %, lane **5**, did not remove any other proteins, and the 65 % fractionation (lane **6**) on this occasion resulted in protein less than 60 % pure as judged by scanning densitometric analysis. Lane **1** shows the molecular mass markers.

v) A further fractionation at 60 %, lane **3**, instead of the 35 % and 55 % fractionations, resulted in the majority of the protein precipitating in this step, at a purity of less than 25 % as judged by scanning densitometric analysis. The sample which precipitated at 65 % was still only 85 % pure (Lane **2**). Molecular mass markers are shown in lane **5**.

the Mono-S was too strong and that a weaker cation exchanger should be attempted.

In the meantime, the Mono-S purified fractions were subjected to gel filtration chromatography on a superdex-75 column, which fractionates on the basis of molecular mass between 3 000 and 75 000. The purification buffer was slightly modified for this step, with 500 mM sodium chloride added to prevent non-specific binding of protein to the matrix. However, the UDGase activity eluted mainly in the void fraction. Activity was also present across the whole range of fractionation, which was at virtually base-line levels except for a single peak in the middle of the run which contained proteins of a variety of molecular weights. The step was repeated with 1 M sodium chloride, but the same effect was observed. A second protein preparation not passed through the Mono-S step also exhibited this behaviour. It appears that at the high concentrations used in this step, owing to restrictions on the volume of sample which can be loaded for optimum resolution, the proteins are non-specifically aggregated. The gel filtration step was thus omitted from future protocols.

The purification protocol was redesigned to make full use of the surface charge properties of the UDGase. The new protocol involved the change of the purification buffer to 20 mM Tris·HCl pH 8.25, 15 % glycerol, 1.5 mM DTT. The cell lysis and nucleic acid precipitation with removal of insoluble debris by high speed centrifugation was still followed by the collection of unbound protein from the DE52 column which had been pre-equilibrated with the new purification buffer. The second step was changed to a weak cation exchanger, carboxy-methyl cellulose to which the protein bound. The UDGase was competed off with an increasing gradient of sodium chloride from 0 to 1.6 M, prior to the active fractions being dialysed twice against 200 volumes of salt-free purification buffer. These were then applied to an Affigel Blue column pre-equilibrated with purification buffer, and the protein eluted with an increasing gradient of sodium chloride from 0 to 1.6 M. The active fractions were then dialysed twice against 200 volumes of salt-free purification buffer.

Analysis of the preparation at this stage on a Coomassie blue stained 12 % discontinuous SDS-PAGE gel showed that the preparation was highly purified (Figure **4.3.2 b**) with only three minor bands of lower molecular mass than UDGase contaminating the preparation (Figure **4.3.2 c**). The next test of the purity of the preparation was to attempt crystallisation of the protein (Section **5.1.1**). Though this was initially successful, it was thought that the quality of the crystals could be improved by removing the trace

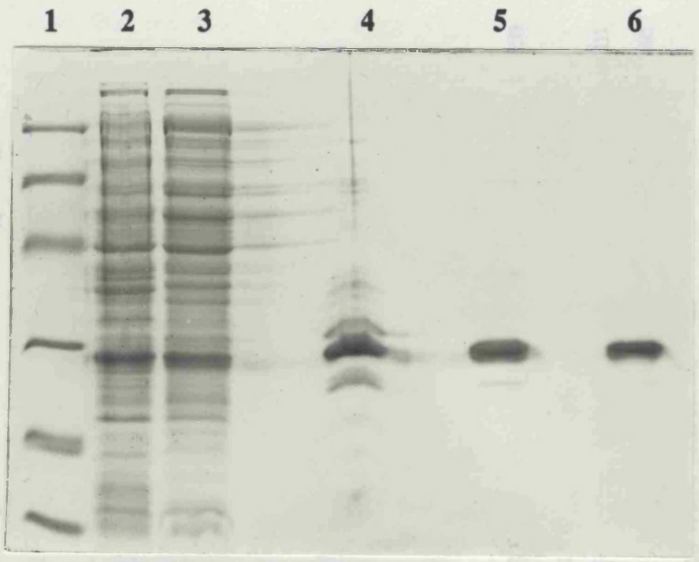


Figure 4.3.2 b

12 % SDS polyacrylamide gel showing the stages in the purification of HSV1 UDGase. The poly.U sepharose step was incorporated at a later date, and examples of the purity achieved with this final step can be seen in figure **4.3.2 c.**

Lane **1** shows the molecular mass markers, lane **2** shows the crude cell lysate at cell harvest. Lane **3** shows the soluble fraction of these cells. Lane **4** shows the unbound fraction following DE52 column chromatography. Lane **5** shows the eluted peak from carboxy-methyl cellulose column chromatography. Lane **6** shows the eluted peak from Affigel Blue column chromatography. The preparation is greater than 95 % pure after this procedure, as judged by scanning densitometry.

Table 4.3.2

Summary of the purification of HSV1 UDGase.1 unit of activity is defined as 1 cpm/ min/ ng

STEP	VOL ml	[Protein] mg/ ml	Protein mg total	Activity units total	Activity units/ mg	Purity (fold)	% yield
Crude Lysate	28	-	-	-	-	-	-
Supernatant	25	80	2000	1.08×10^{13}	5.38×10^{9}	1.0	100.0
DE52	670	0.45	301.5	1.07×10^{13}	3.55×10^{10}	6.6	99.0
CM cellulose	36	2.65	95.4	4.26×10^{12}	4.47×10^{10}	8.3	39.4
Affigel Blue	38	1.52	57.76	2.93×10^{12}	5.07×10^{10}	9.4	27.1
Poly-U Sepharose	68	0.733	49.84	2.75×10^{12}	5.52×10^{10}	10.3	25.5

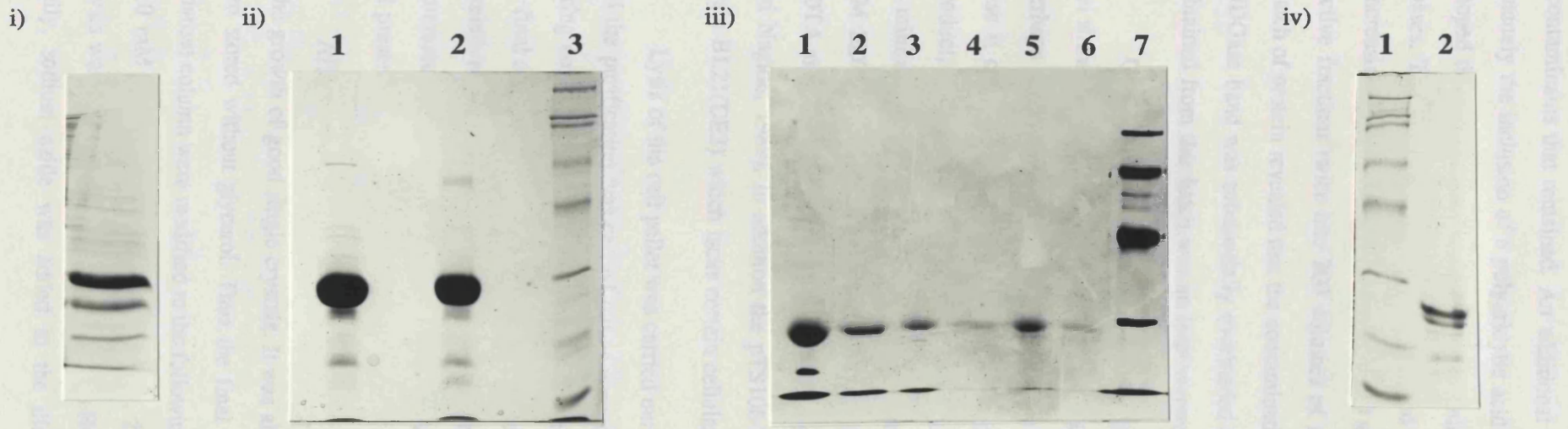


Figure 4.3.2 c

i) A lane from a 12 % SDS polyacrylamide gel showing the appearance of a 2 week old sample of HSV1 UDGase stored at 4°C. This sample has been purified and stored in the glycerol-containing buffer, and was produced in *E.coli* JM109. It was not passed through the poly-U Sepharose column.

ii) Samples purified with the addition of a poly-U sepharose column pass as the final step. The storage buffer still contained glycerol, and the protein was also produced in *E.coli* JM109. Lane **3** shows the molecular mass markers. Lane **2** shows the sample 1 day after purification. Lane **1** shows the sample after 2 weeks storage. The samples were overloaded to emphasise contaminating bands. The samples were stored at 4°C.

iii) Silver stained 12 % SDS polyacrylamide gel showing 3 samples of poly-U sepharose purified HSV1 UDGase, produced in *E.coli* BL21(DE3). The buffer used for purification contained protease inhibitors, and the storage buffer contained protease/ microbial growth inhibitors, and no glycerol. Lane **1** and lane **2** show the first sample after storage at 4°C for a week, the first lane was overloaded. Lanes **3** and **4** show the second sample, after similar storage. The first lane was overloaded with respect to the second. Lanes **5** and **6** show the third sample, after similar storage, with the first lane overloaded compared to the second. Lane **7** shows the molecular mass markers.

iv) 12 % SDS PAGE analysis of a dissolved HSV1 UDGase (native) crystal (lane **2**). Lane **1** shows the molecular mass markers.

contaminants that remained. An additional step was added to the purification protocol, namely the inclusion of a polyuridylic acid sepharose 4B column as a final step. It was hoped that the protein would have an affinity for this resin due to the presence of uracil bases. This turned out to be the case, and the bound protein could be eluted with an increasing sodium chloride gradient from 0 to 0.5 M. This was followed by dialysis of the active fractions twice into 200 volumes of salt-free purification buffer. Analysis of this batch of protein revealed that the contaminating bands were now only visible if the major UDGase band was substantially overloaded (Figure **4.3.2 c**). The quality of the crystals obtained from this batch was an improvement over the less pure batch.

Crystals which had been washed and run out on a 12 % discontinuous SDS-PAGE gel showed a trace of the lower molecular mass band (Figure **4.3.2 c**). It was thought that perhaps the lower molecular mass band was a premature termination product, in which case it could not be removed. The other possibility was that it was a proteolysis by-product, in which case the following modification to the purification buffer would help to minimise its presence. The purification buffer was modified to 20 mM Tris·HCl/ 10 mM EDTA pH 8.25, 15 % glycerol, 1.5 mM DTT, 0.1 mM PMSF. The presence of EDTA and PMSF will inhibit metalloproteases and serine proteases respectively (Salvesen and Nagase, 1989). In addition the pTS106·1 plasmid was transformed into the *E.coli* strain BL21(DE3) which lacks certain cellular protease systems (Section **4.2.1**).

Lysis of the cell pellet was carried out promptly in the presence of 1 mM PMSF, and the purification was carried out as normal except for the modification of the buffer during column steps and dialysis. A silver stained gel of the UDGase after elution from the final column showed that the lower molecular mass bands had decreased substantially in relative intensity (Figure **4.3.2 c**). This indicated that bands remaining were either due to premature termination of translation or, more likely, that proteolysis products, although still present, had been substantially reduced.

After problems with the crystallisation, it was found that EDTA was detrimental to the growth of good single crystals. It was also found that the protein was more stable when stored without glycerol. Thus the final dialysis steps after the polyuridylic acid sepharose column were modified to the following. The first dialysis was into 200 volumes of 10 mM Tris·HCl pH 8.25, 7 % glycerol, 2 mM DTT, 0.1 mM PMSF. The second dialysis was into 200 volumes of 5 mM Tris·HCl pH 8.25, 2 mM DTT, 0.1 mM PMSF. Finally, sodium azide was added to the dialysed, concentrated sample to a final

concentration of 0.02 % w/v. The preparation appears to be completely stable, as judged by enzymatic activity and ability to form large single crystals, for over 8 months when stored at 4°C. Typical yields of the pure enzyme per litre (~6.5 to 7.5 grams wet weight) of *E.coli*, regardless of the strain used, are 20 to 40 mg though some preparations have yielded between 50 and 120 mg. Details of the purification procedure are shown in table **4.3.2**, figure **4.3.2 b**, and figure **4.3.2 d**.

4.3.3 PURIFICATION STRATEGIES FOR THE *E.COLI* URACIL-DNA GLYCOSYLASE OVEREXPRESSED AS A RECOMBINANT PROTEIN IN CELLS OF *E.COLI*

This protein was produced from *E.coli* carrying the plasmid pEU234·2 (Section **4.1.6**). When expressed by induction of the bacterial culture with 1 mM IPTG with agitation at 37°C, the protein is produced in large quantities as judged by Coomassie Blue stained SDS-PAGE gels. It runs at a position intermediate between the 21 000 and 31 500 molecular mass markers, and just a little lower than the overexpressed HSV1 UDGase which has a predicted molecular mass of 27328. The *E.coli* UDGase is a protein of 229 amino acids with a predicted molecular mass of 25664, so this result matches the relative molecular masses of the proteins quite well.

The protein thus produced however, is insoluble, with no visible enhancement of the band position in the supernatant fraction when a Coomassie Blue SDS-PAGE gel is viewed. In fact, the whole cells show a much enhanced band, as does the pellet fraction (Figure **4.3.3**). There is however, a greater than two-fold increase in the levels of detectable UDGase activity in the clarified supernatant fraction of 1 mM IPTG induced *E.coli* cells harbouring pEU234·2, with respect to a control sample from cells without a plasmid. Alternative expression conditions were attempted to try to increase the levels of this soluble, active protein. This protein would be used to compare activity of the *E.coli* enzyme with the HSV1 enzyme under the assay conditions used, as defined in section **2.3.10**. If expression was good enough, the protein would also be purified under a purification protocol similar to a published one (Lindahl *et al.*, 1977) in order to attempt crystallisation of the protein. This protein might prove easier to find isomorphous heavy-atom derivative crystals for (Section **2.4.4**).

The first alternative tried was to attempt overexpression of the protein with 1 mM

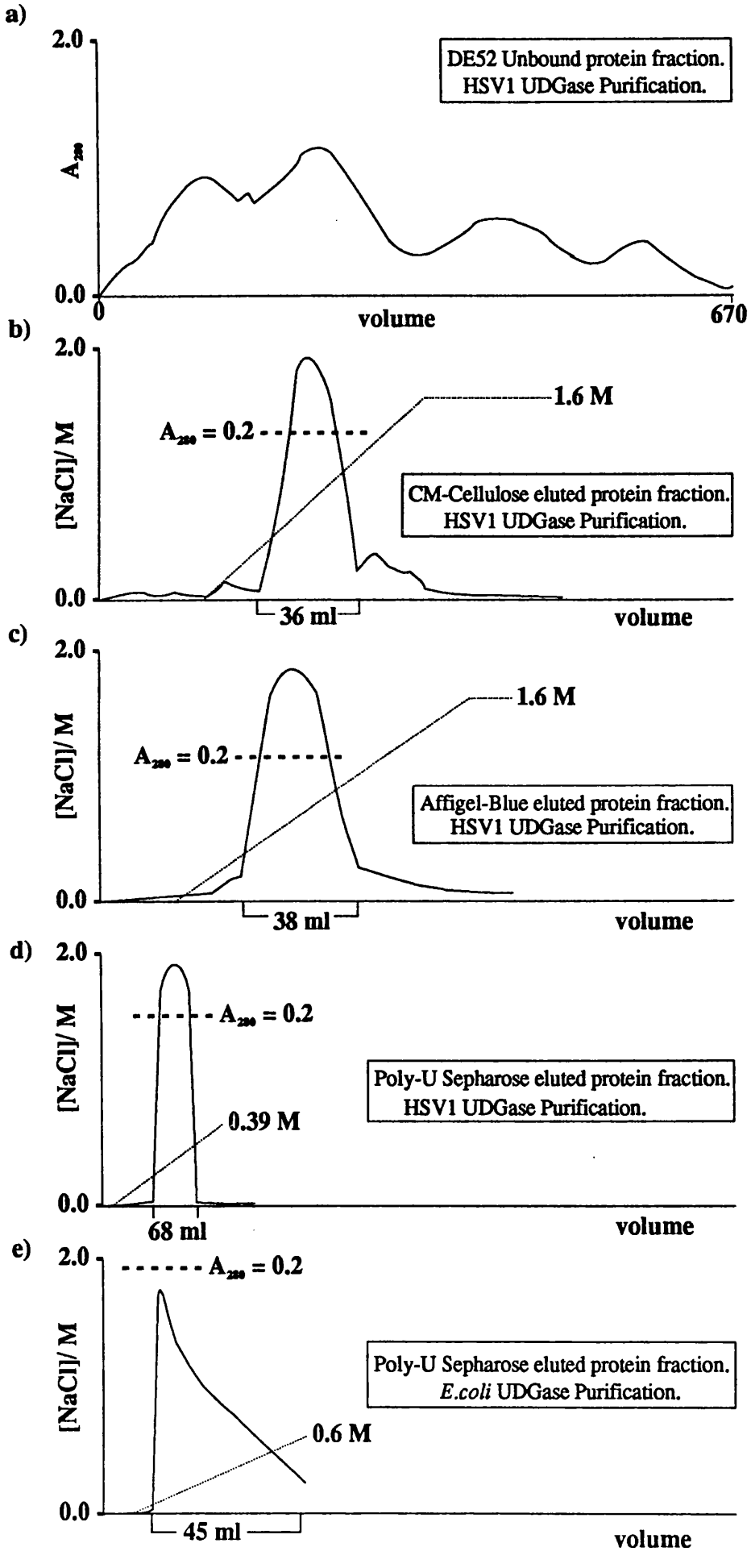


Figure 4.3.2 d
Chromatograms obtained during the purification of HSV1 UDGase (Section 4.3.2) and *E.coli* UDGase (Section 4.3.3).
Dotted lines indicate NaCl gradients (·····).
Dashed lines indicate the full scale deflection of the chart recorder at the detector A_{280} shown (- - - - -).

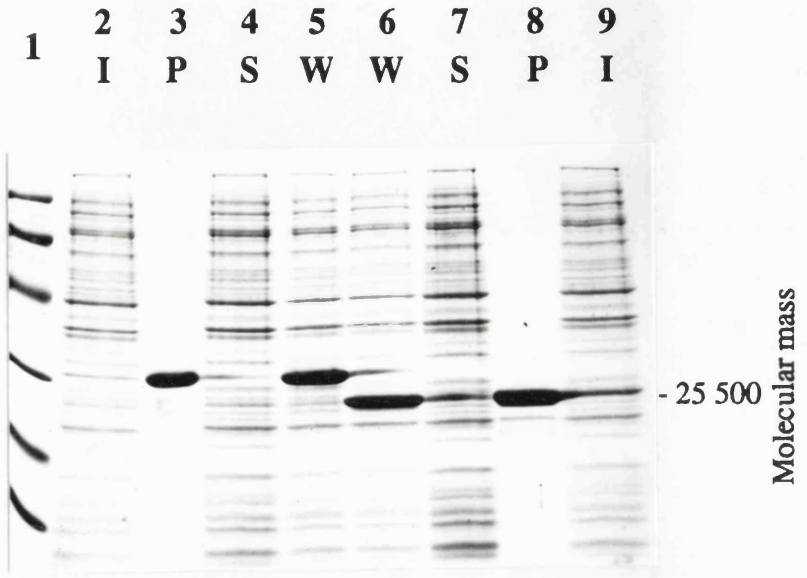


Figure 4.3.3

12 % SDS polyacrylamide gel showing that the new protein expressed in *E.coli* cells containing pEU234.2, is insoluble.

Lane **1** shows the molecular mass markers.

Lanes **2** to **5** show protein expressed from IPTG induced *E.coli* containining another construct, pBU107.1, designed to express the *E.coli* UDGase as a fusion protein for IMAC purification (Results not discussed).

Lanes **6** to **9** show the protein expressed from IPTG induced *E.coli* cells containing pEU234.2.

Codes: **I**= Whole cells before addition of 1 mM IPTG

W= whole cells 4 hours after induction with 1 mM IPTG

S= Soluble cell fraction.

P=Insoluble (pellet) cell fraction.

IPTG at a lower growth temperature of 28°C, which has sometimes resulted in at least part of the total amount of an overexpressed protein product being soluble. However, this again resulted in all the visibly overexpressed protein on a Coomassie Blue stained SDS-PAGE gel being present in the insoluble fraction. The next attempt was a titration of IPTG at 37°C, with three inductions carried out using 1 mM, 0.5 mM, and 0.1 mM IPTG respectively. Once again however, all three samples showed that all the overexpressed protein was in the pellet fraction. This titration could be repeated at 28°C, but this, going on the results of the previous attempts just discussed, at both 37 and 28°C, would probably result in a very minimal amount of soluble protein. There was a better chance of getting large amounts of soluble protein by using a protein solubilisation/ refolding protocol. There has been a report that the activity of *E.coli* UDGase can be successfully recovered after electrophoresis using 8 M urea and subsequent heat treatment (Bennett and Mosbaugh, 1992). It was therefore decided to attempt to resolubilise the protein.

The first attempt involved lysing the cells in buffers of different pH. However, the pH of 7.8 which was originally chosen, was not the cause of the insolubility of the protein, as buffers of pH 6.8 and 8.25 did not result in soluble protein. Next, cells were lysed in the original pH 7.8 buffer (Section **2.3.5**), and the supernatant was discarded following cell lysis and partition of fractions by centrifugation. The pellet was then washed with distilled water to remove any remaining supernatant. The water was removed and the pellet was resuspended by pipetting and vortexing in 30 ml of a wash buffer that contained 2 M sodium chloride (Section **2.3.6**). The pellet was agitated for 30 minutes at room temperature in this buffer, and the tube containing this resuspension was then centrifuged at >12 000 *g* for 30 minutes at 4°C. A Coomassie Blue stained SDS-PAGE gel showed that a number of proteins were re-solubilised by this step, but the *E.coli* UDGase was still in the pellet fraction.

The pellet was then resuspended, washed and recentrifuged twice in 30 ml of the original cell lysis buffer to remove any remaining sodium chloride and re-solubilised proteins. The pellet was then resuspended in 75 ml of de-ionised 8 M urea (Section **2.3.6**) and agitated at room temperature for 4 hours.

The resuspension was then centrifuged at >12 000 *g* for 30 minutes at 4°C, and the pellets discarded. The supernatant was then dialysed as three separate fractions, three times against 200 volumes of the re-folding buffer (Section **2.3.6**), for a minimum of 6 hours per dialysis. There was no precipitate in the dialysis bag at the end of this

procedure. A Coomassie Blue stained SDS-PAGE gel showed that the band corresponding to the size of the *E.coli* UDGase had indeed been re-solubilised along with another 2 bands of higher molecular mass. However, when assayed for UDGase activity, the still dilute fraction appeared to contain none. Thus much, if not all, of the re-solubilised protein was likely to be wrongly folded.

It was suspected that the sample may have been too concentrated with respect to the chances of intermolecular collision during re-folding, thus it was decided to attempt a different approach to the refolding step. The pellet was prepared for resolubilisation in 8 M de-ionised urea as detailed above, but instead of dialysis, the 75 ml of resolubilised protein was slowly added, with rapid mixing, to 5 litres of the re-folding buffer and left stirring at 4°C for 15 hours. It was determined that the best method of concentration of active protein would be to load it onto an affinity column. Thus, any wrongly folded protein and other unrelated proteins should pass through in the unbound fraction. The protein was loaded onto the polyuridylic acid sepharose 4B column described in sections **2.3.6** and **4.3.2**. This column was prepared using the refolding buffer to achieve a stable base-line. The sample was loaded on as a 1 to 1 mixture of refolding buffer and refolded sample. The column was then washed extensively with refolding buffer, before elution was carried out against an increasing gradient of refolding buffer containing 2 M sodium chloride.

The elution resulted in a single peak which trailed off slowly (Figure **4.3.2 d**), as opposed to the sharp descent observed in the case of the HSV1 UDGase. The fractions containing the peak volume were then pooled and dialysed twice against 200 volumes of the following buffers. The first dialysis was against 10 mM Tris·HCl pH 7.8, 5% glycerol, 1.5 mM DTT, 0.1 mM PMSF. The second dialysis was against 5 mM Tris·HCl pH 7.8, 1.5 mM DTT, 0.1 mM PMSF. The yield of soluble protein was approximately 15 mg on two separate occasions that this protocol was carried out. The protein was then concentrated to 18 mg/ ml using pressure ultrafiltration in a stirred cell and centrifugation in a microconcentrator as described in sections **2.3.7** and **5.1.1**. Some protein precipitated out during this process. However the yield of soluble protein at 18 mg/ ml was still 8.5 mg. Sodium azide was then added to a final concentration of 0.02 % as a preservative. The protein in this concentrated stock was found to exhibit specific UDGase activity. Crystallisation of the protein was attempted; this is discussed in section **5.1.2**.

5 CRYSTALLISATION OF HSV1 URACIL-DNA GLYCOSYLASE

The three dimensional structure of a complex macromolecule such as a protein can be determined by X-ray crystallography. The start point for this technique is to determine the conditions under which the most favoured physical state for the protein is crystalline, as opposed to freely soluble or amorphously precipitated. In the crystalline state each molecule forms part of a highly ordered, regularly repeating lattice which can be used to reveal the structure of the constituent molecules (Section **3.1.1**). If conditions that induce a crystalline state are located for a protein, there is no guarantee that this crystalline material will be suitable for X-ray diffraction studies. Thus much optimisation of the initially successful crystallisation conditions may be required. In some cases it is necessary to search for a totally new crystal form. In this chapter the search for, and optimisation of conditions which produce X-ray diffraction quality crystals of the HSV1 UDGase are presented. In addition, further studies on the crystals thus obtained are presented.

5.1 PROTEIN SOLUBILITY AND PHASE TRANSITION

A protein can exist in free solution in a number of dynamic states. In the case of a monomeric protein such as UDGase, it is expected that each molecule is discrete and able to assume a functional conformation in an environment where there is optimal activity in the presence of a substrate. In environments where sub-optimal or negligible activity is observed, the conformation of the enzyme and/ or the charge distribution is no longer suitable to enable a substrate to be turned over. The conformation and charge distribution can be affected to such an extent that it is no longer energetically favourable for the enzyme to remain in solution, and so precipitation occurs.

Precipitation can be reversible or irreversible under the conditions in which it occurs. If the charge distribution is slightly modified a new level of interaction can occur, namely aggregation of monomers; this aggregation can be reversible. If the environment and/ or charge distribution becomes so unfavourable that secondary structure is lost or modified in regions of the molecule, it becomes increasingly unlikely that reversal of the precipitation will occur.

The main environmental factors affecting the charge distribution on the surface of a protein are the pH of the environment (Yang and Honig, 1993), the temperature, the concentration and type of anions and cations present in solution and the local amino acid residues around any surface residues in the protein itself. The charge on the surface of the protein is due to the fact that the acidic and basic amino acids have ionisable groups which have particular pK values. The degree of ionisation is somewhat variable depending on the environmental factors listed, but nonetheless the charge of an ionisable surface residue will be negative at pH values less than its pK and positive above it. Every protein has an isoelectric point where the net charge is neutral; at this point the net positive charge in the molecule is equal to the net negative charge. This point is often, but not always, a point of minimum solubility for the protein molecule. Many more points of minimum solubility potentially exist for a protein molecule and these points, and the isoelectric point, can be exploited for the purposes of crystallising the protein in question.

At a point of minimum solubility the soluble phase as a physical state is often energetically unfavourable or very nearly so, and the alternative physical state is for the protein to precipitate. There exists however, an alternative metastable phase where the precipitating protein is able to aggregate in an ordered manner, namely as a crystal. This crystalline state is very narrow in terms of the area covered in a plot of protein concentration versus precipitating factor (Riès-Kautt and Ducruix, 1992) (Figure **2.4.1**), and can often be enhanced or stabilised by the presence of a variety of different types of molecule that have the potential to interact with the surface of the protein molecule (Riès-Kautt and Ducruix, 1991). Similarly, the crystalline state is very easily disrupted by such potentially interactive molecules. These interactive molecules can be salts and other inorganics, organics, small monomers or large polymers, detergents or zwitterions. The solubility of the protein in certain conditions can be enhanced or reduced and this has the knock on effect of enlarging or decreasing the metastable zone as a result. The types of molecule that can influence protein crystallisation and the means of their interaction are discussed in section **2.4.1.**

5.1.1 A 'SPARSE MATRIX' SURVEY USED TO LOCATE CRYSTALLISATION CONDITIONS FOR THE URACIL-DNA GLYCOSYLASE OF HSV1

The 'sparse matrix sampling' method for screening crystallisation conditions of

proteins (Jancarik and Kim, 1991) is based on the same principles as the incomplete factorial approach, namely that a very large range of factors can be explored with minimal use of protein. As well as sampling a large area of multi-factor space with the minimum of confounding, the 'sparse matrix' is heavily biased toward known or published mixtures that have proved successful in crystallising proteins in the past. The number of conditions is kept to fifty by deleting mixtures that could be at least partially represented by another earlier in the grid. The resulting set of mixtures proved to be very efficient in obtaining at least one crystal form from fifteen previously crystallisable proteins. Since that time, many proteins have been crystallised from mixtures identical to or deriving from the mixtures in the 'sparse matrix'. A number of variations on the original are now in common use, and it is possible to design similar biased incomplete factorial grids.

A 'sparse matrix' survey was carried out when the HSV1 UDGase had been purified to a reasonable level (Section **4.3.2**), that is to greater than 85 % of total protein following elution from the Affigel Blue column. The concentration of protein was measured by reading the absorbance at 280 nm, the aim being to get the sample to supersaturation. To start with, a concentration greater than 15 mg/ ml was desired. The protein was concentrated first by pressure ultrafiltration in a stirred cell until the volume was less than 2 ml. At this point the sample was loaded into a Centricon SR3 microconcentrator. The stirred cell was then washed twice with 0.4 ml purification buffer and both washes were also loaded into the microconcentrator. The concentration was continued by centrifuging the microconcentrator at 7350 *g* until the volume was such that the predicted concentration would be greater than 15 mg/ ml. The sample was recovered by washing the sample up and down over the membrane in the microconcentrator with an air displacement pipette, attaching a cup to the top of the assembly and inverting the entire unit, and centrifuging for one minute at 1 500 *g*. The sample was 0.22 µm filtered and the absorbance at 280 nm was taken, which corresponded to about 19.6 mg/ ml.

The survey was carried out manually as a microbatch procedure under paraffin oil using a 72 well Terazaki plate. The droplets were mixed at a 1:1 v/v ratio of protein (19.6 mg/ ml) to precipitant mixture, with 1µl of each being mixed with an air displacement pipette on a piece of Nescofilm™ prior to being dispensed under the oil into a well in the plate. The droplets were checked under a dark-field microscope after each row of six had been dispensed. This was on average about 10 minutes per row. By the time row five had

been dispensed, the droplet numbered 15, in row 3 well C had produced a shower of fine needles. This is a crystalline form, though at that stage it was not possible to determine whether the crystals were protein or salt in origin. The precipitant in this droplet was from a stock of 30 % PEG 10 000, 0.1 M sodium cacodylate pH 6.5 and 0.2 M ammonium sulphate. Only 3 out of the 50 droplets failed to produce a precipitate within 2 hours. The precipitates varied in thickness and form, and within 2 days a total of 5 of the droplets had produced crystalline forms. The results of this survey are presented in figure **5.1.1 a.** The five successful conditions have since been repeated with a purer protein sample at 25 mg/ ml. The types of crystal form given by each precipitant are shown in figure **5.1.1 b.**

5.1.2 RESULTS OF A 'SPARSE MATRIX' SURVEY CARRIED OUT USING THE REFOLDED *E.COLI* URACIL-DNA GLYCOSYLASE

The insoluble, recombinant *E.coli* UDGase was purified and refolded as described in section **4.3.3**, and a 'sparse matrix' survey was carried out with the protein stock concentration at 18 mg/ ml. At first, it appeared that very small needles had formed within a few minutes in the very first droplet, corresponding to a precipitant of 0.02 M calcium chloride, 0.1 M sodium acetate pH 4.6, 30 % v/v 2-methyl-2,4-pentanediol. This condition was repeated subsequently, but no crystalline material was observed in these droplets. All droplets produced a fine, fibrous precipitate which did not differ in appearance from one condition to the next. This tends to suggest that the re-folded preparation, although soluble and containing activity, was probably heterogeneous in terms of the protein homogeneity required for crystallisation.

5.2 OPTIMISATION OF THE SUCCESSFUL CRYSTALLISATION CONDITIONS

5.2.1 LOCATING THE OPTIMUM PRECIPITANT STRENGTH AND pH

The first thing that was done following the needle formation in droplet number 15, was to set up duplicates of the droplet, and to vary the relative volume of protein and precipitant at three different protein concentrations, 19.6 mg/ ml, 15 mg/ ml, and 9.5 mg/ ml. These were observed 30 minutes later and two hours later. The results of this are

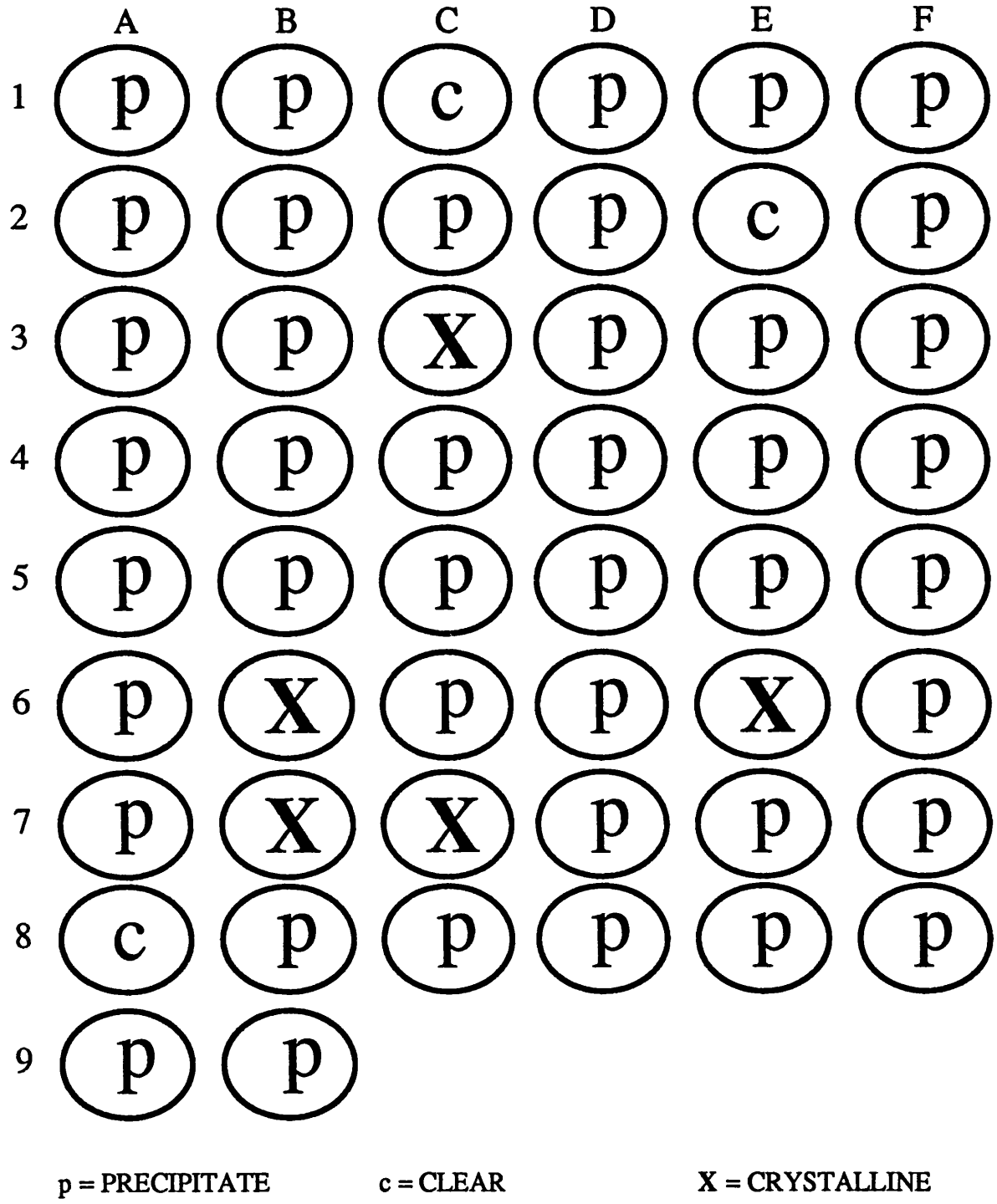


Figure 5.1.1 a

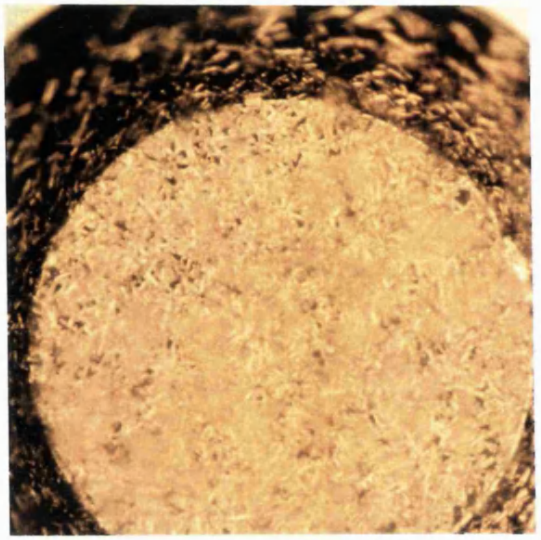
Representation of the results of the initial sparse-matrix survey carried out using the herpes simplex virus type1 uracil-DNA glycosylase at 19.5 mg/ml. All drops were a 1:1 mixture of of protein and one of fifty precipitant solutions. The precipitant mixtures were added such that the first was in row 1/ column A, and the fiftieth was in row 9/ column B. The results are as seen after a week at a constant temperature of 16 celsius.

	SALT	BUFFER	PRECIPITANT
1	0.02 M $CaCl_2$	0.1 M Na·acetate pH 4.6	30 % v/v MPD
2	None	None	0.4 M K Na·tartrate
3	None	None	0.4 M NH_4·phosphate
4	None	0.1 M Tris·HCl pH 8.5	2.0 M NH_4·phosphate
5	0.1 M Na·citrate	0.1 M Na·HEPES pH 7.5	30 % v/v MPD
6	0.2 M $MgCl_2$	0.1 M Tris·HCl pH 8.5	30 % w/v PEG 4000
7	None	0.1 M Na·cacodylate pH 6.5	1.4 M Na·acetate
8	0.2 M Na·citrate	0.1 M Na·cacodylate pH 6.5	30 % v/v propan-2-ol
9	0.2 M NH_4 acetate	0.1 M Na·citrate pH 5.6	30 % w/v PEG 4000
10	0.2 M NH_4 acetate	0.1 M Na·acetate pH 4.6	30 % w/v PEG 4000
11	None	0.1 M Na·citrate pH 5.6	1.0 M NH_4·phosphate
12	0.2 M $MgCl_2$	0.1 M Na·HEPES pH 7.5	30 % v/v propan-2-ol
13	0.2 M Na·citrate	0.1 M Tris·HCl pH 8.5	30 % w/v PEG 400
14	0.2 M $CaCl_2$	0.1 M Na·HEPES pH 7.5	20 % w/v PEG 400
15	0.2 M $(NH_4)_2SO_4$	0.1 M Na·cacodylate pH 6.5	30 % w/v PEG 8000
16	None	0.1 M Na·HEPES pH 7.5	1.5 M Li_2SO_4
17	0.2 M Li_2SO_4	0.1 M Tris·HCl pH 8.5	30 % w/v PEG 4000
18	0.2 M Mg·acetate	0.1 M Na·cacodylate pH 6.5	20 % w/v PEG 8000
19	0.2 M NH_4 acetate	0.1 M Tris·HCl pH 8.5	30 % v/v propan-2-ol
20	0.2 M $(NH_4)_2SO_4$	0.1 M Na·acetate pH 4.6	25 % w/v PEG 4000
21	0.2 M Mg·acetate	0.1 M Na·cacodylate pH 6.5	30 % v/v MPD
22	0.2 M Na·acetate	0.1 M Tris·HCl pH 8.5	30 % w/v PEG 4000
23	0.2 M $MgCl_2$	0.1 M Na·HEPES pH 7.5	30 % w/v PEG 400
24	0.2 M $CaCl_2$	0.1 M Na·acetate pH 4.6	20 % v/v propan-2-ol
25	None	0.1 M imidazole pH6.5	1.0 M Na·acetate
26	0.2 M NH_4 acetate	0.1 M Na·citrate pH 5.6	30 % v/v MPD
27	0.2 M Na·citrate	0.1 M Na·HEPES pH 7.5	20 % v/v propan-2-ol
28	0.2 M Na·acetate	0.1 M Na·cacodylate pH 6.5	30 % w/v PEG 8000
29	None	0.1 M Na·HEPES pH 7.5	0.8 M K Na·tartrate
30	0.2 M $(NH_4)_2SO_4$	None	30 % w/v PEG 8000
31	0.2 M $(NH_4)_2SO_4$	None	30 % w/v PEG 4000
32	None	None	2.0 M $(NH_4)_2SO_4$
33	None	None	4.0 M Na·formate
34	None	0.1 M Na·acetate pH 4.6	2.0 M Na·formate
35	None	0.1 M Na·HEPES pH 7.5	1.6 M K Na·tartrate
36	None	0.1 M Tris·HCl pH 8.5	8 % w/v PEG 8000
37	None	0.1 M Na·acetate pH 4.6	8 % w/v PEG 4000
38	None	0.1 M Na·HEPES pH 7.5	1.4 M Na·citrate
39	None	0.1 M Na·HEPES pH 7.5	2 % v/v PEG 400 + 2.0 M $(NH_4)_2SO_4$
40	None	0.1 M Na·citrate pH 5.6	20 % w/v PEG 4000 + 20 % v/v propan-2-ol
41	None	0.1 M Na·HEPES pH 7.5	20 % w/v PEG 4000 + 10 % v/v propan-2-ol
42	0.05 M K·phosphate	None	20 % w/v PEG 8000
43	None	None	30 % v/v PEG 1500
44	None	None	0.2 M Mg·formate
45	0.2 M Zn·acetate	0.1 M Na·cacodylate pH 6.5	18 % w/v PEG 8000
46	0.2 M Ca·acetate	0.1 M Na·cacodylate pH 6.5	18 % w/v PEG 8000
47	None	0.1 M Na·acetate pH 4.6	2.0 M $(NH_4)_2SO_4$
48	None	0.1 M Tris·HCl pH 8.5	2.0 M $(NH_4)_2SO_4$
49	1.0 M Li_2SO_4	None	2 % w/v PEG 8000
50	0.5 M Li_2SO_4	None	15 % w/v PEG 8000

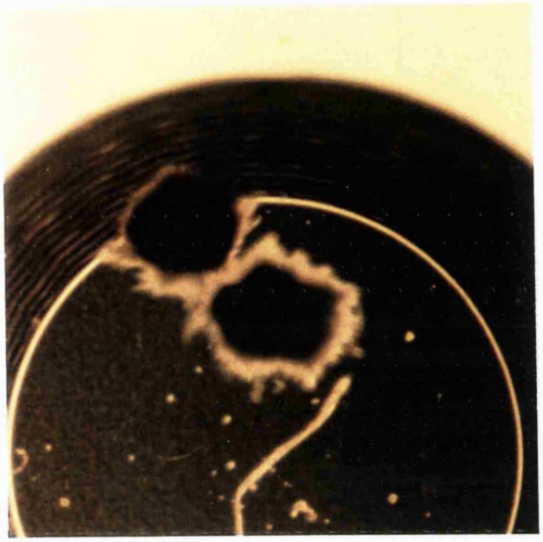
Figure 5.1.1 a (Continued . .)

Precipitant/ salt/ buffer mixtures used at 1:1 ratio with protein in a sparse matrix survey.

i)



ii)



iii)

iv)

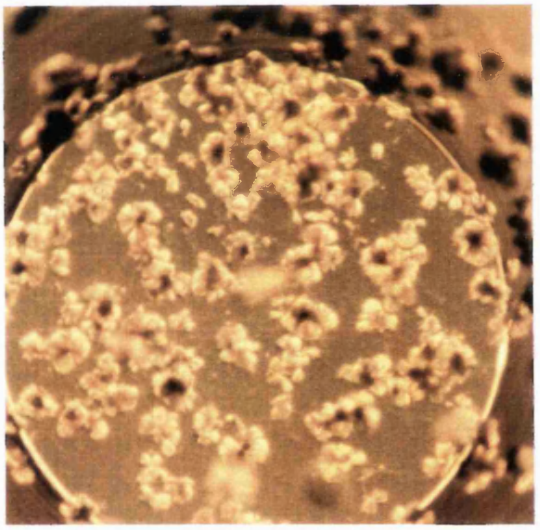


Figure 5.1.1b

Crystalline forms of HSV1 UDGase, as seen 2 days after a repeat of the initial Sparse-Matrix Survey, using 25 mg/ml protein stock.

The different crystal forms were obtained with:

i) **Condition 15 (Well 3C)**; hundreds of tiny needles, approx. 0.01 x 0.05 mm (w x l) each.

ii) **Condition 32 (Well 6B)** - Not Shown, but resembling many smaller examples of the type shown here, which were formed with - **Condition 39 (Well 7C)**; One large baton with large, sheaf-like ends.

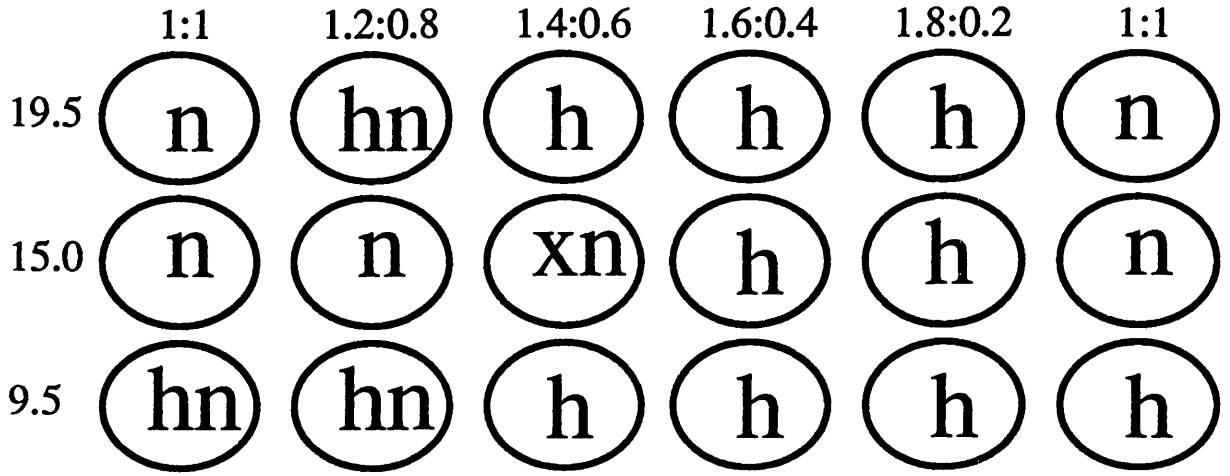
iii) **Condition 35 (Well 6E)**; Many clusters of thin plates, approx. 0.05 x 0.2 mm (w x l), each (less than 0.005mm thick).

iv) **Condition 38 (Well 7B)**; Many clusters of tiny thin plates (about 1/5 the size of those in condition 35).

presented in figure **5.2.1 a**. Figure **5.2.1 b** shows the results of the first gradient search carried out to test pH with different average molecular weight PEGs. The results of these searches were that a relative volume ratio of between 1.2 and 1.4 of protein to 0.8 and 0.6 of precipitant gave the thickest needles. Also, the pH of 6.2 to 7.4 and PEG 4 000 to 8 000 seemed to give the best needles at this ratio in terms of length and face width (200µm x 20µm). The largest were found mainly between pH 6.6 and 7.2. The two highest protein concentrations gave the best crystals (Figure **5.2.5**).

The next tests were done to investigate whether the type of buffer used would affect the size or quality of the crystals in terms of external morphology. The first grid that was set up varied the percentage of PEG 8 000 on one axis (30, 20, 10), and the pH on the other for two buffers: sodium cacodylate (6.4, 6.6, 7.0) and HEPES·NaOH (6.8, 7.1, 7.5), while keeping the ammonium sulphate concentration at 0.2 M. The droplets were set up in both 1:1 and 1.3:0.7 relative volume ratios of protein to precipitant. The results showed that the PEG 8 000 at 30 % in the stock precipitant provided nucleation conditions, whereas it did not at 20 % and 10 %. The 1.3:0.7 protein:precipitant ratio gave triangular batons that had dimensions of 60 µm wide by 200 µm long at pH 6.6 with sodium cacodylate buffer. The HEPES·NaOH at pH 6.8 and 7.1 gave very large needles over 0.5 mm in length and up to 50 µm wide but these were highly branched. In fact, the general observation with the sulphonic acid buffers (MES·NaOH, PIPES·NaOH, and HEPES·NaOH) in this pH region of 6.8-7.0 is that they tend to induce twinning effects in the crystals. Needles were observed in all the 1:1 ratio experiments between pH 6.4 and 7.1. The next experiment involved assessing another buffer, sodium phosphate at pH 6.6, 6.8, 7.0 and 7.2, against HEPES·NaOH at pH 6.8 and 7.1, and sodium cacodylate at pH 6.6. The effect of sodium phosphate buffer was to produce large batons in the 1.3:0.7 protein to precipitant ratio droplets at pH 6.8 and 7.0 that were of similar dimensions to the HEPES·NaOH crystals, namely 50 to 70µm wide by up to 0.5 mm in length. These crystals, unlike those grown in the presence of HEPES·NaOH were not branched, and the external appearance was that of hexagonal batons as opposed to the trigonal batons obtained in the presence of sodium cacodylate.

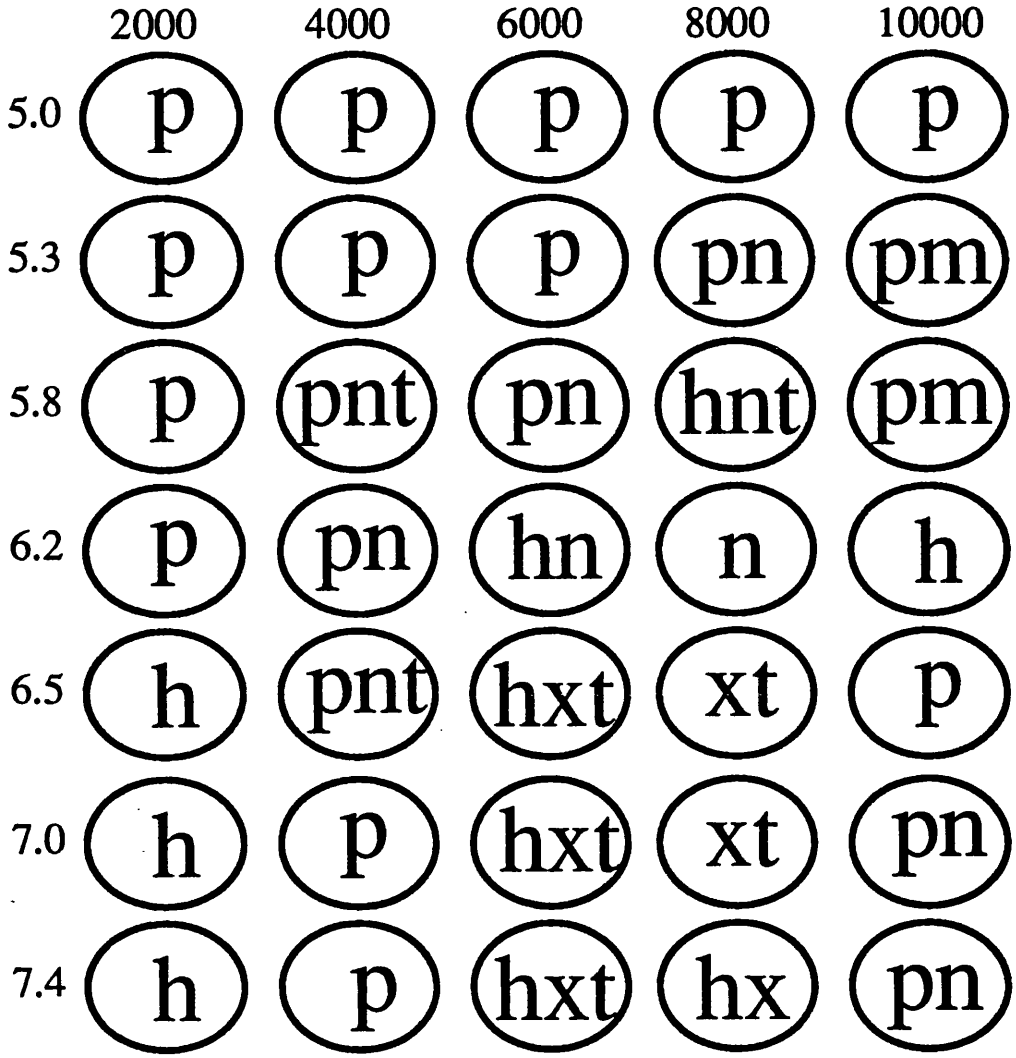
The size of the crystals was enhanced when the additional step of a polyuridylic acid sepharose column was included in the protein purification, and they typically reached 90 to 100 µm wide by over half a millimetre long. Increasing the protein stock concentration to 25 mg/ ml also resulted in increased dimensions. Seeding of the crystals



h= hazy or cloudy n= needles (<30 microns wide) x= crystals (>30 microns wide)

Figure 5.2.1a

Representation of the results of the first experiment to follow the sparse-matrix survey, to assess the effect of varying both the initial concentration of protein, and the relative protein:precipitant concentration. The figures along the top of the diagram refer to the relative volumes in microlitres, in the form of protein:precipitant. The figures down the side refer to the concentration of protein prior to mixing, in mg/ml. The results are as seen after two hours at a constant temperature of 16 celsius.



p= precipitate h= hazy or cloudy t= twinned
x= crystals (>30 microns wide) n= needles (<30 microns wide)
m= microcrystalline appearance

Figure 5.2.1b

Representation of the results of the test carried out to observe the effect of pH and average molecular mass of polyethylene glycol on the crystallisation of herpes simplex virus type 1 uracil-DNA glycosylase. The concentration of ammonium sulphate in the precipitants was constant at 0.2M. The numbers along the top of the diagram represent the average molecular mass of polyethylene glycol in the precipitant mixture. The numbers down the side show the pH of the 0.1M sodium cacodylate in the precipitant. The drops were 1:1 mixtures of protein and precipitant. The results are as seen after a week at a constant temperature of 16 celsius.

improved the ultimate size still further; this is discussed further in section **5.3.2.** The crystals were of diffraction quality, with an 80 µm wide crystal diffracting to better than 3.0Å with a CuKα radiation/ MAR image plate system, which is discussed further in section **3.2.3.**

5.2.2 THE EFFECT OF GLYCEROL AS AN ADDITIVE

The effect of glycerol as an additive was investigated at 10 % and 20 % v/v in the precipitant stock using the buffers discussed in the previous section, with PEG 8 000 at 30 % w/v and ammonium sulphate unchanged at 0.2 M. This was done to see whether the presence of glycerol in the protein stock at 15 % was likely to be affecting the crystallisation in any way. The results showed needle formation at all 1:1 ratios of protein to precipitant without added glycerol. With 10 % added glycerol, the number of nuclei was limited to less than 20, as opposed to a shower. With 20 % added glycerol, nucleation was not present in some drops and only one or two nuclei were present in others. Despite this, the size of the crystals was not enhanced. A sample of protein which had been dialysed twice against 200 volumes of glycerol free buffer and re-concentrated showed no noticeable difference in the crystals that were formed to crystals that are produced from the usual 15 % glycerol containing protein stock.

5.2.3 THE EFFECT OF COUNTER-IONS ON CRYSTAL GROWTH AND STABILITY

A final test was to gauge the effect of ammonium sulphate. Droplets were set up using the 1.3:0.7 protein to precipitant ratio, and precipitant consisting of 30 % w/v PEG 8 000, 0.1 M sodium phosphate buffer at pH 6.8, with or without 0.2 M ammonium sulphate. It was found that, in the absence of ammonium sulphate, only a precipitate would form. A further set of experiments was done, varying the level of ammonium sulphate from 0.05 M in the stock precipitant to 0.6 M, everything else remaining constant. It was found that crystals were absent or very small at 0.05 M, smaller than average at 0.1 M, and progressively smaller and more branched at concentrations higher than 0.3 M.

To test whether either the ammonium ions or sulphate ions were dispensable, the

precipitant was altered to include 0.2 M ammonium chloride instead of ammonium sulphate. There were no crystals, though the ammonium chloride gave discrete sphere-like masses of loose precipitate. The precipitant was again altered to include 0.2 M sulphates of lithium, sodium, magnesium, and copper, instead of ammonium sulphate. In the case of copper sulphate there was nothing but heavy precipitate, but in all other cases there were crystals. In the case of magnesium sulphate, background precipitate levels were comparable to those obtained in the presence of ammonium sulphate. However, in the case of lithium and sodium sulphates, there was far less background precipitate, and if any was present at all, it was more powdery than the rather fibrous precipitate that formed in the ammonium sulphate containing droplets. The lithium sulphate was somewhat less soluble in the precipitant stock than sodium sulphate, so the sodium sulphate was favoured over the others.

To further test the importance of sulphate ions in the crystallisation of the HSV1 UDGase, some more experiments were conducted. The first involved transferring a crystal into the standard stabilising solution for the crystals, 0.2 M sodium sulphate, 12 % w/v PEG 8 000, 0.05 M HEPES·NaOH pH 6.8. The same procedure was carried out for a second crystal, except that the stabilising solution contained no sodium sulphate. After about 30 minutes, the crystals were observed under the microscope using a cross-polariser. Though the external appearance of the crystals was no different under normal light, it was clear to see that there was no longer any birefringence from the crystal that was in buffer without sulphate. When the crystal was manipulated with an animal hair, it deformed readily and had become a gel. An equal volume of sodium sulphate containing stabilising solution was added to this droplet to bring the final sulphate concentration to 0.1 M. After 30 minutes, the crystal was manipulated with an animal hair again, and was found to have re-set, though the birefringence was no longer present and it was, strictly speaking, a glass. Thus it is concluded that the sulphate ions play a key role in stabilising the crystal lattice.

The next experiment involved observing whether a precipitant containing no other anions but sulphate would be sufficient to crystallise the protein. The buffer in the precipitant was brought to the correct pH of 6.8 by using concentrated sulphuric acid. The buffer consisted of 0.2 M imidazole pH 6.8, with a concentration of sulphate ions approximately equal to 94 mM, and 30 % w/v PEG 8 000. This precipitant was successful in producing long, narrow crystals of trigonal morphology. Similar precipitants which

contained no sulphate ions, but which were produced by adjusting the pH of the imidazole buffer to 6.8 with nitric acid, orthophosphoric acid, or hydrochloric acid failed to induce crystallisation of the protein.

5.2.4 THE EFFECT OF TEMPERATURE ON CRYSTAL GROWTH

The crystallisation is normally carried out in a constant temperature cabinet set to 18°C, apart from setting up the droplets and observing them which is done at room temperature. The crystallisation was repeated at 4°C and 30°C. The crystallisation at 4°C has the effect of increasing the number of nuclei and thus limiting the size of the crystals. The amount of background precipitate is slightly diminished on average, though the skin on the surface of the droplet is still present. At 30°C, the crystallisation produces the normal number of nuclei on average and the crystals are of typical dimensions observed at 18°C, but nucleation does not occur at all sometimes. In addition, the background precipitate is increased from the normal levels seen at 18°C, and the protein skin is still present. It is seen that the level of background precipitate begins to increase even under the heat of the illumination of the microscope stage, and so exposure to this must be minimised during the growth stages.

5.2.5 THE EFFECT OF PROTEIN CONCENTRATION ON CRYSTAL GROWTH

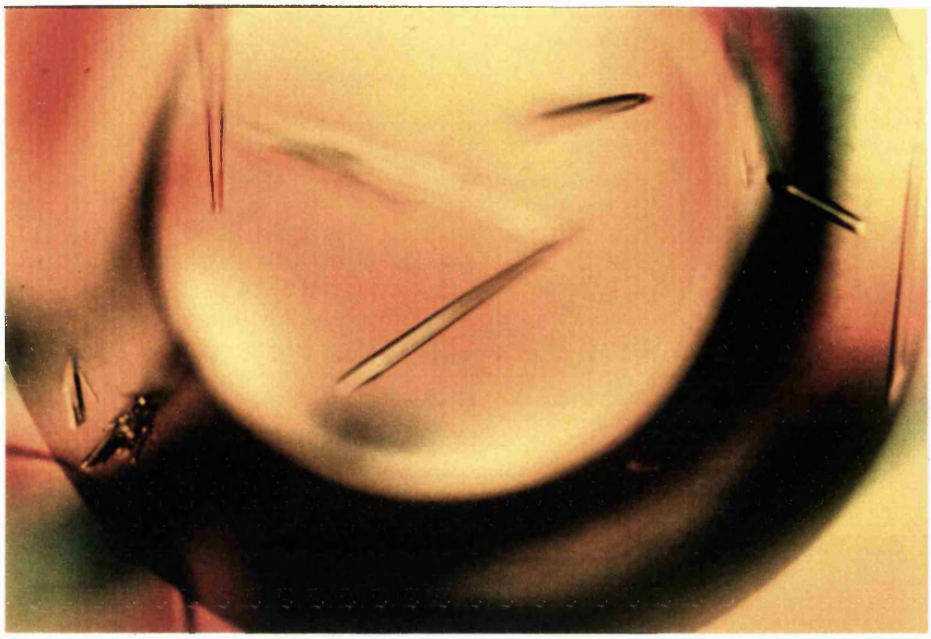
It was found that when the initial refinements were being done on the successful 'sparse-matrix' condition number 15, the protein concentration could be either 15 mg/ ml or 19.6 mg/ ml and the crystallisation would look virtually identical. It was also found that 9.5 mg/ ml was not enough to produce sizeable crystals, though nucleation could occur. When purer supplies of the enzyme were prepared and crystallisation parameters were optimised, it was noted that crystal size could be improved still further by increasing the protein stock concentration to 25 mg/ ml. This seemed to be the highest concentration the protein could tolerate however, because the protein would begin precipitating only minutes after being brought to this concentration. It could be resuspended, but this became less successful with the age of the sample. It was noted however, that a six month old sample which lacked glycerol was still in solution though it did not crystallise well any more. This prompted the modification to the last dialysis stages of the purification, to reduce the

glycerol gradually to nothing, and reduce the general ionic strength and include preservatives and protease inhibitors. With a batch of protein so produced, the protein was found to remain soluble at 65 mg/ ml for up to nine months with no decrease in the quality of crystals produced. The effect of such a supersaturated protein solution is that the protein crystallises at just 5.2 % PEG 8 000, 34 mM sodium sulphate, 17 mM sodium phosphate buffer pH 6.8 with protein at 54 mg/ ml. Previously, crystallisation would occur at 10.5 % PEG 8 000, 70 mM sodium sulphate, 35 mM sodium phosphate buffer pH 6.8 with protein at 12.5 mg/ ml.

The change in concentration has the effect of producing crystals that are much larger, and which have morphologies that are far clearer to see (Figure **5.2.5**). Typical sizes of crystals at the old protein concentrations were between 40 µm x 40 µm x 500 µm and 110 µm x 100 µm x 1 mm. At the new protein concentration the typical sizes are between 150 µm x 150 µm x 300 µm and 600 µm x 600 µm x 2 mm. In addition, the number of nuclei per droplet is seen to vary with the buffer used, as does the external appearance of the crystals. The precipitant containing sodium sulphate and sodium phosphate buffer tends to produce droplets containing 1 to 3 nuclei with crystals of definite trigonal form that can reach the upper limit of the sizes quoted. The crystals tend to grow on the walls of the well, though they can be seen to form in the central region of the droplet and sink to the bottom as they become larger. They are usually squat in appearance with a characteristic hollow at the growing end. The crystals take a variable time to form nuclei, from a couple of seconds to several days, but growth is extremely rapid once this commences with 50 % of the final size of the crystal laid down within 3 hours. Growth is generally complete within 12 hours of the first observation of a nucleus.

The precipitant containing imidazole sulphate produces long crystals, which are definite trigonal prisms that can reach up to two millimetres in length but rarely exceed 250µm in width. One or two nuclei per droplet is typical for these, and they grow either from the skin of the droplet or on the walls of the well. There is no hollowing with these crystals, rather a three-sided pyramidal point that protrudes from the growing end. It is of note that this point is also present during the initial growth stages of the phosphate, sodium sulphate grown crystals, but is caught up by the outer edges and overtaken by them during the latter stages of growth in that precipitant. A precipitant containing 0.1 M imidazole brought to pH 6.8 with hydrochloric acid, and sodium sulphate at 0.2 M produces between 15 and 30 nuclei per droplet with resulting sizes of about 60 to 90µm

i)



ii)

iii)

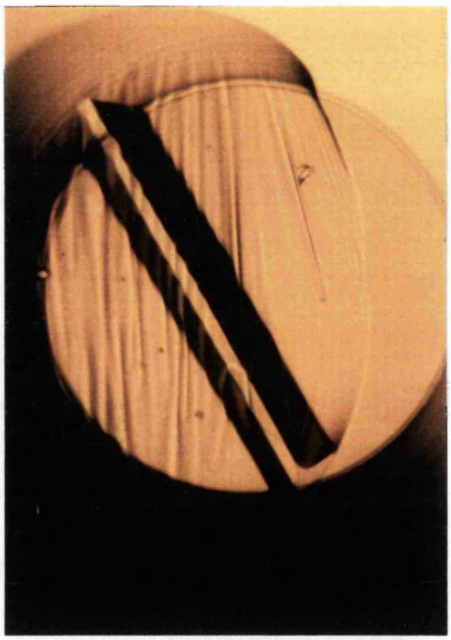


Figure 5.2.5

i) Initial improvement of the size of needles to approx. 0.04 x 0.2 mm (w x l).

ii) Typical HSV1 UDGase native crystal morphology obtained with sodium sulphate/ sodium phosphate buffer pH 6.8/ PEG 8000 precipitant.
The crystal (approx. 0.3 x 0.6 mm, w x l) is seen end on.

iii) Typical HSV1 UDGase native crystal morphology obtained with imidazole sulphate buffer pH 6.8/ PEG 8000 precipitant. The crystal (approx. 0.2 x 1.0 mm, w x l) is seen side on. The thick skin (Section **5.3.1**) which covers the droplets is clearly visible.

wide by about 200 µm long. The external appearance is trigonal rods with blunted corners and the pyramidal point at the growing end. These nucleate from the skin or the surface of the well. A similar effect is produced in precipitant containing a 0.1 M imidazole buffer brought to pH 6.8 with orthophosphoric acid, 0.2 M sodium sulphate. With a precipitant containing 0.1 M imidazole buffer brought to pH 6.8 with nitric acid, and 0.2 M sodium sulphate, crystals such as those produced with sodium sulphate/ sodium phosphate buffer are formed. Imidazole phosphate or imidazole nitrate are not able to induce or support crystals without the presence of sulphate ions (Section **5.2.3**).

A re-optimisation of the crystallisation parameters was done with the more concentrated protein stock, but apart from the weaker precipitant strength required, the pH and temperature requirements are the same. The crystals are still stable in the same stabilisation solution.

5.2.6 THE EFFECT OF ADDITIVES ON CRYSTALLISATION

Before the effect of increased protein concentration in the absence of glycerol was observed, an attempt was made to increase the width of the existing crystals by including various organic solvents in the crystallisation droplet (Derewenda *et al.*, 1989). It was hoped that one of these would act as a selective poison for the rapidly growing c-axis of the crystal, enabling the width to expand by accretion of the protein that would otherwise go onto the c-axis. This type of effect has been seen upon addition of non-ionic detergents (Walker *et al.*, 1990). Some non-ionic detergents were tested to see whether they would have an effect on the crystallisation. It was initially reported (McPherson *et al.*, 1986a; McPherson *et al.*, 1986b) that these act by limiting the number of nuclei and keeping the protein from forming non-specific aggregates as the crystallisation proceeds, thus increasing the ultimate size of any crystals. Table **5.2.6** shows the results of these tests. None of the organics or detergents had the required effect, though phenol appeared to modify the crystal size more than any of the others.

The results of sections **5.2.1** to **5.2.6** are summarised in figure **5.2**.

ADDITIVE name and final concentration	TYPICAL CRYSTAL SIZE/ µm	NUMBER OF NUCLEI	TYPE OF PRECIPITATE
No additive	90	4	v.light
1mM PHENOL	120	3	light
1mM QUINOL	50	3	heavy
1% v/v ETHANOL	80	4 twinned	v.light
1% v/v DICHLORO-METHANE	60	15 irregular edges	medium
1% v/v CHLOROFORM	60	12 irregular edges	medium
1% v/v TOLUENE	60	9 irregular edges	medium
1% v/v 2-METHYL-2, 4-PENTANDIOL	n/a	0	light
1% v/v ACETONE	80	3 twinned	medium
1% v/v 2-BUTANONE	90	9 twinned irregular edges	medium
1% v/v TETRACHLORO-METHANE	80	7 twinned irregular edges	medium
0.1% v/v TRITON X-100	n/a	0	heavy
0.3% v/v β-OCTYLGLUCO -PYRANOSIDE	n/a	0	light

Table 5.2.6

Effects of the addition of various organic solvents to the precipitant mixture, to give the final concentrations listed above in the mixed droplet. A heavy precipitate is one that appears as clumps, a very light one appears as a fine dust on the walls of the well.

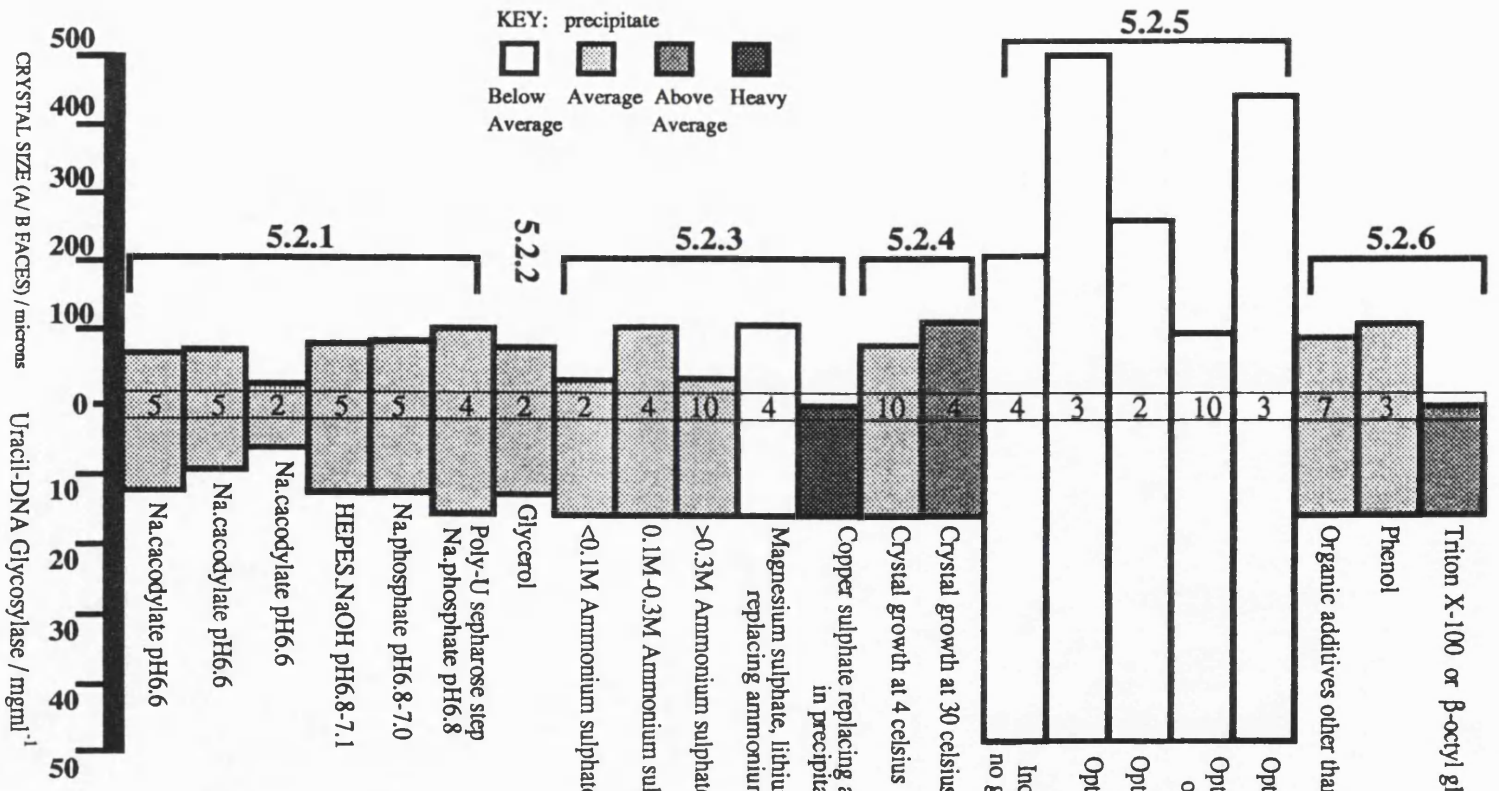


Figure 5.2

Summary of crystallisation results presented in sections **5.2.1** to **5.2.6**. The relevant section for any particular result is shown above the bar which depicts that result. An average precipitate is one which appears as a light dust on the walls of the well. Figures in the central bar represent typical numbers of nucleation points per drop.

5.3 TECHNIQUES USED IN THE PRODUCTION OF CRYSTALS OF URACIL-DNA GLYCOSYLASE

During the early stages of growing crystals of this enzyme, various techniques were attempted to try to optimise the reproducibility of crystal growth and size of the crystals. This section details the methods used.

5.3.1 THE MICROBATCH METHOD IS USED TO GENERATE CRYSTALS OF URACIL-DNA GLYCOSYLASE

The method of choice in generating crystals of this enzyme is the microbatch method, carried out under paraffin oil in 72 well Terazaki plates (Chayen *et al.*, 1992). The reason for staying with this method since the 'sparse matrix' was carried out is that the protein, with or without the presence of precipitant, tends to form a tough skin over any exposed surface (Figure **5.2.5**). In attempts to grow the crystals by vapour diffusion, the skin prevented the free exchange of water vapour between the reservoir and droplet, resulting in small nuclei that would begin to grow but then stop and disintegrate over a long period of time. Liquid/ liquid free interface diffusion in capillaries was also attempted, but this method always generated a large number of nuclei, thus limiting the useful size of the crystals obtained.

5.3.2 THE ENLARGEMENT OF CRYSTAL SIZE BY MACROSEEDING

Crystals of the HSV1 UDGase, described earlier in this chapter, were on the lower size limit of crystals useful for X-ray diffraction before the effect of reduced glycerol/ increased protein concentration was gauged. Thus, in order to improve the limit of diffraction of the crystals when using a laboratory CuKα rotating anode radiation source, the technique of macroseeding was used (Stura and Wilson, 1991).

Crystals could be removed from the droplets by first adding a small amount of stabilisation solution to the surface of the droplet. This would often result in the protein skin becoming ruptured. Even if it did not, however, the skin was easier to remove if there was liquid above it as well as below. The skin was removed using an animal hair. The droplet was then swollen to a volume permissible in the Terazaki plate well, about 20 to

30 µl. Crystals were nudged gently with an animal hair prior to aspiration to dislodge them from the surface of the well. This was done at the very point of contact of the crystals to the well because it is easy to break them due to their long, thin shape.

The crystals were transferred to a petri dish of about 3 ml capacity, filled with de-ionised water. After about 30 minutes, the crystals looked very etched at the surface, and any amorphous precipitate that was adhered to the surface had come away. The crystals were blown around with an air-displacement pipette to dislodge much of the loosely associated surface. The crystals were then transferred into stabilisation solution and left for 5 minutes. The crystals were then transferred to another droplet containing stabilisation solution. Each transfer step was carried out to minimise liquid carry over, thus the crystal was allowed to sink to the very rim of the pipette tip before being placed in the new solution. This often resulted in the crystal dropping in by a surface tension effect with very little transfer of liquid. This is important, because during the first couple of transfers there are still tiny fragments of the crystal dropping off the surface which could themselves, act as nuclei in the seeding droplet. This is undesirable, because the aim of macroseeding is to boost the size of an existing crystal and not to allow the protein to be depleted by new crystal formation.

The droplet to be seeded was initially of such composition that spontaneous nucleation was very unlikely to occur within a week or ten days. This corresponded to a droplet containing 9.75 % PEG 8 000, 65 mM sodium sulphate, and 32.5 mM sodium phosphate buffer pH 6.8, with UDGase at about 8.4 mg/ ml. A test grid proved that seeding was effective between 6 to 10.8 mg/ ml of UDGase. It was later also found that seeding was effective even in drops where spontaneous nucleation could immediately take place, though obviously protein was exhausted by the presence of new nuclei. Larger seeds were obtained in droplets with larger volumes. Seeding took between 12 hours to 72 hours to be complete. The resulting crystals were anywhere between 100 µm and 200 µm wide by variable lengths between 1 and 2 mm. The problem associated with seeding was that the seed crystals would invariably land on their sides in the target droplet and grow such that they were anchored tightly along their sides to the well. Consequently, any attempts to move the crystals with a hair would invariably result in the crystals being stressed at several points along their length and often the fragments were too small to do anything with other than to use them as seeds.

5.3.3 INTRODUCTION OF NUCLEATION POINTS BY MICROSEEDING

Microseeding was also used. This involves setting up a droplet that is not likely to spontaneously nucleate within a couple of days. After some time has elapsed, there is a protein skin on the droplet. An animal hair is pushed onto a crystal in another droplet such that the tip makes good contact. The hair is then drawn through 1 ml of distilled water several times and dried off with tissue. The point of the hair is used to prod the skin on the target droplet. This often results in a cluster of small crystals, but sometimes results in a single crystal nucleating from the skin. If there is no nucleation, then the skin can be prodded again and again until it occurs. A crystal has usually appeared within 12 hours of touching the skin of the droplet with the hair. Growth is at the normal rate for crystals of UDGase.

5.3.4 THE EFFECT OF DROPLET 'FEEDING' ON CRYSTAL GROWTH

In an attempt to grow larger crystals of the enzyme without the need to remove the crystals from their original growth positions, a study was made of the effects of continually adding fresh protein, or protein/ precipitant mixtures not likely to spontaneously nucleate, to droplets in which crystals had already begun to grow. This was done in attempt to by-pass the usual seeding problem of crystals falling onto their sides and adhering to the well. It was also thought that arrest of nucleation could be achieved, such that once a single nucleus had been seen to form, immediate dilution of the droplet might force crystal growth to continue only in the existing nucleus.

In practice, it was seen that some increase in size occurred with crystals that had been growing for a day or two, the increased growth continuing slowly for up to a week. This effect was seen only if a mixture of protein and precipitant was added to the droplet. The ratio of the added protein/ precipitant mixture was such that nucleation would not normally be seen to occur without seeding. However, additional nuclei did appear in these droplets in the majority of cases, and in many of the droplets which were 'fed', the target crystals for enlargement developed branch points. In the worst cases, the crystals became covered in needle-protrusions.

It is thought that these effects are to do with the slight variation in precipitant strength when the feeding solution is mixed. It is also thought that the appearance of new

nuclei is an apparent effect, and that these nuclei were in fact already present in the target droplet, their growth arrested by the accretion of available protein in the droplet onto the larger nuclei present. On addition of fresh protein to the droplet, they become surrounded with new material with which to continue their growth in much the same way as the target crystals. It was observed that in droplets significantly older than 4 or 5 days, the crystals became increasingly more prone to amassing outgrowths and less prone to smooth overall growth. In such older droplets, the appearance of nuclei which were not previously visible was not diminished.

Finally, in droplets which had just been set up, when the first nuclei were just becoming visible, protein was added to reduce the concentration of precipitant in the droplet to levels that are not normally sufficient to support nucleation, but are sufficient to support growth. In these droplets, it was seen that the number of crystals present when any droplet reached a near steady-state was not much greater than the number of nuclei seen when the initial protein addition was carried out. In these droplets, the growth was slower than is typical for crystals of UDGase and the branching and other splitting effects were more predominant than is normally the case. The average size for any crystal from such a droplet was not significantly greater than for a crystal in the normal 'un-fed' situation from a droplet which contained a similar number of nuclei. The aim was to try and stop the nucleation when one or two nuclei were present, which can sometimes occur spontaneously, but then to try and force more protein into these nuclei. It would appear from these trials, that it is not the total amount of protein in the droplet that is the factor, but it is the presence of a constant driving growth force which can only be supplied by the correct ratio of protein and precipitant concentrations in the droplet. Successful crystal size-optimisation using a very similar approach has recently been reported (Saridakis *et al.*, 1994).

5.3.5 USE OF AGAROSE GELS AS MATRICES FOR CONTROLLED CRYSTAL GROWTH

Very low percentage agarose gels as matrices to limit convective forces in crystallisation droplets have been described elsewhere (Robert and Lefaucheux, 1988; Robert *et al.*, 1992). Before the effect of increased protein concentration in the absence of glycerol was known, the precise published protocol (Robert *et al.*, 1992) was followed

in an attempt to grow larger crystals of UDGase. Microbatch, vapour diffusion, and liquid/ liquid diffusion were set up using precipitant stocks containing 0.05 % and 0.1 % agarose. The effect of the agarose was to inhibit crystallisation unless precipitant concentrations were higher than normal (>13 % PEG 8 000, and associated increases in sodium sulphate and buffer concentrations), and also to slow the growth rate significantly from the normal 12 hours to several days. The agarose did not prevent the formation of a protein skin, thus vapour diffusion was unsuccessful. The liquid-liquid diffusion produced a shower of needles of varying size, getting larger the further away from the precipitant laden agarose they grew. One crystal was obtained under microbatch; it seemed to be a perfect trigonal prism, but after a day of growth it began to split at the base, and ended up as a cluster of rods. However, these were good seed crystals when they were separated.

5.4 CO-CRYSTALLISATION EXPERIMENTS

UDGases have been shown to bind to single or double stranded DNA in a non-specific manner (Higley and Lloyd, 1993), and to short oligomers of DNA (Varshney and van de Sande, 1991), also to be inhibited by uracil - the end product of the reaction, and to an extent by certain uracil analogs (Krokan and Wittwer, 1981). It was hoped that co-crystallisation of an enzyme/ DNA or enzyme/ inhibitor complex could be achieved. This section presents the findings of the attempts to obtain such crystalline complexes.

5.4.1 ATTEMPTED CO-CRYSTALLISATION OF URACIL-DNA GLYCOSYLASE WITH URACIL AND SOME URACIL ANALOGS

While UDGase was being crystallised from stocks at less than 30 mg/ ml, the crystallisation precipitant was modified to include 1 to 3 mM uracil, dithiouracil and 6-aminouracil, as well as cibacron blue F3GA, the dye used in the Affigel Blue columns. The cibacron blue droplet caused the protein to precipitate heavily, and any small crystals obtained were colourless, rather than blue. These results may indicate that the method of UDGase binding to cibacron blue is not via a specific nucleotide binding site interaction. However, this may not be the case because the protein had not been pre-incubated with the dye before precipitant addition.

In the case of the uracil and its analogs, there was inhibition of crystallisation,

even in the presence of high precipitant concentrations. On one occasion, a droplet that contained uracil at 1 mM yielded 2 crystals, which were both about 100 µm wide by 2 mm long, but the morphology was not any different to that of normal crystals, and they were both shattered during attempted removal for mounting due to their being anchored along their length to the well surface. Seeding small crystals into droplets containing uracil at 2 mM causes striations to appear on the surface of the crystals within a couple of days, and the crystals appear to erode, without loss of birefringence, over the course of a few weeks. If the protein is pre-incubated with uracil prior to setting up the crystallisation, to give a final uracil concentration of 2 mM, then nucleation is delayed for anywhere between 10 days to a month. After this delay, there is the appearance of a sheaf of crystalline plates. These plates appear to be less than 5 µm thick, but up to 1 mm long and about 20 to 80 µm wide. Crystallisation of protein pre-mixed with uracil as described, can be induced by increasing the relative concentration of precipitant by a factor of five over the normal conditions. This results in the appearance of large crystals of trigonal morphology, though these have many smaller crystals growing randomly from them.

5.4.2 CO-CRYSTALLISATION EXPERIMENTS WITH URACIL-DNA GLYCOSYLASE/ DNA OLIGOMER MIXTURES

These studies were carried out with the 60 mg/ ml and over, concentrated glycerol-free protein. The oligomeric DNA used was initially a trimeric single stranded thymidine 5'-phosphorylated DNA (Sigma Chemical Company Ltd).

Protein was added to vacuum desiccated DNA such that molar ratios of 1:1 through to 4:1 of DNA to protein were made. These were mixed thoroughly, placed at 37°C for 60 seconds and left on ice for a minimum of 5 minutes. The protein/ DNA mixture was then added to precipitant in the normal manner on Nescofilm™, and dispensed under oil into Terazaki plate wells. The precipitant used was the 0.2 M imidazole sulphate pH 6.8, 30 % PEG 8 000 mixture. Some drops were also set up using the 0.1 M sodium phosphate buffer pH 6.8, 0.2 M sodium sulphate, 30 % PEG 8 000 mixture.

Only 2 out of 8 droplets nucleated, resulting in dense clusters of tetragonal rods, one cluster in each droplet, one being with the phosphate precipitant, the other with the imidazole. Both were 3:1 molar ratios of DNA to protein. The rods looked very hollow, and some were extracted for examination. Their mechanical strength was greater

than that of the native trigonal crystals of similar dimensions (100 µm wide by 1 mm long), and they were definitely hollow (Figure **5.4.2**). One such rod was used as a macroseed, but was not etched in water, rather it was quickly rinsed in water and added with minimal liquid transfer to another drop containing the DNA/ protein/ precipitant mixture. Over the course of 2 or 3 days, the crystal grew to dimensions of 150 µm wide by 1 mm long, and the hollow was filled in. The X-ray diffraction data for this crystal (Section **6.3.2**) revealed that it was different in its symmetry and packing from native trigonal crystals. The crystal must have been dried very well during mounting, because it was subsequently found that the crystals disintegrated into bundles of tiny thin needles in the normal stabilisation buffer (Figure **5.4.2**).

Thin plate clusters are another morphology that occurs with the protein/ trimeric thymidine DNA mixture. With dimeric thymidine (Sigma Chemical Company Ltd), the plates are the favoured morphology (Figure **5.4.2**), and with tetrameric thymidine (Sigma Chemical Company Ltd), it is difficult to generate anything but clumps of crystalline material. Droplets were also set up using pre-mixed trimeric single stranded 5'-phosphorylated deoxycytidine (Sigma Chemical Company Ltd), but these precipitated badly. This could be due to the fact that the deoxycytidine is indicated to be a less pure product (~86 % by HPLC, as opposed to >96 % by HPLC for thymidine), or it may be that the sample was vacuum desiccated too long, which tends to make the sample insoluble in the protein. Thus the experiment was repeated with a sample of deoxycytidine trimer that was vacuum desiccated until it was just dry. Even in this case the droplets were either clear or produced small (~100 µm wide by <500 µm long) crystals of native trigonal morphology. This experiment was repeated a number of times and the results were always the same. In addition, some droplets were set up where the protein had been pre-mixed with trimeric single stranded 5'-phosphorylated deoxyadenosine. The appearance of these droplets, within 12 hours, was clear with a dense suspension of small (~10µm) opaque points with whisker-like protrusions. These objects may well have been crystalline in nature, though it was difficult to manipulate them. Thin, needle-like bunches have grown in these experiments.

Some trials were also carried out with double stranded dodecameric oligonucleotides made for co-crystallisation trials with another protein, the *Bsp* RI (cytosine-5) methyltransferase. A survey was initially carried out to determine the approximate stoichiometry of the complex for optimal crystallisation. Judging by the degree of precipitation when the

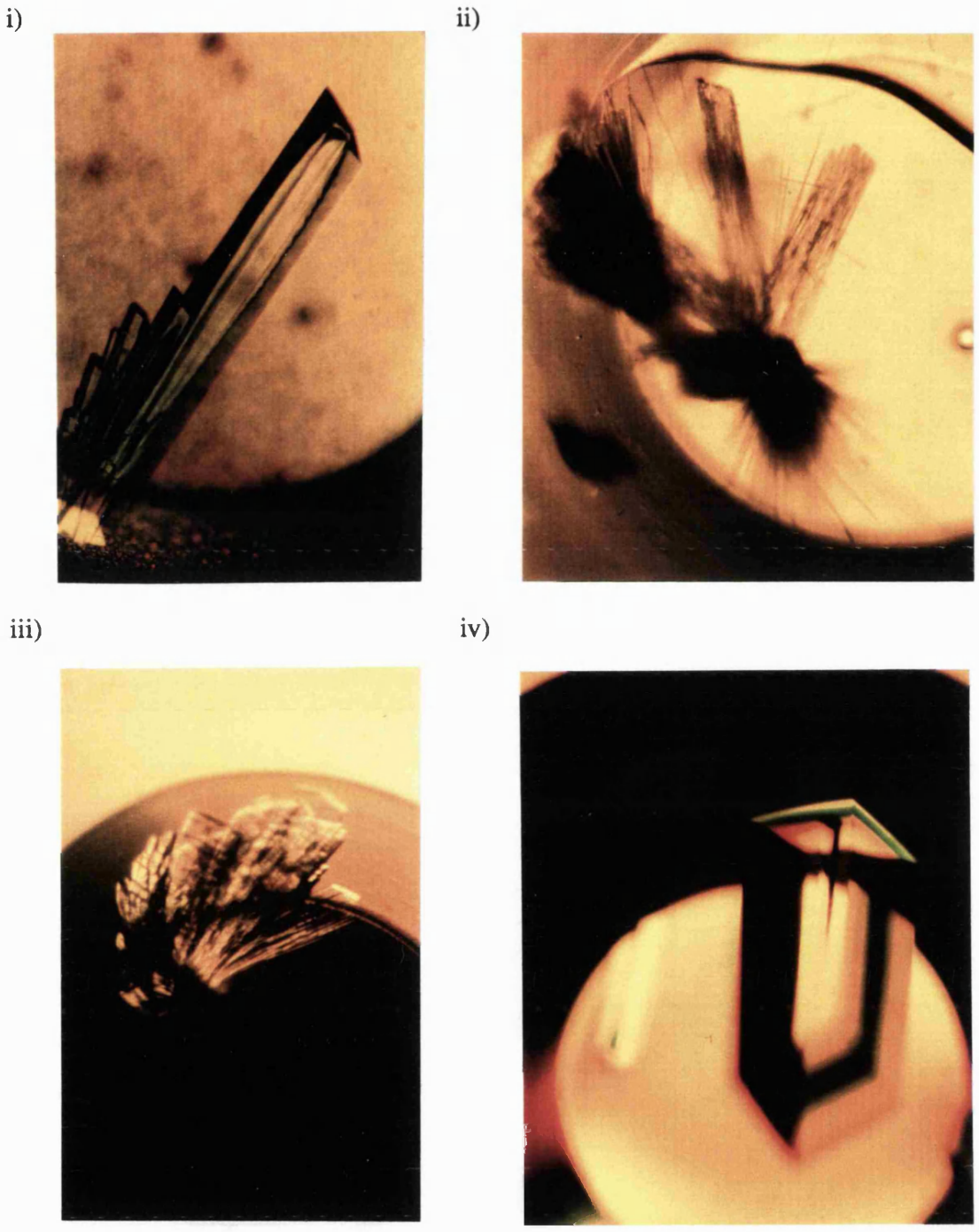


Figure 5.4.2

i) A cluster of rhombohedral rods, showing characteristic hollow appearance. Largest is approx. 0.15 x 0.8 mm (w x l).

ii) Disintegration of the HSV1 UDGase/ oligonucleotide co-crystals in the standard stabilisation solution used for native crystals.

iii) A cluster of thin plates obtained after mixing with a di-thymidine oligonucleotide.

iv) Isomorphous HSV1 UDGase/ pdT ^{l}U T co-crystal. This crystal (approx. 0.4 x 0.9 mm, w x l) was grown singly by careful microseeding (Section **5.3.3**).

complex was mixed with the precipitant used for the native crystallisations, it was decided that about 2 molecules of UDGase per oligonucleotide was best. These UDGase/ oligonucleotide complexes have so far proved resistant to crystallisation under similar conditions to those that produce the other crystal forms. Instead, irregularly shaped gel-like masses are immediately formed. These do not form crystals, though if the precipitant strength is low enough they tend to gradually dissolve. A very fine, sparse granular precipitate is observed in these droplets. After some weeks, some long (generally > 0.5 mm), thin (generally less than 200 µm wide) crystals with identical morphology to the unbound protein are seen to appear. It is likely that these are native, unbound protein crystals, though this has yet to be confirmed by X-ray diffraction studies. A 'sparse matrix' survey carried out on the complex showed that crystals could be obtained under conditions where the native enzyme has been crystallised. However, these conditions contain sufficient salt in the precipitant to disrupt a non-specific protein/ DNA complex, suggesting that the crystals are of unbound enzyme. Further evidence for this is the fact that the morphology of these crystal forms is identical to that obtained when unbound enzyme alone is crystallised.

In an attempt to generate isomorphous heavy-atom derivatives (Section **2.4.4**), DNA oligomers analogous to the tri-thymidine oligomer were synthesised. In one of the oligomers, the middle thymidine moiety is replaced by a 5-iodouridine moiety. In the second, the 5' thymidine is replaced by a 5-iodouridine. The third in this series of oligomers has both the 5' thymidine and the middle thymidine moieties replaced by 5-iodouridine moieties. The spatial configuration and size of 5-iodouracil is comparable to that of thymine, but iodine is big enough (iodine has 53 electrons, which is 44 electrons more than the methyl group it replaces) to provide a difference which should also be isomorphous to the tri-thymidine co-crystals. The pre-incubated protein/ oligomer mixtures, in the same molar ratio as that which was successful for the tri-thymidine/ UDGase co-crystals, were mixed with the 0.1 M sodium phosphate buffer pH 6.8, 0.2 M sodium sulphate, 30 % PEG 8 000 precipitant mixture. Crystals of identical visual appearance to those obtained with the tri-thymidine oligomer were obtained within the normal time-period of 24 hours (Figure **5.4.2**). The crystals were found to be isomorphous with the tri-thymidine co-crystals, and solved for a single heavy-atom site per monomer in the case of the first two oligomers, and for two sites per monomer in the double-iodinated oligomer (Section **6.3.2**).

6 STRUCTURAL STUDIES OF HSV1 URACIL-DNA GLYCOSYLASE

The production of strongly diffracting crystals of UDGase, and the identification of isomorphous heavy atom derivative crystals, paves the way for the detailed three dimensional molecular structure to be resolved. In addition the enzyme has also been used in studies such as circular dichroism analysis, which using protein in free solution, can give information about the type of secondary structure. Computer prediction of protein secondary structure is not a totally reliable technique, mainly owing to the relatively small number of known conformations in solved structures, compared to the potentially much larger number of possible structures for unknown sequences (Jones and Thornton, 1993). However, such programs have been used in an attempt to find likely secondary structure motifs in UDGase. In the end, it will be the three dimensional molecular structure determined by X-ray crystallography that will tell how effective such prediction methods are.

6.1 CIRCULAR DICHROISM ANALYSIS OF HSV1 URACIL-DNA GLYCOSYLASE

Circular dichroism (CD) analysis is a spectroscopic technique in which the wavelength dependent absorption of circularly polarised light is measured. This analysis can be used to estimate the type and extent of secondary structure features in a macromolecule that is optically active. Due to the optically active state of most biological macromolecules, especially proteins and nucleic acids, this technique is well suited to the study of the structures and interactions of such molecules. The sensitivity of the technique to structural features means that it is particularly well suited to the detection of conformational changes upon binding of one optically active molecule to another molecule. Since the technique is a useful tool in structural analysis, a CD spectrum showing the likely type and extent of the secondary structure of the UDGase from HSV1 was collected. In addition, the interaction of this enzyme with DNA was studied, as was the effect of the end product of the enzymatic reaction, free uracil base.

The following is a short description of the theory and application of this technique,

which has been more fully dealt with elsewhere (Bayley, 1980; Freifelder, 1982).

6.1.1 THEORY AND APPLICATION OF CIRCULAR DICHROISM ANALYSIS

Circular dichroism is the difference in absorption of right circularly polarised light and left circularly polarised light. Molecules sensitive to the specific left and right polarisation of light are known as optically active molecules, and they are asymmetric in that their mirror images cannot be superimposed. In most cases, optically active molecules do not have identical extinction coefficients for both left and right circularly polarised light, thus light which emerges from them will have been absorbed to a greater extent in either one of the left or right polarised directions. This will result in elliptically polarised light of various angles. When the ellipticity is due to an amplitude change (absorption) in the electric field vector of light, the result is called the circular dichroism (CD) of the sample.

The circular dichroism is represented by $\Delta\varepsilon$, where $\varepsilon_L - \varepsilon_R = \Delta\varepsilon$ (ε = the molar extinction coefficient of the sample, L and R represent the left and right polarised light respectively), and the circular dichroism is called positive when $\varepsilon_L - \varepsilon_R > 0$, and negative when $\varepsilon_L - \varepsilon_R < 0$. Experimentally, $\Delta\varepsilon$ is measured, but traditionally the ellipticity (θ) is plotted. A CD spectrum is the ellipticity vs wavelength. CD is superior to absorption spectroscopy in that it gives information not only about the extent of absorption of light, but also about the degree of optical activity in the absorbing matter and the polarity of the absorption.

The far UV CD spectrum of a protein may be thought of as an additive spectrum of the individual spectra of the constituent amino acids. However, the interactions between amino acids in the complex secondary, and higher, structures modify the spectral pattern. Thus, various regular structures such as α-helices and β-sheets have characteristic spectral signatures. However, the pattern is complicated by less readily distinguishable motifs, like β-turns, and unpredictable contributions by prosthetic groups, as well as from certain spatial configurations of aromatic amino acids. With nucleic acids, the spectra are more easily predictable. The interaction between a base and its neighbouring base makes the strongest contribution to the CD, thus summation of the CD of all neighbouring bases gives a very nearly accurate CD spectrum for the polynucleotide. The bases become

optically active once attached to the ribose via the N-glycosidic bond. This is with the exception of uracil. Also, it appears that all the polynucleotides have stacked bases, again with the exception of uracil. This stacking is less obvious in DNA. With double-stranded nucleic acids, the prediction of CD spectra is at a much less developed stage.

The near UV spectrum is composed mainly from the contributions of amino acid side chains, and in particular, aromatic side chains, and will thus be expected to alter if the conformation, or any other property which involves the aromatic environment, such as binding of ligands or other macromolecules, changes. The technique is thus suitable for biochemical studies of a protein and any molecules which it binds. In the case of uracil, which has no CD signal under normal circumstances, the observation of an induced CD signal will indicate that it has entered into an optically active environment, and thus may possibly be bound. An experiment was conducted to determine whether this occurs with UDGase. This, and other experimental data are now presented.

6.1.2 RESULTS OF THE CIRCULAR DICHROISM ANALYSIS CARRIED OUT ON THE URACIL-DNA GLYCOSYLASE OF HSV1, WITH, AND WITHOUT, DNA OR URACIL

Circular dichroism analysis was carried out to determine the likely secondary structure features of the HSV1 UDGase. The far ultraviolet (UV) CD spectrum (Figure **6.1.2 a**) indicates that the protein is likely to consist of both α-helix and β-sheet secondary structure elements.

In addition, the interaction of this protein with DNA was studied by CD analysis. The near UV CD spectra of the protein, and of the DNA molecules it was to be mixed with, were measured. The near UV spectrum for one or both components should show a change in CD signal if there is a conformational change on binding. A protein/ DNA mixture was studied. In the first experiment, the protein was mixed with a double stranded dodecameric oligonucleotide. A 3.5-fold excess of protein over nucleic acid was used, and the mixture was made up in 25 mM sodium phosphate buffer at pH 8.0. There was a small change in the near UV CD spectra of both the protein and the nucleic acid. The spectra produced are analysed for changes in the CD of both the nucleic acid, and of the protein, by subtracting the individual CD spectra of these components from the CD of the mixture.

The instrument used for the CD analysis was a Jasco 600 Spectropolarimeter. The buffer in all experiments for which a trace is shown was 25 mM sodium phosphate pH 8.0. A reference line was obtained prior to each reading by running a spectrum of the buffer alone. This was subsequently subtracted from the spectrum of the buffer with sample (protein and/ or DNA). The path lengths of the cells used for CD analysis were 0.5 cm for the far UV analysis of HSV1 UDGase (Figure 6.1.2a), 0.1 cm for the near UV analysis of HSV1 UDGase (Figure 6.1.2b black line), and 1cm for the protein/ DNA mixtures (Figure 6.1.2c blue line) and protein/ uracil mixtures (Figure 6.1.2d coloured lines). The concentration of the samples prior to CD analysis was determined from absorbance at 280 nm (protein), and at 260 nm (DNA) using a double beam spectrophotometer (Shimadzu Graphicord UV-240). The concentration of the HSV1 UDGase used for far UV analysis (Figure 6.1.2a) was equivalent to 0.4 mg/ ml. The concentration of the HSV1 UDGase used for near UV analysis was equivalent to 0.5 mg/ ml (Figure 6.1.2b black line) or 0.8 mg/ ml (Figure 6.1.2d black line). The concentration of HSV1 UDGase in the protein/ DNA mixtures was equivalent to 0.385 mg/ ml. The DNA stocks were diluted for the mixtures such that there was 3.5 times more protein than DNA in the mixtures.

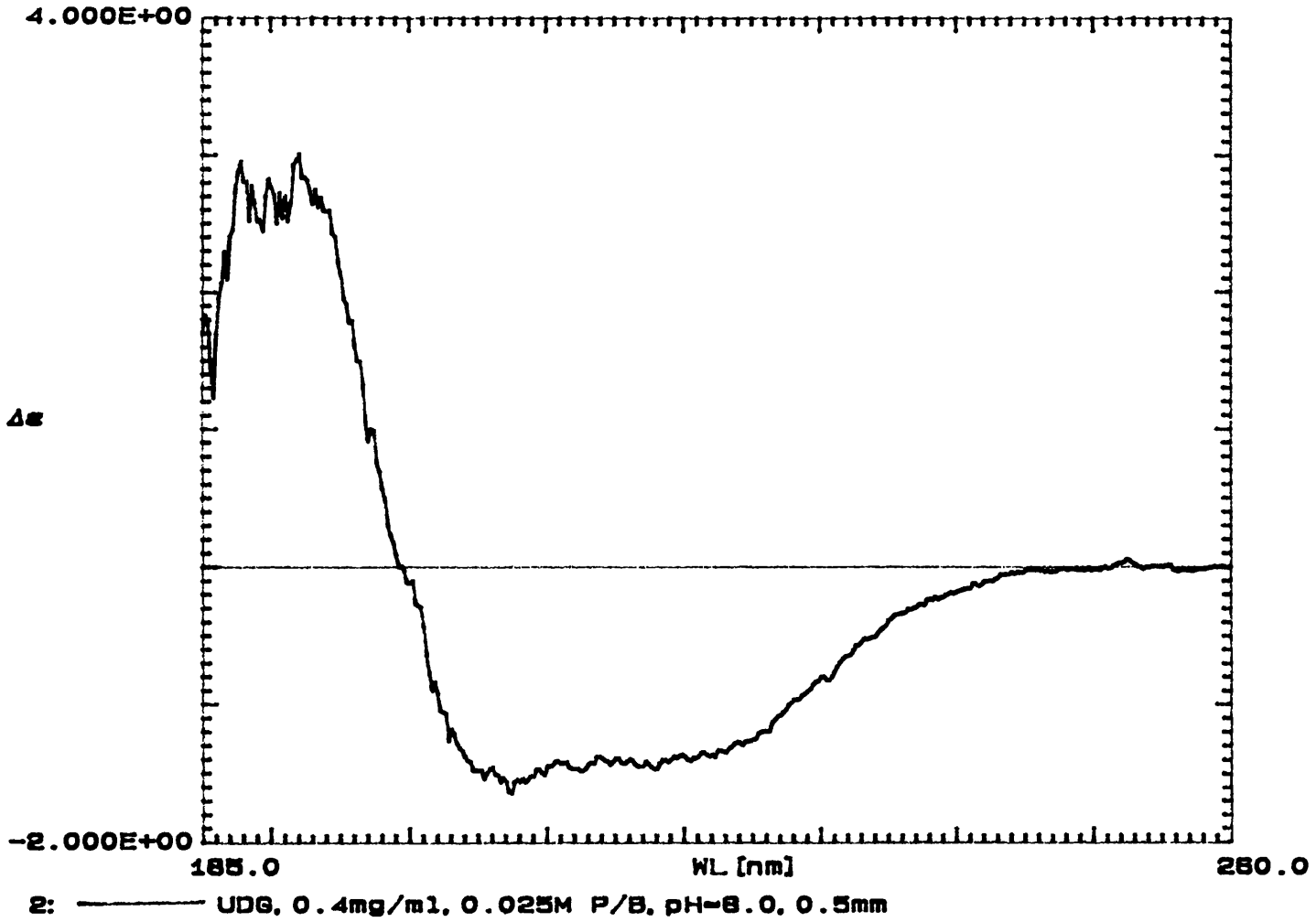


Figure 6.1.2a

The far–UV UDGase CD spectrum indicates that the protein is likely to be composed of a mixture of α–helix and β–sheet secondary structure elements. The helix is indicated by the positive peak around 190–200 nm, the β–sheet contribution is the flattened negative trough in the CD.

The protein is not very stable in phosphate buffer, and so a later set of experiments were performed with the phosphate buffer replaced by a low (10 mM) concentration of Tris·HCl buffer at pH 8.0. The original oligonucleotide was used again, at the original protein/ DNA ratio, and also at an approximately equimolar ratio. However, the change in the near UV CD spectra was no clearer. A second mixture was also made, with a single stranded dodecameric oligonucleotide (one of the strands of the first oligonucleotide). This also showed very little change in the near UV CD spectrum. Finally, a third mixture was made, comprising the protein and a single stranded trimeric oligonucleotide composed only of thymidine (Section **5.4.2**). Again, there was very little change in the CD spectrum. Representative example spectra from one of the mixtures are presented, showing the near UV spectra of UDGase before, and after, mixture (Figure **6.1.2 b**); and the double stranded dodecameric oligonucleotide before, and after, mixture (Figure **6.1.2 c**).

An attempt was also made to study the effect of uracil, one of the products of the reaction of UDGase, on the protein. The free base uracil is only sparingly soluble, and so a concentrated stock has to be made up in an organic solvent such as dimethyl sulphoxide. Various concentrations of uracil in the mixture were tested, though very high concentrations of uracil were not possible due to the large absorbance effect of the base at these concentrations interfering with the CD detection. The concentrations used were 100 µM and 250 µM. The CD spectra of UDGase before, and after, addition of uracil, at both concentrations, are shown (Figure **6.1.2 d**). There appears to have been an enhancement of the CD signal, especially noticeable in the region around 294 nm where there is a contributory bump from the protein's aromatic residues. This bump is elevated upon addition of uracil. The initial change seen upon addition of 100 µM uracil, is not further enhanced by moving to the higher concentration.

The significance of these findings is discussed in section **7.2.3**.

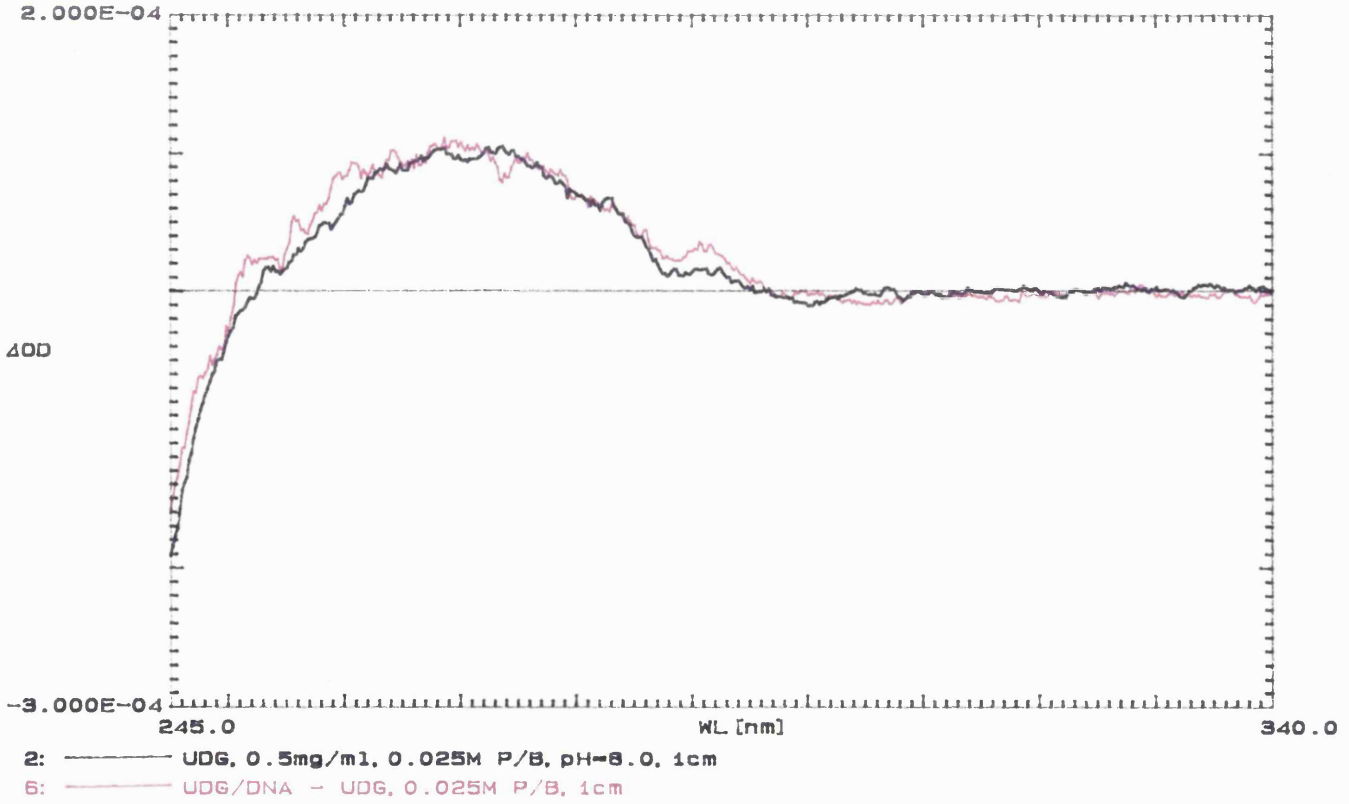


Figure 6.1.2b

The effect on the near–UV UDGase CD spectrum upon binding to a double stranded dodecameric oligonucleotide. The red line shows the CD of the UDGase component of the mixture. The black line shows the original CD spectrum of the UDGase, before adding the DNA.

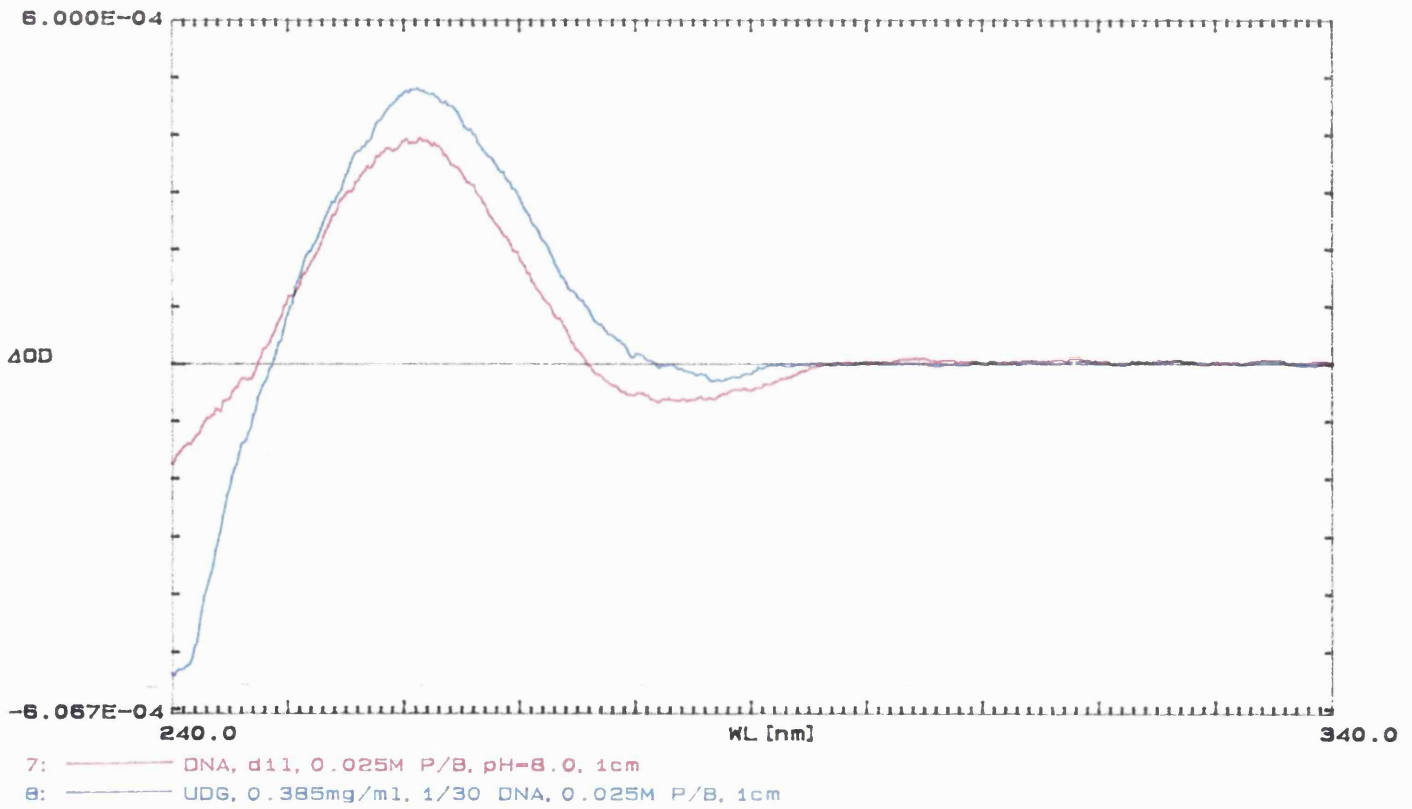


Figure 6.1.2c

CD spectrum showing the effect of UDGase binding on the secondary structure of a double stranded dodecameric oligonucleotide. The blue line shows the CD spectrum of the DNA component of the mixture. The red line shows the original CD spectrum of the DNA, before adding UDGase.

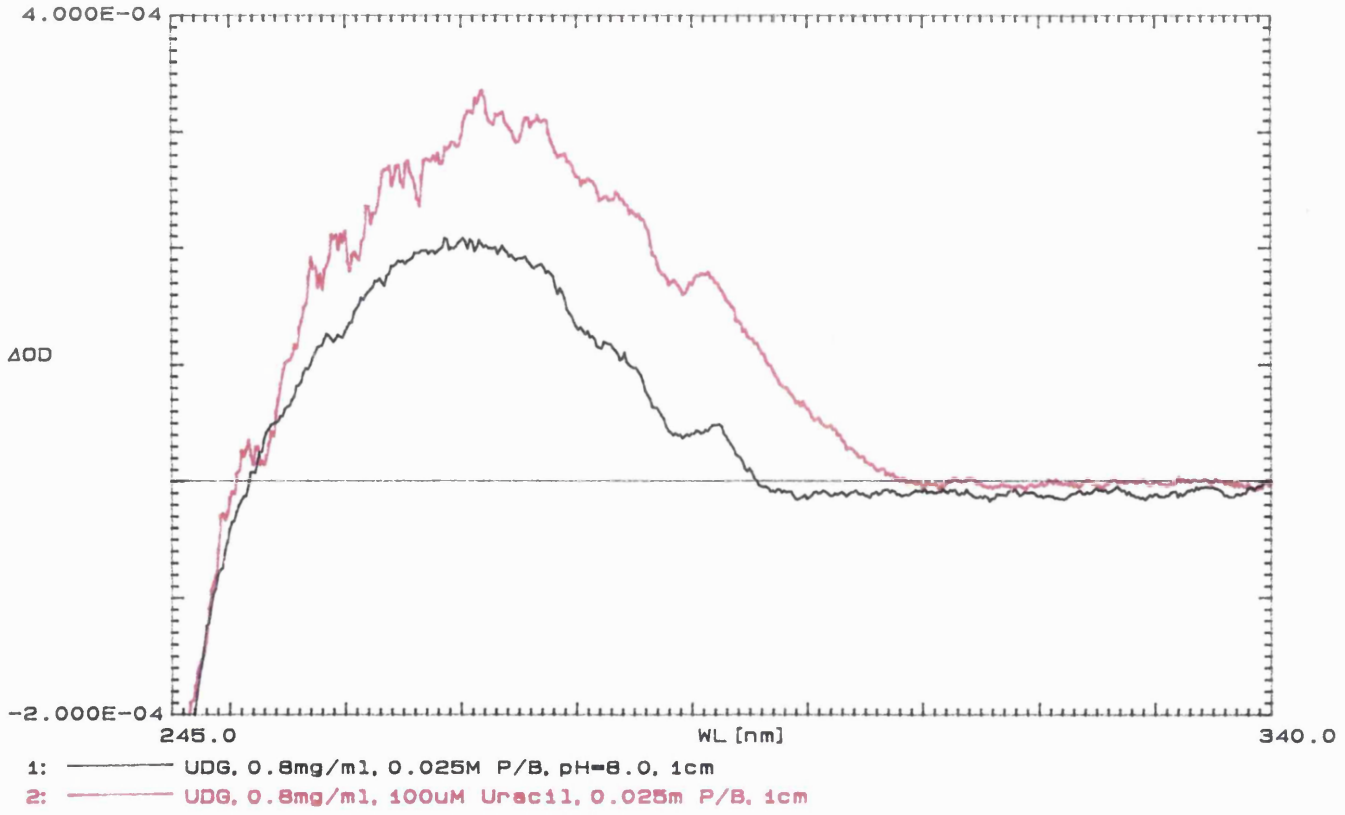


Figure 6.1.2d

The effect on the near–UV UDGase CD spectrum upon addition of 100 μM free uracil. The red line shows the CD of the UDGase component of the mixture. The black line shows the CD spectrum of UDGase prior to uracil addition.

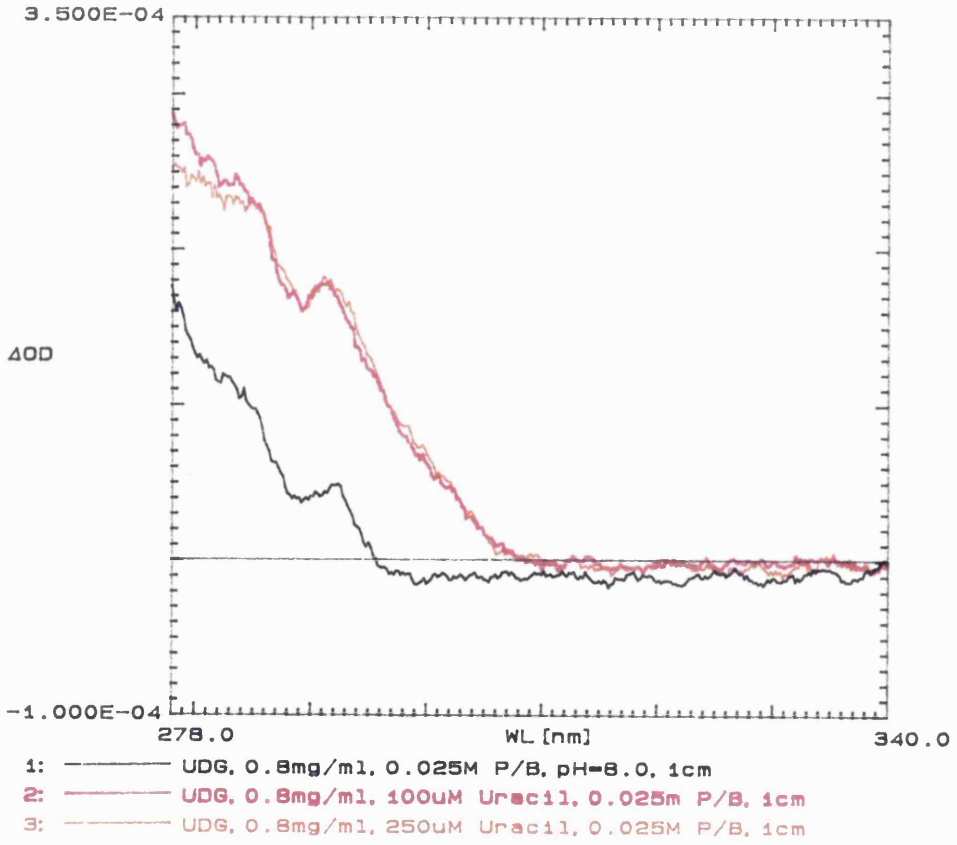


Figure 6.1.2 d (Continued . .)

The enhancement of the near-UV UDGase CD spectrum upon addition of free uracil is not markedly changed by increasing the concentration of free uracil from 100 μM to 250 μM (orange line). The enhancement of the aromatic region of the CD, around 293 nm, is shown here.

The unit cell and space group assignments for the two crystal forms of HSV1 UDGase were made in the following manner. A first estimate of the likely space group is made from visual inspection of the crystals. In the case of the native unbound UDGase crystals, it was thought that a trigonal or hexagonal lattice would be likely. At this stage, although it was thought that the crystals were triangular prisms, some crystals grew with blunted corners which looked rather hexagonal. Upon inspection of the diffraction oscillation images, some indications of a trigonal or hexagonal symmetry were apparent. The unit cell dimensions were measured using the program *hklview*, which requires the input of the crystal to detector distance and the wavelength of the incident X-rays. Since the crystals were rods, they tended to mount with the long axis almost perpendicular to the beam. This enabled a clear view of two of the cell edges, a and b. The c axis was not immediately visible, but the program REFIX was run with initial cell estimates and the trigonal space group P3. The data processed well in P3 or $P3_1$ / $P3_2$, and so the cell parameters were refined by the REFIX and MOSFLM postrefinement software.

With the DNA co-crystals, it was thought that the space group was likely to be orthorhombic or tetragonal, because the crystals were rectangular rods. A similar approach to that taken with the unbound crystals was used. Like the unbound form, these crystals were rods and tended to mount with the long axis perpendicular to the beam. The initial cell parameters refined best in space group $P2_12_12$ or $P2_12_12_1$. The data processed well in either space group.

The ambiguities remained because the c axes were perpendicular to the beam, and so the systematic absences produced by symmetry related atoms that indicate that a space group has a further screw axis were not visible. In both cases the use of a FAST area-detector resolved the ambiguity by allowing the crystal to be offset using the kappa geometry, this allowed the c axis of the crystal to be viewed, and systematic absences were observed in both cases. The trigonal crystals were thus assigned $P3_1$ / $P3_2$, and the orthorhombic crystals were assigned $P2_12_12_1$.

The number of molecules in the unit cell, the number of molecules in the asymmetric unit, and the proportion of solvent by volume in the unit cell were estimated by the procedure of Matthews (1968). Briefly, the crystal volume per unit molecular weight, V_M , is used to determine the fractional volume of the unit cell occupied by the protein and/ or solvent. The number of molecules in the asymmetric unit is initially assumed to be 1, and the V_M is calculated by dividing the volume of the asymmetric unit by the molecular mass of the

protein.

The fractional solvent content is then calculated by taking into account a constant related to the partial specific volume of an 'average' protein, the formula being $1-(1.23/ V_M)$. If the calculated proportion of solvent in the crystal is in agreement with the observed diffraction characteristics and mechanical strength of the crystals, then it is assumed to be a reliable estimate. If the value is not in agreement, then a larger number of molecules per asymmetric unit should be used in the calculation. Further evidence, such as the presence of non-crystallographic symmetry will support the notion that there is more than one molecule per asymmetric unit.

In the case of the trigonal crystals, V_M was found to be 2.16, and the solvent content was 43 % with one molecule per asymmetric unit (unit cell a=b=64.8Å, c=48.7Å $\alpha=\beta=90°,\gamma=120°$). With the orthorhombic crystals, V_M was 2.22, and the solvent content was 44.6 % with one molecule per asymmetric unit (unit cell a=46.7Å, b=61.8Å, c=86.7Å, $\alpha=\beta=\gamma=90°$).

6.2 AN ATTEMPTED PREDICTION OF THE SECONDARY STRUCTURE OF URACIL-DNA GLYCOSYLASE FROM ITS PRIMARY STRUCTURE BY USING A SEQUENCE/ STRUCTURE THREADING PROGRAM

A program which performs *optimal sequence threading* (Jones *et al*, 1992; Jones and Thornton, 1993) was used in attempt to find a likely folding pattern for UDGase. The program used 102 known folds, from the Brookhaven structure databank, as scaffolds upon which the sequence of UDGase was 'threaded' in order to identify a low energy match with a known fold. The results of the threading analysis did not reveal any highly likely similarities with known folds in the Brookhaven databank. The best score was a very poor fit with a Walker motif, a common fold in nucleotide binding proteins (Walker *et al.*, 1982).

Optimal sequence threading takes into account detailed pairwise interactions between residues and their nearest neighbour environments. The potentials used are derived from Sippl potentials (Jones and Thornton, 1993), which can measure short (residue to residue) to long (such as disulphide bonds) range effects residue interactions. A solvation potential is used, taking into account the solvent exposed area of a particular residue. Finally, loop regions in structures are ignored, on the basis that these are the least conserved regions of proteins. The process is achieved by optimally fitting the test sequence, in thirteen residue blocks, to each library fold, with the energy of each threading being the sum of the proposed pairwise interactions, using C_β atom positions which are less variant in their positions from structure to structure. The lowest energy folds are the most likely matches, these would result for instance, from hydrophobic residues being in buried rather than exposed positions, and *vice versa* for hydrophilic residues. The results are presented in ascending order.

6.3 RESULTS AND ANALYSIS OF CRYSTALLOGRAPHIC DATA COLLECTED FOR URACIL-DNA GLYCOSYLASE

The results obtained for all the X-ray diffraction and heavy atom soaking work in this project are presented in table **6.3**. Main abbreviations used in this table are as follows:

NAT Native HSV1 UDGase crystals (unbound form)

DNACO HSV1 UDGase crystals (DNA bound form)

M	MAR Research Image-Plate Detector System
R	Rigaku R-axis Image-Plate Detector System
F	Enraf-Nonius FAST Area Detector System
S	MAR Research Image-Plate Detector System with Synchrotron radiation (with wavelength used)
NR	Data not retrievable
***ND	No diffraction observed
***NP	Data Unprocessable
Grad.	Deviation from the least-squares straight line calculated by normal probability analysis (Howell and Smith, 1992). Raw= Including all data, Corr.= using corrected data (without the strongest 10 % of reflections)
?	Not Recorded
n/a	Not applicable

Minor abbreviations in table **6.3**:
Hr = Hours; min = Minutes; Comp. = Heavy atom compound; conc. = Concentration of heavy atom compound; res. = Resolution; Deg. = Degrees; ddH_2O = deionised water; Na-Caco = sodium cacodylate; Prot = protein.

The strategies used in the heavy atom soaking work are described in section **6.3.1** and section **7.2.4.**

6.3.1 AN ATTEMPT TO GENERATE ISOMORPHOUS HEAVY ATOM DERIVATIVES BY DIRECT MIXING OF THE PROTEIN WITH THE COMPOUND PRIOR TO CRYSTALLISATION

Due to the slow progress made with attempting to generate isomorphous heavy atom derivatives of HSV1 UDGase, it was decided that some attempts would be made at reacting the protein stock directly with heavy atom reagents prior to crystallisation. This type of approach is not novel, but is not the usual first choice method, as it has been known to result in non-isomorphous crystals (Petsko, 1985).

The approach taken was to mix the concentrated 60 mg/ ml stock of HSV1 UDGase with concentrated solutions of heavy atom compounds, such that there was a minimal volume increase, and the final molar ratios of protein to compound were 1:1 and

The data presented in table **6.3** were processed from diffraction oscillation images by Dr.Laurence Pearl. All crystallisations, soaks, and data collection were performed by Renos Savva, with the exception of data collected on the FAST area detector with Dr.Pearl.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
Native 1	NAT n/a	n/a	n/a	90 M	11.2-3.0 3.2	3768	2.8	97.6	0.100	n/a	n/a	n/a
Native 2	NAT n/a	n/a	n/a	90 M	11.2-3.0 3.0	4442	2.1	94.7	0.093	n/a	n/a	n/a
Native 3	NAT n/a	n/a	n/a	90 M	11.2-3.0 3.0	3817	3.3	81.4	0.784	n/a	n/a	n/a
Native 4	NAT n/a	n/a	n/a	120 M	9.8-2.5 2.5	NR	NR	99.6	0.077	n/a	n/a	n/a
Native 5	NAT n/a	n/a	n/a	40 S0.88	7.5-1.8 1.8	16431	1.3	77.5	0.074	n/a	n/a	n/a
Native 6	NAT n/a	n/a	n/a	80 S0.88	3.6-1.75 1.75	19073	2.4	82.4	0.055	n/a	n/a	n/a
PCMBS	NAT 1.0	HEPES 6.8	2 Hr	M	***ND							
PCMBS	NAT 1.0	HEPES 6.8	17 Hr (40 min)	48 M	9.8-2.5 2.5	7465	1.5	93.2	0.088	2.859	2.126	0.103
PCMBS	NAT 2.0	HEPES 6.8	12 Hr	? F	6.0-2.73 2.73	4067	1.3	66.6	0.042	3.584	2.831	0.104
PCMBS	NAT 1.5-10.0	HEPES 6.8	41 Hr tot. (10 min)	63 M	9.8-2.5 2.5	7670	2.0	96.4	0.047	6.162	6.341	0.205

Table 6.3

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
PCMBS	NAT 1.5-10.0	HEPES 6.8	41 Hr tot. (16 Hr)	68 M	10.7-2.8 2.8	4791	1.2	83.4	0.038	3.888	3.049	0.122
PCMBS	NAT 10.0	Tris-HCl 8.0	16 Hr (4 Hr)	S0.87	***ND							
PCMBS	NAT 1.5-10.0	HEPES 7.1	53 Hr tot. (90 min)	90 S0.87	9.8-2.5 2.5	8126	2.8	99.4	0.027	3.584	2.712	0.094
PCMBS	NAT 2.0	HEPES 7.1	28 days	25 M	***ND							
PCMBS	NAT 1.5-10.0	HEPES 7.1	41 Hr tot. (12 Hr)	90 S0.91	9.8-2.5 2.5	9010	2.8	97.1	0.043	3.680	2.920	0.104
PCMBS	NAT 1.5-5.0	HEPES 7.1	72 Hr tot. (12 Hr)	90 S0.88	***NP							
PCMBS	NAT 1.5-10	HEPES 7.1	41 Hr tot. (12 Hr)	44 S0.87	***NP							
Satd. 5 mM $Hg_2(CH_3CO_2)$	NAT 1/6 - 1/2	HEPES 6.8	19 Hr tot. (20 min)	86 M	9.8-2.5 2.5	7859	2.7	98.3	0.036	3.739	2.690	0.101
$Hg_2(CH_3CO_2)$	NAT 1.5	Tris-HCl 8.0	24 Hr (90 min)	90 M	9.8-2.5 2.5	7910	2.8	97.7	0.028	4.109	3.151	0.102
$Hg_2(CH_3CO_2)$	NAT 2.0	Na-Caco. 7.4	28 days	M	***ND							

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.	
										% Grad. Raw	% Grad. Corr.	R-Fac.
$Hg_2(CH_3CO_2)$	NAT 6.1	Na-Caco. 7.4	28 days	M	***ND							
Mersalyl Acid	NAT 1.0	Tris-HCl 8.0	48 Hr (90 min)	90 M	9.8-2.5 2.5	7745	2.9	95.3	0.025	4.357	3.393	0.099
Ethyl-Hg-PO_4	NAT 9.5	HEPES 6.8	12 Hr	M	***ND							
Ethyl-Hg-PO_4	NAT 4.0	HEPES 6.8	12 Hr	M	***ND							
Ethyl-Hg-PO_4	NAT 0.35	HEPES 6.8	2 Hr (24 Hr)	90 M	9.8-2.5 2.5	7807	2.9	97.6	0.040	3.430	2.568	0.097
Ethyl-Hg-PO_4	NAT 1.0	HEPES 6.8	8 Hr (16 Hr)	85 M	9.8-2.5 2.5	7007	3.0	86.5	0.040	3.812	2.981	0.098
Ethyl-Hg-PO_4	NAT 2.0	Tris-HCl 8.4	28 days	M	***ND							
Ethyl-Hg-PO_4	NAT 2.0	Na-Caco. 7.4	28 days	M	***ND							
Baker's Di-Mercurial	NAT 3.5	HEPES 6.8	12 Hr	60 M	9.8-2.5 2.5	7740	1.9	95.4	0.065	2.565	1.970	0.095
Phenylmercury chloride	NAT 2.0	Tris-HCl 8.4	28 days	M	***ND							

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
K_2PtCl_4	NAT 5.0	HEPES 6.8	22 Hr	M	***ND							
K_2PtCl_4	NAT 0.5	HEPES 6.8	21 Hr	M	***NP							
K_2PtCl_4	NAT 0.5	HEPES 7.15	16 Hr	61 M	10.7-2.8 2.8	5529	1.9	96.6	0.091	2.906	2.150	0.125
K_2PtCl_4	NAT 2.0	Glycine-NaoH 9.2	28 days	M	***ND							
K_2PtCl_4	NAT 2.0	HEPES 7.1	28 days	M	***ND							
K_2PtCl_4	NAT 1.0	HEPES 7.1	19 days (3.5 Hr)	M	***ND							
K_2PtCl_4	NAT 0.25	HEPES 7.1	19 days (4 Hr)	60 M	10.7-2.8 2.8	5331	2.0	94.3	0.050	3.638	2.995	0.134
K_2I_4Pt	NAT 3.0	HEPES 7.1	22 Hr (7.5 Hr)	85 M	10.7-2.8 2.8	5665	2.2	98.7	0.032	3.758	3.035	0.099
cis-$PtCl_2$	NAT 2.0	HEPES 7.1	22 Hr (7.5 Hr)	90 M	10.7-2.8 2.8	5632	2.6	99.2	0.041	4.364	3.111	0.106
K_2PtCN_4	NAT 2.0	MES 6.6	30 days	? F	9.8-2.5 2.5	6940	1.6	86.5	0.049	3.075	2.423	0.108

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
$KAuCl_4$	NAT 0.5	HEPES 6.8	24 Hr	M	***ND							
$KAuCl_4$	NAT 0.5	HEPES 6.8	45 min (30 min)	26 M	***NP							
$KAuCl_4$	Prot+Au 1:2 Mix	n/a	2 day growth	34 M	***NP							
$KAuCl_4$	NAT 2.0	Glycine-NaOH 9.2	28 days	M	***ND							
$KAuCN_2$	NAT 2.0	MES 6.6	30 days	M	***ND							
$AgNO_3$	NAT 37.4	HEPES 7.1	17 Hr (45 min)	M	***ND							
$AgNO_3$	NAT 6.0	HEPES 7.1	15 Hr (40 min)	M	***ND							
$AgNO_3$	NAT 5.0	HEPES 7.1	90 min (25 min)	M	***ND							
$AgNO_3$	NAT 4.9	HEPES 7.1	45 min (35 min)	33 M	10.7-2.8 2.8	3376	1.7	58.5	0.041	3.211	2.387	0.092
$AgNO_3$	NAT 0.25	Tris-HCl 8.0	9 days (90 min)	90 M	9.8-2.5 2.5	7773	2.6	95.7	0.039	3.179	2.469	0.098

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
K_3IrCl_6	NAT 2.0	HEPES 7.1	28 days	67 M	9.8-2.5 2.5	7730	2.1	95.7	0.043	5.805	4.846	0.204
K_3IrCl_6	NAT 2.0	HEPES 7.1	28 days	90 S0.92	9.8-2.5 2.5	7571	2.8	96.0	0.040	7.204	6.095	0.267
K_2OsCl_6	NAT 1.2	HEPES 7.1	19 days (3 Hr)	? F	10.7-2.8 2.8	4234	1.3	74.2	0.054	6.038	6.166	0.243
K_2OsCl_6	NAT 1.0	HEPES 7.1	3 days (2 Hr)	90 R	10.7-2.8 2.8	5324	2.9	97.8	0.042	3.566	2.785	0.106
$KReO_4$	NAT 1.0	HEPES 7.15	12 Hr (30 min)	30 M	9.8-2.5 2.5	4804	1.6	59.3	0.085	3.031	2.042	0.121
Na_2WO_4	NAT 1.4	HEPES 7.1	19 days (4 Hr)	90 M	10.4-2.7 2.7	6216	2.9	99.5	0.051	3.386	2.635	0.103
$NaBiO_3$	NAT 1.0	HEPES 7.1	19 days (4 Hr)	90 M	10.1-2.6 2.6	6198	3.1	91.0	0.030	4.449	3.317	0.098
$UO_2(NO_3)_2$	NAT 1.0	Tris-HCl 8.0	77 Hr (90 min)	89 M	9.8-2.5 2.5	7722	2.9	95.0	0.067	2.686	1.970	0.111
$Eu_2(CH_3CO_2)$	NAT 1.0-3.0	HEPES 6.8	16.5 Hr tot (40 min)	75 M	9.8-2.5 2.5	8078	2.4	99.5	0.048	2.890	2.012	0.108
$Sm_2(CH_3CO_2)$	NAT 1.5	HEPES 7.1	21 days (3 Hr)	74 M	10.1-2.6 2.6	6767	2.3	97.3	0.026	4.240	3.193	0.105

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

Heavy Atom Compound Used or *Name of Native Data Set*	Type of Crystal (Comp. conc. in mM)	Type of Buffer and pH in Soak	Length of Soak (Back -soak)	Deg. Data + Det. Type	Resolution Range (Peak Res. of Data) in Å	Number of Reflections	Multiplicity	% Completeness	R-Merge	Isomorphous Diff.		
										% Grad. Raw	% Grad. Corr.	R-Fac.
$Ce_2(SO_4)_3$	NAT 2.0	HEPES 7.1	3 Hr (30 min)	41 M	9.8-2.5 2.5	6529	1.5	80.2	0.036	3.261	2.388	0.104
K_2PtCl_4 + Satd. 8.6 mM $Ce_2(SO_4)_3$	NAT 0.5 / 1:1	HEPES 6.8 / 7.1	20 / 5Hr (30 / 20 min)	M	10.7-2.8 2.8	5544	1.9	97.7	0.048	4.181	3.330	0.113
UDGpdTTT 1	DNACO n/a	n/a	n/a	86 M	9.8-2.5 2.5	8669	3.3	95.4	0.061	n/a	n/a	n/a
PCMBS	DNACO 1.0-10.0	HEPES 6.8	30 Hr tot. (14 Hr)	16 M	***NP							
PCMBS	DNACO 2.0	ddH_2O	12 Hr (1 Hr)	44 S0.88	10.7-2.8 2.8	4757	3.8	74.0	0.172	***NP		
Ethyl-Hg-PO_4	DNACO 1.0	ddH_2O	2 Hr (12 Hr)	34 M	11.2-3.0 3.0	3432	2.0	65.3	0.085	2.535	2.130	0.133
UDGpd^{t}UT	DNACO n/a	n/a	n/a	58 M	11.2-3.0 3.0	4528	2.6	86.2	0.049	3.685	3.096	0.102
UDGpd^{t}UT	DNACO n/a	n/a	n/a	107 M	8.9-2.2 2.2	10259	5.2	80.1	0.050	3.224	2.636	0.088
UDGpd^{t}UTT	DNACO n/a	n/a	n/a	81 M	11.2-3.0 3.0	3860	4.3	74.2	0.098	1.848	1.494	0.075
UDGpd^{t}UTT	DNACO n/a	n/a	n/a	85 M	8.4-2.5 2.5	8558	3.4	95.8	0.060	2.501	2.122	0.087
UDGpd^{t}U^{t}UT	DNACO n/a	n/a	n/a	57 M	11.2-3.0 3.0	3791	3.0	72.8	0.106	2.357	1.949	0.109

Table 6.3 (Continued . .)

X-Ray Diffraction Data for Native and Heavy Atom Soaked Crystals of uracil-DNA glycosylase in both the Free and DNA-bound forms.

See the note preceeding this table for a description of abbreviations and terminology.

1:2, in two separate tubes. The compounds used were as follows:

mercury(II) acetate
mercury(II) chloride
silver nitrate
potassium tetrachloroaurate(III)
potassium tetrachloroplatinate(IV)

The tubes were allowed to stand on ice for 10 minutes, whereupon it was observed that heavy precipitation had occurred in all cases. Each mixture was then clarified by passage through a 0.22 µm membrane in a microcentrifuge. The clarified protein/ metal solution was then used to set up 2 µl volume crystallisations using the microbatch method. All the mixtures yielded crystals of the usual trigonal morphology, though they were of diminished size. This was most severe with mercury(II) chloride, and with silver nitrate. The droplets prepared from mixtures using these compounds continued to precipitate after filter clarification. At the 1:2 protein to compound molar ratio there were no crystals at all in these droplets. In the 1:1 molar ratio droplets, the crystals were of sub-100 µm width dimension, and highly branched.

The droplets prepared using the other 3 compounds resulted in crystals from both ratios of protein to compound. There was a greater degree of precipitation in these droplets than is usual for HSV1 UDGase crystallisations, and the crystals exhibited a substantial amount of branching. There was at least one crystal in each case with dimensions close to 150 µm wide. Only one crystal was used for X-ray diffraction analysis. This was a crystal about 120 µm wide by 400 µm long from the droplet generated from a 1:2 molar ratio mixture of HSV1 UDGase and potassium tetrachloroaurate(III). The crystal diffracted tolerably, in terms of spots visible on the diffraction images, and intensity maxima from these images. However, the data was not processable due to an apparent unit cell change. This change was not defineable from the collected data, and in addition the crystal appeared to have degraded rapidly during collection, thus attempts at processing were abandoned.

It was concluded that the direct mixing of protein and heavy atom compounds in this manner was not a suitable method for derivatisation, as it resulted in heavy losses of protein in the formation of insoluble precipitate, and at least in one case also resulted in non-isomorphous crystals which were more prone to radiation damage than is normally

the case. A different approach to 'direct' introduction of heavy atoms into native HSV1 UDGase crystals was attempted by replacing part, or all, of the sodium sulphate in the stabilisation solution with a heavy atom sulphate. This was achieved by transferring fully grown crystals into a stabilising solution where:

a) all the sodium sulphate had been replaced with cerium sulphate, and
b) adding 2 mM cerium sulphate to the existing stabilisation solution. This latter case constituted a normal heavy atom soak.

In the first case, the result, an hour after the change of stabilising solution, was that the crystal had become pale white and opaque in appearance. When attempts were made to move the crystal, it was found that it had become a gel. This method was clearly not suitable for the total replacement of the entire sodium sulphate content of the stabilisation solution. In the second case, the crystal was still birefringent after 3 hours, and had a slightly milky appearance. The data obtained from X-ray diffraction analysis revealed no heavy atom sites were observable above the general background.

A full discussion of the strategies used in heavy atom derivatisation trials in this project, along with suggestions for future experiments may be found in section **7.2.4.**

6.3.2 HEAVY ATOM DERIVATIVES OBTAINED IN THIS PROJECT WITH CALCULATED SITE POSITIONS AND PHASING STATISTICS

The successful heavy atom derivatisation experiments resulted in the location of the heavy atom sites using the programs outlined in section **3.3.** In addition, graphical output of the calculated Patterson Difference maps was studied. The peaks corresponding to vectors between the heavy atom site and its symmetry related equivalents are found on the Harker Sections for spacegroup $P3_1$ in the case of unbound native HSV1 UDGase crystals, and $P2_12_12_1$ in the case of the DNA co-crystals. Significant peaks on these sections will be a strong indication that a heavy atom site is present. The Harker Sections of these Patterson Difference maps are shown in Figure **6.3.2.**

Table **6.3.2** shows the refinements (if any) carried out on the heavy atom sites. Sites in the DNA co-crystals have been refined, and have phasing power, and temperature factor values assigned to them. The electron density map for the DNA bound UDGase is interpretable only in a few places, thus the phase estimate still requires improvement.

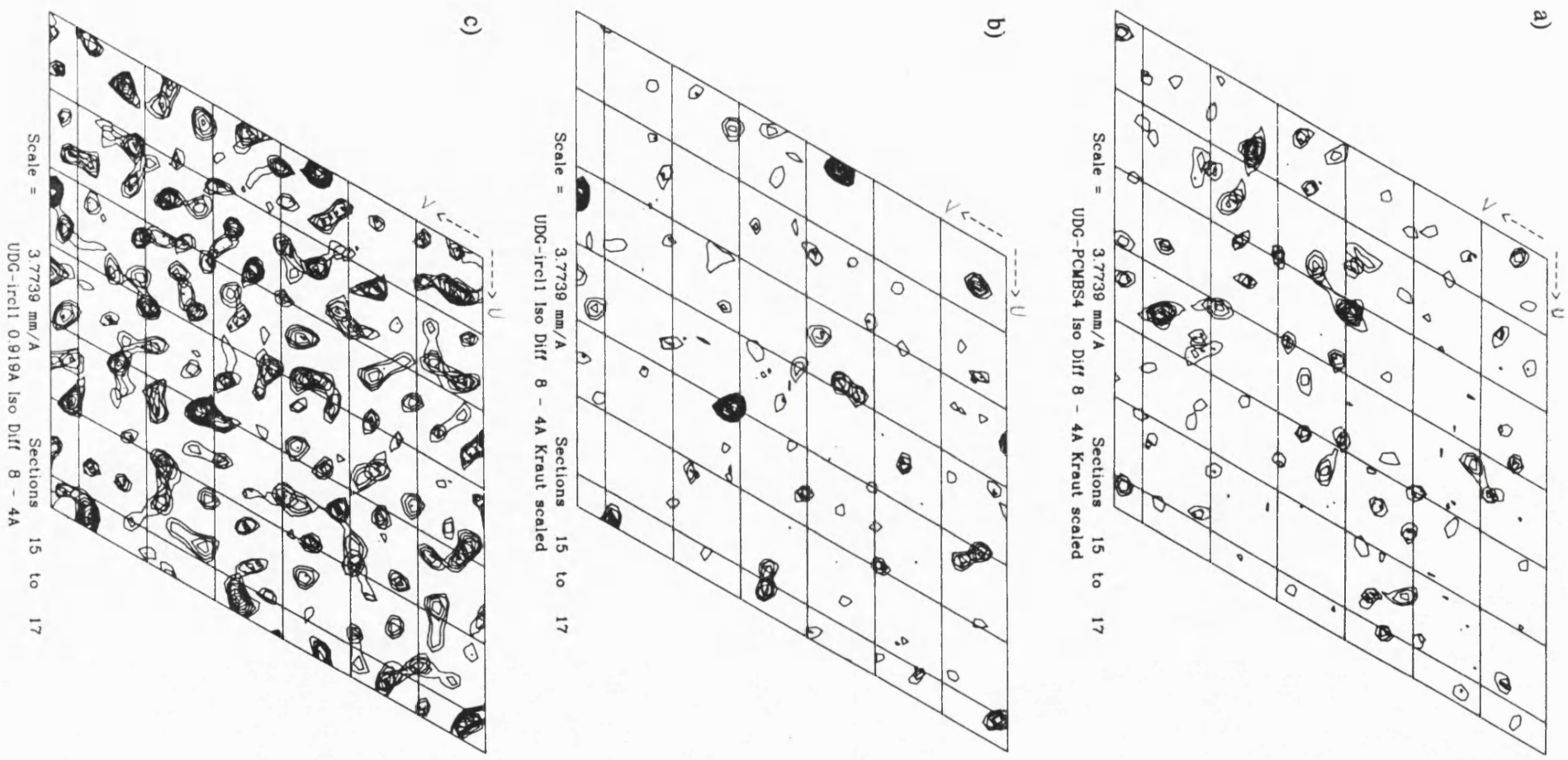


Figure 6.3.2

Harker Sections of Patterson Difference maps between the native data set and, a) PCMBS, b) iridium - CuKα, and c) iridium - 0.919 A, data sets, using unbound UDGase crystals.

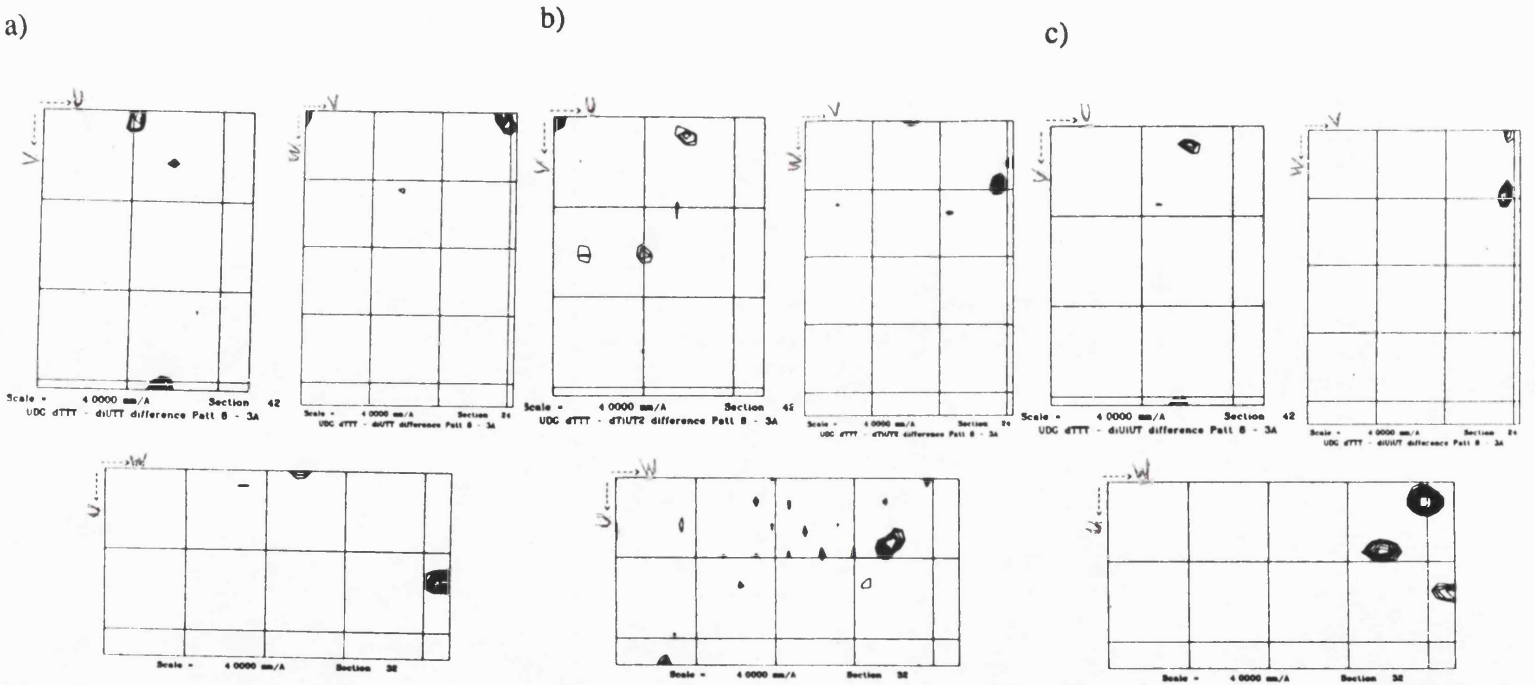


Figure 6.3.2 (Continued . .)
Harker Sections at 1/2 U, 1/2 V, and 1/2 W for the DNA co-crystals incorporating iodine. a) pd^{I}UTT, b) pdT^{I}UT, c) pd^{I}U^{I}UT

Atom Name oligonucleotide or compound	No. of Sites	Site co-ordinates (in Fraction of unit cell)			Occupancy	Anomalous Occupancy	B-Factor	Phasing Power (Cullis R Factor)		<FOM>
		x	y	z				Centric	Acentric	
Mercury PCMBS	1	0.932	0.743	0.990	12.716	4.088	30.000	NR	NR	NR
Iridium K_3IrCl_6	1	0.082	0.549	0.708	25.822	6.447	47.906	NR	NR	
Iodine pd^{I}UTT	1	0.355	0.488	0.009	9.931	7.930	32.931	0.7 (0.77)	1.0 (0.77)	0.5459
Iodine pdT^{I}UT	1	0.091	0.016	0.050	15.485	12.283	37.749	0.9 (0.65)	1.1 (0.73)	
Iodine pd^{I}U^{I}UT	2	0.354 0.094	0.488 0.015	0.009 0.050	11.088 16.362	9.780 13.426	33.061 44.392	1.1 (0.54)	1.5 (0.62)	

Table 6.3.2

Real space co-ordinates (in fractions of unit cell axes) of the heavy atoms obtained as isomorphous derivatives, and refinement and phasing statistics. The mercury and iridium sites were obtained with the native HSV1 UDGase unbound crystal form, and the iodine sites from the DNA bound co-crystal form. The iodouracil at the 5' end of the oligonucleotide appears to be held more tightly than the residue in the middle of the oligonucleotide, as judged by the relative temperature factors.

x, y, z refer to the real space unit cell axes. The B-factor is the temperature factor.

6.4 PRELIMINARY MODEL BUILDING OF URACIL-DNA GLYCOSYLASE

The phase information obtained from the iodinated oligonucleotides co-crystallised with HSV1 UDGase has been used to construct a preliminary electron density map. Some secondary structure features are visible, though the map is very discontinuous (Section **7.2.4**). However, a very clear solvent boundary between the molecules is visible, and this enables the relative positions of each molecule in the unit cell to be seen. In addition, a region of density which fits the C-terminal 10 residues has been modelled. A portion of α-helix in the density, and the skeletonised view of molecules of HSV1 UDGase around a superimposed unit cell boundary (looking down the c-axis) are shown in figure **6.4**.

Currently, the phases are being improved by collection of more complete data sets and the collection of data from brominated oligonucleotide co-crystals. The map is much improved in appearance, and more secondary structure features are visible.

i) ii)

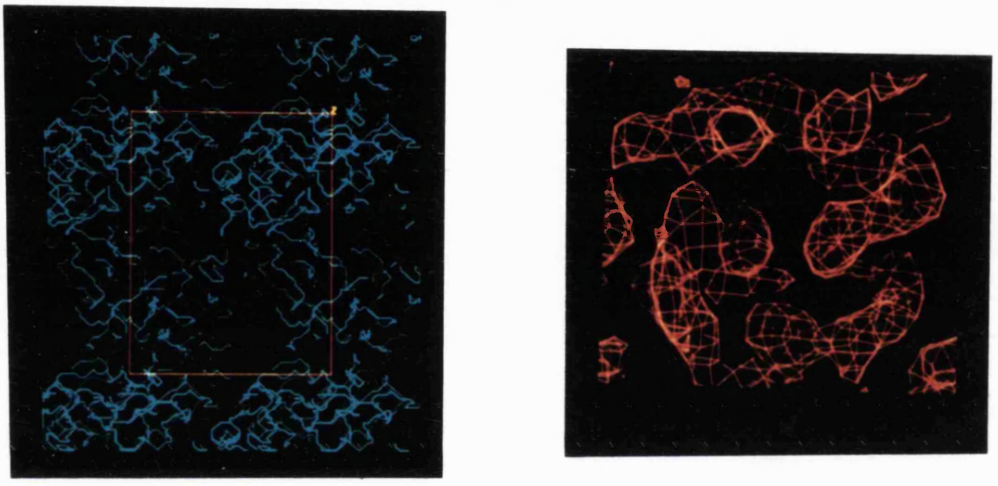


Figure 6.4

i) View of skeletonised molecules down the c-axis, shown with a superimposed unit cell boundary. The solvent space around the molecules is clearly defined.

ii) A preliminary 3 Å electron density map reveals a few fragmented secondary structure details. A piece of density with the shape, dimensions, and pitch of a typical α-helix is shown here.

7 GENERAL DISCUSSION

This project was undertaken with the aim of understanding, at the molecular level, the basis of the specificity and efficiency of a highly conserved enzymatic mechanism. This reaction is carried out in all organisms with DNA genomes by a highly conserved group of enzymes, the UDGases (Section **1.2**). In this final chapter the results obtained in the course of the project will be discussed with respect to several criteria. These are, the methods used to investigate the problem, the conclusions which may be drawn from the results obtained, and the direction in which the project could be extended to further the knowledge of how molecular structure dictates function. As a prelude, a discussion of how structural studies may complement and enhance biochemical studies of UDGases carried out in the past, will be presented.

7.1 STRUCTURAL STUDIES OF HSV1 URACIL-DNA GLYCOSYLASE CAN REVEAL THE MOLECULAR BASIS FOR KNOWN BIOCHEMICAL DATA

UDGases have been extensively studied with regard to their biochemical properties; this has been reviewed in chapter **1**. These studies have revealed that the UDGases, as a class, are highly invariant in terms of their specificity for uracil in DNA, and thus the manner in which it is removed. This inference can be made because of a number of similarities. Firstly, it has been shown that the rate of uracil removal from single stranded DNA is tenfold faster than the rate of uracil removal from double stranded DNA for all of these enzymes thus far studied. Secondly, a very tightly binding, and highly specific protein inhibitor of the *B.subtilis* UDGase is an equally potent inhibitor for any other UDGase thus far incubated with it. In addition, studies where UDGases from more than one organism have been directly compared have shown that the kinetics for all the enzymes are very similar. These findings, taken together with the high degree of conservation at the amino acid level, indicate that extrapolation of any one result to the entire group is unlikely to be far in error. This will almost certainly be found to be the case with the tertiary structure (Figure **7.1**).

In the light of the ever increasing number of solved protein structures, it is now known that even distant sequence relationships can result in structural similarities. This

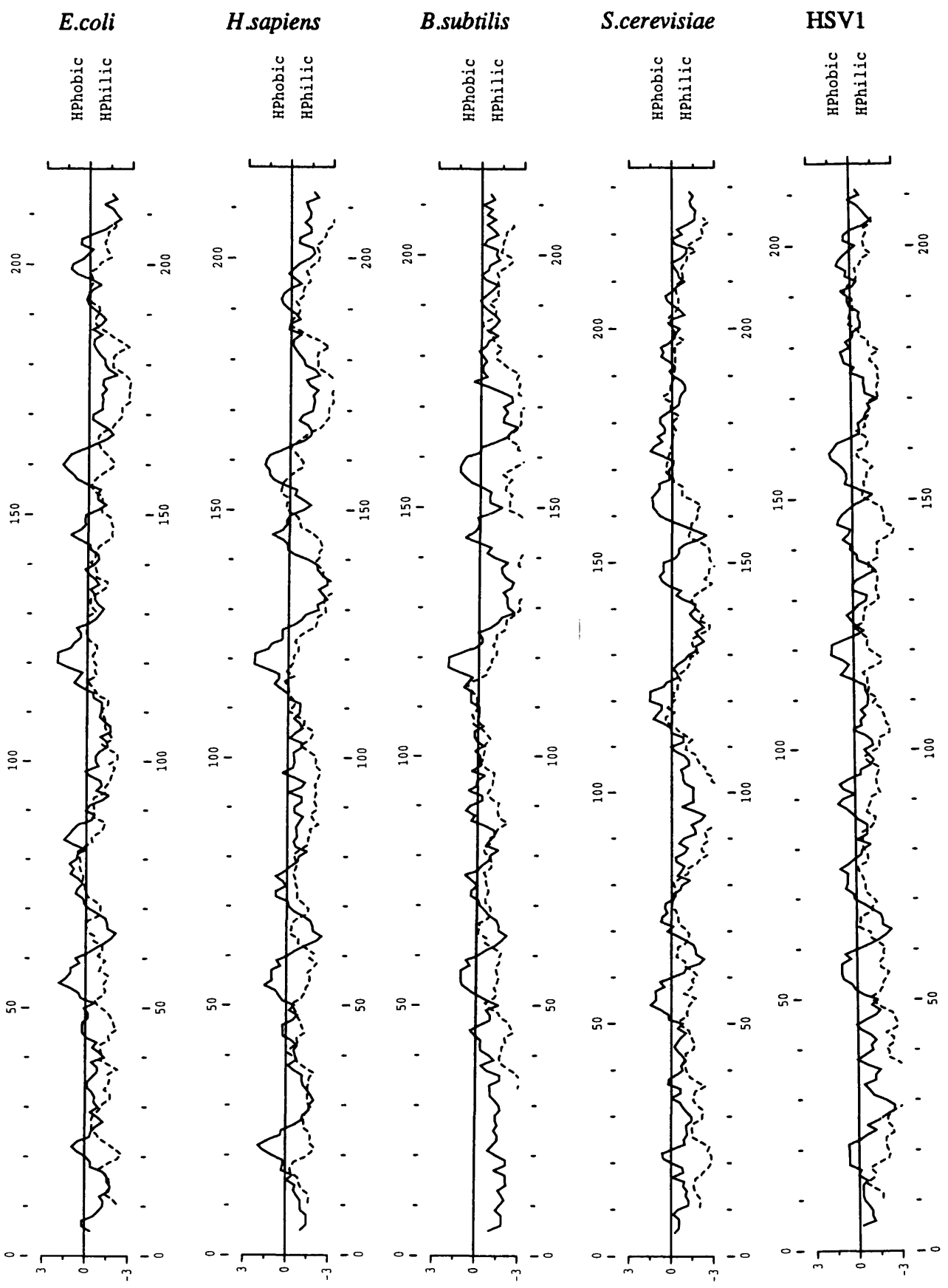


Figure 7.1

Comparison of hydropathy/ hydrophilicity profiles, spanning residues within 117-351 of the UDGase sequence alignment of figure **1.2.2.** The similar profiles indicate that the tertiary structures of these enzymes are also likely to be very similar. The profiles were calculated and plotted using the PEPPLOT program in the GCG package.

may often be traced to a motif which is able to carry out the enzymatic function. In such cases the rest of the structure may only bear a loose resemblance, and there may be little in the way of immediately obvious sequence conservation at the amino acid level. Yet upon solution of the tertiary structure, the link in function is seen to be clearly linked to the motif or domain features. In the case of the main class of UDGases the tertiary structures are likely to be as easily superimposable as the primary structures. The solution of one UDGase structure could very easily be extrapolated to the rest of the group. The most interesting point in the solution of the structure of these enzymes will be to identify the residues, motifs, or domains, responsible for the highly selective reaction mechanism. This mechanism is seemingly also reproduced in UDGase enzymes which share no obvious sequence homology with the main class. In this case, it is likely that the tertiary structures will not necessarily be directly superimposable, but they may well be strikingly similar in comparison. It will be particularly interesting to compare the structure of HSV1 UDGase with that of human G3PDase. The latter has been crystallised, though a structure has not yet been reported.

The HSV1 UDGase has been studied in this project with the primary purpose of understanding how the function of the main class of UDGases is linked to their molecular structures. Much kinetic data is available on the enzyme from HSV1 isolated from eukaryotic cell culture, but it is unclear whether this enzyme is identical to that produced for structural studies in this project. The possibility exists that the HSV1 UDGase produced in eukaryotic cells is the product of translational initiation from the first UL2 ATG, at 9886 in HSV1 (Figure **1.3.2**), and that this enzyme retains an N-terminal leader sequence (Section **7.1.2**). The likelihood is that the difference in the function of these two enzymes is minimal. Other cases where such differences exist, for example between nuclear, and mitochondrially, localised forms of a UDGase, result in discernable, yet marginal, differences in rates of uracil removal. The fact that UDGases from prokaryotic sources are often devoid of N-terminal sequences extending far beyond the homologous core region of this class of enzyme, and yet display rapid and accurate uracil-removing ability seems to suggest that the N-terminal sequence is no more than a localisation signal. This has certainly been shown to be the case with the main class human UDGase (Slupphaug *et al.*, 1993). In addition, the enzyme studied in this thesis can be shown to be active with and without an additional 40 amino acid fusion at its N-terminus (Section **4.2.4**, Section **4.2.5**), suggesting that the larger enzymes will have the same fundamental

tertiary structure (Section 7.3) and that the N-terminus can be highly variable.

The extrapolation of structure to function can then be easily made, and when other UDGase structures become available, the differences in kinetics may become explainable in terms of molecular structure differences. The structure of HSV1 UDGase is not yet solved, but the quality of the phase estimation needs only a small contribution from a further isomorphous heavy atom derivative to change this situation.

The structural data for HSV1 UDGase will show not only the native enzyme structure, but also the structure of the enzyme in complex with a single stranded DNA oligomer. This data will be the first glimpse of the residues and motifs involved in DNA binding and substrate recognition. The specificities of the *E.coli* enzyme for different lengths of DNA oligomer have been studied, and in addition, non-specific scanning of DNA has been shown to occur. It is likely that the DNA binding site, and the orientation of a DNA oligomer in this site, will be revealed. This information will enable an understanding of the preference of the enzyme for single stranded DNA, the reason for the type of scanning motion and difference in kinetics on different length substrates, and possibly the reason that the enzyme only removes uracil, and not the structurally similar thymine.

7.1.1 HSV1 URACIL-DNA GLYCOSYLASE MAY BE ESSENTIAL FOR THE MAINTENANCE OF THE VIRAL GENOME DURING LATENCY

Very recently, new evidence has emerged which implies that the HSV1 UDGase is essential in maintaining the fidelity of viral DNA during latency (Pyles and Thompson, 1994). It was observed that following a hyperthermic shock to murine neuronal cells harbouring latent viral genomes, a large decrease in reactivation levels was observed in mutants with disrupted UL2 ORFs. This was not the case in cells in which the mutants were restored to wild type. The mutant and wild-type viruses have similar replication efficiencies during virulent infection (Mullaney *et al.*, 1989), but it appears that the mutant virus replicates at a diminished level in neuronal latency (Pyles and Thompson, 1994). This evidence is strongly supportive of the hypothesis that the viral UDGase has an essential role in restoring the viral genome prior to reactivation of the virus from latency. It also indicates that the viral UDGase may have an essential role in the replication of the viral genome during latency in neuronal cells.

If this is so, then it would appear that some other regulatory elements exist in the viral genome. Thus, apart from UL2 being an early β-gene, it is also a 'latency-on' gene. It can be envisaged that many ORFs in the HSV1 genome might be 'latency-on' type genes, and that these genes could be activated at different stages in latency. These stages might be designated 'latency-immediate early', 'latency-on', and 'reactivation'. Thus from the data of Pyles and Thompson discussed above, the HSV1 UDGase would most likely be latency-on/ reactivation, in the latent phase of the viral lifecycle. This means, as the data suggests, that it is probably required for replication during latency, and for final proofreading prior to reactivation. This latter role may be indistinguishable from the former, since it is known that reactivation of the virus from the latent state is preceded by large-scale replication of the viral genome (Roizman and Sears, 1991). The way in which the latent phase of the viral lifecycle is regulated has yet to be fully understood.

It is also possible that the viral UDGase may be present only because such an activity is required at a precise time in the replicative cycle. This has been shown not to be the case (Mullaney *et al.*, 1989), though in latency this enzyme seems to be associated with replication. This is akin to the cell cycle regulation of the human uracil-DNA glycosylases, with the peak activity being associated with replication. This may indicate that misincorporation of uracil is a major problem in DNA replication, but may just as easily indicate that pro-mutagenic cytosine deamination is an ever-present danger to the fidelity of a DNA genome. It is possible that inhibitors directed against HSV1 UDGase might enable sufficient mutagenic damage to amass in the viral genome to render it unviable (Focher *et al.*, 1992a; Focher *et al.*, 1993). Compounds which selectively inactivate the HSV1 UDGase have been developed (Focher *et al.*, 1993; Botta *et al.*, 1994), and their specific interaction with HSV1 UDGase on a structural level is being investigated (Section **7.3**).

A final question that could be asked is whether the neuronal cells of individuals with latent herpes simplex infections are less prone to aging effects due to the presence of the viral UDGase activity. If this is the case, then gene-therapy to negate the effects of senility in neuronal cells could be achieved with a constitutively expressing UDGase gene. The possibility exists, when the biology of latency is better understood, that HSV1 could be modified so that it is the vehicle for such gene transfer.

7.1.2 THE LIKELY START CODON FOR THE UL2 ORF ENCODING HSV1 URACIL-DNA GLYCOSYLASE

The UL2 ORF has been assigned co-ordinates of 9886-10888 (TAA stop codon is 10887-10890) in the HSV1 genome (McGeogh *et al.*, 1988). SDS-PAGE analysis of *in vitro* translation of the analogous UL2 ORF in HSV2 (Caradonna *et al.*, 1987) indicate that the 5'-most start codon (corresponding to 9886 in HSV1) is the likely start point of translation, though shorter RNA transcripts are detectable, and other translation products of lower molecular mass are produced. These products are of similar molecular mass to those expected with translation from start codons further downstream from the putative start codon.

In this project, the first possible start codon at 9886 in HSV1 was not found to produce a new protein in an *E.coli* expression system, but the next possible start codon at 10156 did give rise to an active uracil-DNA glycosylase in *E.coli*. It is known that the analogous second start codon in HSV2 does not exist (McGeogh *et al.*, 1991), but that a second possible start codon is found 11 codons upstream from this position. The analogous position in HSV1, 10123, is an ACG codon which may still function as a start codon (McGeogh *et al.*, 1991). The amino acid sequence homology with other main class UDGases begins close to an isoleucine, encoded at 10237 in HSV1. The ACG at 10123 in HSV1, and the analogous ATG in HSV2 are both downstream of the respective UL1 ORFs. However, the ATG at 9886 in HSV1 is found 244 nucleotides into the 3' end of the UL1 ORF. It is known that overlapping ORFs exist in HSV1, and so this does not rule out the possibility of this ATG being the start codon of UL2.

The possibility that more than one transcript is produced that can encode a uracil-DNA glycosylase (Caradonna *et al.*, 1987), makes it reasonable to assume that the expression of this protein could be regulated by more than one promoter. This would result in UDGase proteins of different molecular masses. It is possible that the N-terminal extension may not be required in a particular part of the viral life-cycle, for example during infectivity or alternatively, during latency. Thus the form of UDGase expressed during latency might be different from that produced in the replicative cycle, and the difference would be that a different start codon is preferentially employed for translation. Thus the difference could be analogous to the nuclear and mitochondrial forms of human UDGase (Slupphaug *et al.*, 1993), where the N-terminal sequence could be a localisation

signal. Alternatively, it could be a sequence that allows interaction with other proteins. In the case of HSV1, the N-terminal sequence encoded from 9886-10156 is particularly proline rich, a common feature of protein-protein interface regions (Williamson, 1994).

An experiment to determine the molecular mass of the UDGase expressed in infected eukaryotic cells was conducted by raising polyclonal antibodies to the protein overexpressed in this project. The result was inconclusive however, because the antibodies cross-reacted with a number of other proteins. This could have been overcome by blocking the antibodies which cross-reacted with a control sample prior to incubation with an infected cell lysate. This is not ideal however, and is not always reliable due to discrepancies between infected cell lysates and control lysates. A much more workable method would have been to raise monoclonal antibodies. However, this is more expensive and time was not available to do so by the time the results of the polyclonal antibody data were obtained. It was hoped the result would reveal the molecular masses of proteins in the infected cell lysates which cross-reacted with an antibody specific for the UDGase produced in *E.coli* in this project.

7.2 AN APPRAISAL OF THE METHODS USED TO STUDY HSV1 URACIL-DNA GLYCOSYLASE IN THIS PROJECT, AND THE CONCLUSIONS WHICH MAY BE DRAWN FROM THE RESULTS PRESENTED

In order for a molecule to be studied at the structural level by crystallography, it must be abundantly available in a very pure and homogeneous form. If the structural study is aimed at understanding the function in terms of the structure, then the molecule must be in an active conformation at the very least, and preferably fully active. The HSV1 UDGase was prepared in *E.coli* cells with intention of expressing it to high levels for such studies. Scanning densitometry of SDS polyacrylamide gels on which a crude cell lysate (of cells grown as described in section **4.3.2**) has been run shows that HSV1 UDGase can account for between 9 and 15 % of total *E.coli* cell protein. In addition, this protein can be shown to be fully active with the assay method used, up to 9 months after purification. Specific UDGase activity has been recovered from partially purified samples stored at 4°C for over 2 years (results not shown), though the levels of this activity are very low.

7.2.1 THE ACTIVITY ASSAY FOR URACIL-DNA GLYCOSYLASE

Detailed kinetics have appeared for UDGases, using base-release methods similar to the one employed in this project and also by fluorimetry. These substrates have tended to be composed of calf-thymus or bacteriophage DNA randomly labelled with tritiated uracil in the case of base-release assays, or of closed circular DNA containing incorporated uracil, or deaminated cytosine, for fluorimetric assays. These assays have been used with small quantities (typically nanograms) of enzyme to obtain steady state kinetics for UDGases.

In this project, the substrate was made primarily in order to track the active fractions of UDGase during purification. For this purpose it proves to be very sensitive. However, the assay appears to be unsuitable for kinetic studies. The problem encountered is that steady state kinetics are not reproducibly attainable. Typically, if total protein is diluted with a stabilising mixture (Section **2.3.10**) at stages in the purification prior to elution from the Affigel Blue column, then a dilution can be reached at which it appears that steady state kinetics are obtained. Once the elution from Affigel Blue has been carried out however, the enzyme cannot be diluted to a level at which steady state kinetics result. The problems are as follows: firstly, if the enzyme is diluted to picogram levels, then it still appears to turn the substrate over non-linearly before 1 minute has elapsed. This is an apparent 'burst' effect. Secondly, at these levels a consistent dilution is difficult to achieve in order to repeat the assays. The problems here are that the margins of error between what should be identical assays are unacceptably wide.

A single dilution of enough enzyme to carry out several assays in series is not possible because the enzyme activity diminishes rapidly over a period of minutes at such dilutions. In order to attempt to get around this problem, up to 8 assays have been run in parallel, staggered at 30 second intervals. The kinetics are always non-linear. There are two main possibilities. Firstly that the substrate is present in sufficient concentration for the enzyme to become end product inhibited during the assay. Secondly that the enzyme activity is declining in the assay due to its apparent low stability at very low concentrations. The fact that impure preparations of the enzyme can be seen to display steady state kinetics indicates that end product inhibition during the course of an assay is an unlikely reason for the behaviour of more pure preparations. Also the levels of uracil present in the reaction after non-linear kinetics have been observed should not be

sufficient to inhibit the reaction. It is clear that the enzyme is not very stable at low concentrations, even in the presence of a stabilising BSA containing buffer. For these reasons, a complementary kinetic study of the HSV1 UDGase was not performed. However there is no doubt that the enzyme which is being crystallised is an active UDGase, and that it is obtained in a very pure form at high levels.

7.2.2 THE PURIFIED HSV1 URACIL-DNA GLYCOSYLASE IS CORRECTLY FOLDED AND IN A HOMOGENEOUS STATE

The purified UDGase appears to be at a level of near homogeneity. The protein can be crystallised, and these crystals are of suitable quality for X-ray diffraction studies. This shows that the HSV1 UDGase produced in this project is homogeneous following purification. The batches produced by the latest modifications to the purification protocol are particularly stable, and can be used for crystallisation for at least 9 months after purification. The quality of the crystals produced is not seen to diminish during this time. The modifications carried out to the purification protocol included the complete re-assessment of the likely causes of instability on long term storage, and the likely nature of contaminants observed after the original purification protocol. These are discussed in section **4.3.2**, along with the reasons for selecting the various purification steps.

It is likely that HSV1 UDGase contains one or more protease sensitive sequences, since the change of bacterial growth strain to a less proteolytically active one, and the incorporation of protease and microbial growth inhibitors in purification and storage buffers resulted in a substantial increase in stability, and a significant loss of contaminating bands on SDS PAGE analysis. It was also found that glycerol, although known to act as a stabilising medium due to its interaction with the surface of proteins, was found to decrease the useful storage time of the protein at 4°C. The exact reason for this is unclear, though the type of instability observed was that the protein could not be concentrated to very high levels, and would precipitate out over the course of weeks. This type of effect could be due to the glycerol replacing surface water molecules on the protein, and altering its solubility trait.

Once the protein has been crystallised, it appears that proteolysis can take place within the crystal. This is apparent on SDS PAGE analysis of dissolved crystals. However, an observable change in the appearance of the crystals is not apparent, and so the

proteolytic cleavage is either not in a part of the protein essential in the packing of the crystal, or the cleavage products are still stably maintained in their respective packed positions in the lattice. The latter might be expected to result in a degree of disorder in the crystal, and thus affect the resolution limit of diffraction. This type of effect is observed with crystals which are several months old. The diffraction limit is not dramatically affected, but the degree of spreading in the spots (the 'mosaicity', indicating slight perturbations in the lattice) increases, and the general appearance of the diffraction images is less clean.

To test that the protein is correctly folded, CD analysis was carried out. This indicated that the protein contained an ordered secondary structure signature, and was likely to have substantial α-helical content. This information, taken together with the fact that the enzyme appears to be an active UDGase, in a highly homogeneous state, indicates that the recombinant HSV1 UDGase is correctly folded.

7.2.3 CIRCULAR DICHROISM STUDIES OF URACIL-DNA GLYCOSYLASE

In an attempt to gain a preliminary insight into the interaction of UDGases with DNA and uracil, CD analysis was carried out. As explained in section **6.1**, this should show up any changes in conformation induced by binding. The results showed that there was a slight change in the conformation of DNA when UDGase binds. The mode of binding is likely to be electrostatic, with minor unwinding effects. The CD spectrum of single stranded DNA showed less apparent change in signature upon UDGase binding. This does not enable any insight as to how the enzyme identifies uracil in DNA. The DNA molecules used in the CD analysis did not contain uracil, and so it is possible that major changes in CD would occur if such DNA were used. The possible modes of detection of uracil in DNA are:

1) The enzyme unwinds double stranded DNA as it travels along it, and the bases pass through the active site. The detection and removal of uracil occurs here. For example, the passing of a pyrimidine through the enzyme active site might cause certain interactions with active site residues, but the N-glycosyl bond may not be cleaved unless the base is uracil. This may be because the other pyrimidines sterically interfere with the key residues in the active site.

2) The enzyme scans DNA processively, and flips each base out of the helix in turn into the active site, where a process similar to the one just mentioned occurs. This type of action would be akin to the mechanism of base access used by the *Hha* I cytosine-5 methyltransferase, though that enzyme only flips out one base in a specific recognition sequence (Klimasauskas *et al.*, 1994).

3) The enzyme scans the major groove. When uracil is detected (probably by detecting the absence of a methyl group at the 5 position of the pyrimidine ring, which protrudes into the major groove in thymine), the base is flipped out, or the helix is unwound, and the base is passed through the active site where it is removed.

4) The enzyme detects any backbone distortions caused by the presence of uracil. This is not very likely to be the case, since uracil is not a bulky lesion.

In the first case, there would be a definite change in the CD spectrum of both double and single stranded DNA, since the bases would become unstacked by the action of the enzyme, and separated in order to pass through the enzyme active site. In the active site environment, various induced CD changes might take place due to the change in the local environment of a base. The enzyme CD might be expected to alter slightly since the DNA comes in very close contact with the protein at the active site, and induced CD changes might again be expected.

In the second case, the CD spectrum of the DNA would again be expected to change, since unstacking of a single base would have to occur to enable processing of the base by the enzyme. The enzyme might be expected to undergo some sort of conformational change, in order to expose the active site, and stabilise the DNA structure.

In the third case, the bases can all be identified easily without significant distortion of the DNA structure. Thus, this is the most likely mode of action of UDGase in the light of the CD data. However, no firm conclusions can be made, because, the CD of a uracil containing DNA was not studied. The preference of UDGase for single stranded DNA can only be explained if:

a) The entire strand being read is processed (as in 1 and 2 above).
b) The uracil, once detected, must be acted upon by another part of the molecule (compatible with 3 above).

If major groove scanning is the mode by which uracil is detected, then option b

above would explain the difference in rates of uracil removal from single and double stranded DNA. If a double stranded substrate is encountered, then the base must be unstacked prior to its removal from the molecule. In the case of a single stranded substrate, this does not need to be done, and the reaction may proceed more rapidly.

Though the CD signature for the trimeric single stranded deoxythymidine oligonucleotide indicates that it has a stacked conformation, the preliminary structural data obtained from the initial electron density map of HSV1 UDGase indicates that the bases in this oligonucleotide are unstacked when bound to the enzyme. A distance of 12.2 Å separates the first and second thymine methyl groups (identified by location of the iodine atoms which isomorphously replace them in the derivative crystals), when the stacked distances should be about 4.9 Å. It appears that one of these methyl groups is completely flipped out of the general stacking plane. This may indicate that the enzyme is primed to act at every base in a single stranded substrate. That is, the bases pass through the active site one by one, and uracil may be acted upon instantly. On the other hand, in the case of double stranded DNA, the major groove may pass the active site, and uracil may still be detected, but it has to be flipped into the active site before it can be processed.

An alternative mechanism, is that uracil is detected by a group of residues adjacent to the active site, and that local unwinding, or flipping of the base out of the helix, may occur at this point, with the base then passing through the active site as the enzyme scans along the DNA. The elucidation of the mechanism will have to await further studies (Section 7.3) however, since the lack of any appreciable CD signal change upon binding of UDGase to apparently stacked single stranded oligonucleotides appears to be in conflict with the preliminary crystallographic data.

The CD analysis of the effect of free uracil on the enzyme did yield interesting data. There was an apparent enhancement of the CD signal of the UDGase, especially in the aromatic region of the spectrum upon addition of 100 µM uracil. This enhancement was unchanged at 250 µM. There are two possible explanations for this effect. Firstly, the uracil, which has no CD signal of its own, develops an induced CD from non-specific interactions with the protein. Secondly, the uracil is bound by the enzyme in a specific, and biologically significant manner by the aromatic residues of the enzyme to cause an induced CD of the uracil, and/ or an induced enhancement of the CD of one or more aromatic residues in the protein. If the second effect is true, then crystals which have been soaked with uracil should contain a modified UDGase structure. This certainly appears to

be likely, since native UDGase crystals soaked with uracil appear to shrink slightly in width, and develop definite and visible striations. UDGase crystals grown in uracil are either of a different morphology, or have many minor outgrowths on the surface. Thus, some definite interaction appears to be taking place between uracil and UDGase. Data sets from suitably sized uracil/ UDGase crystals will shortly be collected.

7.2.4 CONCLUSIONS CONCERNING THE CRYSTALLISATION OF HSV1 URACIL-DNA GLYCOSYLASE, AND DERIVATISATION EXPERIMENTS WITH HEAVY ATOM CONTAINING COMPOUNDS

The HSV1 UDGase, without bound DNA, has been shown to crystallise under a number of conditions, and the crystal form chosen for this study is found to have an absolute requirement for sulphate ions as part of the crystal lattice (Chapter 5). The sulphate ions are likely to bridge adjacent molecules in the lattice by ionic linkages. The requirement for PEG in the crystallisation is probably due to the fact that this molecule will compete for water with the protein and cause precipitation. In the absence of PEG, the crystals will slowly dissolve. The reason why phosphate or imidazole sulphate buffers seem to produce good single crystals is likely to be because they do not interfere with the interaction of sulphate ions in the lattice building process. This is probably the problem encountered when trying to grow crystals in the presence of sulphonic acid buffers. Once crystal growth is complete however, the crystals appear to be stable in a wide range of buffers for at least a month, in the pH range 6.0 to 9.5. This is judged by maintenance of birefringence of the crystals. It has been a general observation with the native crystals that loss of birefringence is accompanied by loss of measurable X-ray diffraction.

A number of different buffers of varying pH have been used in the stabilisation solution in order to carry out heavy atom compound reactions by soaking the crystals. This is because various amino acid side chains are ionised to different extents at different pH values, and the potency of certain compounds is increased or diminished to varying extents depending on the type of buffer used (Petsko, 1985). The use of mercury containing compounds was the first choice, because these compounds react preferentially with cysteine (and also with histidine, and to a lesser extent with other amino acids at pH values above 9), and the recombinant HSV1 UDGase contains six cysteine residues, none of which is conserved overall in the main group of UDGases. This is important, because

most amino acid sequence substitutions in highly conserved enzyme groups occur in loop regions rather than buried regions which usually constitute the invariant core of the enzyme. Thus it was thought likely that the cysteines would be exposed, and able to react with mercury and its compounds.

The initial experiments looked promising, with the compound PCMBS showing a weak site in Patterson maps from two data sets. The isomorphous differences were not big however, thus more derivatives were required. Attempts to repeat the PCMBS interaction for collection of optimised anomalous scattering data at the synchrotron were not successful however. From the effects of other mercury compounds which were used, it was clear that small, penetrating compounds, such as ethylmercury phosphate were reacting too vigorously, and causing loss of birefringence, and that the bulkier compounds such as PCMBS were not reacting very strongly, if at all. The same sort of effect was found for compounds that are known to interact predominantly with histidine groups (Blundell and Johnson, 1976), of which there are 13 in the recombinant HSV1 UDGase. The compounds used were silver nitrate, potassium tetrachloroaurate(III), and potassium tetrachloroplatinate(IV). The latter compound is also used to react with methionine residues, of which there are four in the recombinant HSV1 UDGase.

Two main conclusions can be drawn from this data. Firstly, it is possible that there are one or more cysteine or histidine residues involved in the packing interfaces of the crystal lattice, making them inaccessible to large compounds, but causing disruption of the lattice if a small reactive species preferentially reacts with them. The second possibility is that the reactive residues are not directly involved in the lattice contacts, but are placed very close to adjacent molecules such that binding of a heavy atom species will cause disruption of the lattice.

Other reactive compounds were also used. These compounds were mainly lanthanide compounds containing cerium, samarium, europium, and uranium, and transition metal compounds containing tungsten, rhenium, osmium and iridium. The full list of compounds searched is given in table **6.3**. A further isomorphous derivative was obtained with potassium hexachloroiridate(III). This produced one strong site in a Patterson map. The same crystal was used to collect a second data set at the synchrotron, to make use of optimised anomalous scattering using a wavelength just below the L_I absorbance edge of iridium. Cross phasing of these two iridium data sets with strongest PCMBS derivative recovers one mercury, and one iridium site. The phases are not good

enough to generate an interpretable electron density map, and so further isomorphous heavy atom derivatives will be required.

The HSV1 UDGase has also been crystallised with a single stranded oligonucleotide, and a preliminary electron density map has been produced, using phases from oligonucleotides which have iodine atoms replacing the 5-methyl group on the thymine pyrimidine ring (Section **5.4.2**). Certain parts of the density are recognisable as secondary structure features, but the map is very fragmented and noisy. This is mainly due to incompleteness of the data sets, and the inherently weak phasing power of iodine. The incompleteness means that certain areas of reciprocal space have not been observed due to the alignment of the crystal being near-perfectly set along one axis with respect to the X-ray beam.

The UDGase/ oligonucleotide co-crystals display very strange stability characteristics. If the crystals are removed from the droplet in which they grew, and placed in a solution which should have the same ratio of precipitating agents as were present in the growth droplet, then the crystals reduce to bundles of fine needles. If the precipitants are made progressively weaker, the same effect occurs. It appears that the crystals are only stable in deionised water, and in this lasts for only a couple of days. The long term storage of these crystals can be achieved by dilution of the growth droplet with deionised water. The ratio of water volume to the original droplet volume does not appear to be critical.

7.2.5 CLONING AND EXPRESSION PROBLEMS ENCOUNTERED DURING THE PROJECT

At the outset of the project, a great deal of difficulty was encountered in attempting to clone the PCR product of UL2 (9886-10890) into the expression vector plasmid pET3a. The fault was eventually inferred to be in the non-coding strand PCR primer. The way in which this was done was firstly, to observe the extent of restriction digestion of both ends of the PCR product, and secondly, to attempt to clone one, or both halves of the gene independently. When the end labelling experiment was carried out, in order to check the extent of digestion of the ends of the PCR product with *Bam*H I, it was not clear whether digestion was only 50 %, or whether it was much more complete. The cloning of the 5' end of the UL2 ORF PCR product confirmed that the coding strand

primer did specify a *Bam*H I site, and sequencing of this end of the PCR product proved this to be correct, and also showed that *Taq* DNA polymerase, which had been used for that particular PCR had not incorporated any incorrect bases.

It is now known that the ATG at 10156 in HSV1 is the start codon for a functional uracil-DNA glycosylase which can be expressed to high levels in *E.coli*. However, the potential start codon at 9886 may well be functional in eukaryotic cells. However, it is also known that the HSV1 UL1 ORF ends at 10013 in HSV1, and thus would overlap the potential first start codon at 9886. This does not mean that the real start codon for UL2 must be at 10156. The constructs pHS5 and pAH1 did not appear to produce any new protein at all, but this may be due to codon preference in *E.coli*. The product of an ORF starting at 9886 would likely be a UDGase with an extended N-terminus.

The other expression problem encountered in this project was with the overexpression of the *E.coli* UDGase. This protein has been overexpressed in the past (Bennett and Mosbaugh, 1992), as a soluble protein, and purified by conventional methods. In this project however, it proved to be insoluble, even though it was produced to very high levels. The same was true when the gene was cloned into pRSET B (results not discussed, but protein product shown in figure **4.3.3**). The refolded and purified preparation (from cells directing the expression of the protein from pEU234·2) was seen to be the only visible band on a Coomassie Brilliant Blue stained SDS polyacrylamide gel. However, upon attempting crystallisation, the appearance of the sparse-matrix survey precipitates indicated that the protein was not of a homogeneous nature. The precipitates were all virtually identical in appearance, showing a fine fibrous appearance within seconds of being set up. The trace of the elution from the poly-U sepharose column used to isolate the refolded protein did not return smoothly to the baseline, suggesting that more than one species was present. However, upon concentrating the sample, there was a lot of precipitation, thus many of the most unstable conformers may have been lost before the crystallisation was attempted.

The PCR of the *ung* gene was carried out with Vent® DNA polymerase, which has a 3'→5' proofreading exonuclease function. The PCR product was not sequenced, and there is a possibility that the sequence was in error leading to the insolubility problem. The PCR has since been repeated with a much shorter extension time of 45 seconds, as opposed to the original 5 minutes. This should ensure that the polymerase does not chew back and fill in several times, as it may do if the extension cycle is too long. Longer than

necessary extension times with Vent® polymerase are not recommended in the accompanying literature. This new PCR product has been cloned into pRSET B, but has not yet been characterised for expression in BL21(DE3).

7.3 FUTURE DIRECTIONS FOR STRUCTURAL STUDIES OF URACIL-DNA GLYCOSYLASES

The work in this thesis has not resulted in any firm answers to the original question which was asked, but has succeeded in producing a system which needs only a small amount of further work to provide these answers. The UDGase/ oligonucleotide co-crystals have now also been grown with oligomers containing bromine instead of the thymine methyl group, and optimised anomalous scattering below the K absorption edge of bromine was recently carried out at the synchrotron, to a resolution of 2.3 Å. The phases were extended to the point that some more secondary structure has become visible in the map. Once again, the data from the crystals was not very complete, so this needs to be addressed. The two singly iodinated oligonucleotide co-crystals (5' end iodinated, and middle iodinated) have also been subsequently recollected, with better completenesses, to resolutions of 2.5 Å, and 2.2 Å respectively.

In addition, the co-crystals have been soaked in the original growth droplets with potassium hexachloroiridate(III), though no site has been found. A large new batch of tri-thymidine oligonucleotide is now available, and this will allow thorough parameterisation of the co-crystals in terms of stabilising solutions. The scarcity of the oligonucleotides was the reason why thorough trials had not been carried out previously. It is hoped that a suitable stabilising solution will be found to enable heavy atom soaks to be carried out on these types of crystals, as even a weak site will probably be enough to solve the phase problem now.

The native crystals have been soaked in a number of new compounds, and have proved resistant to derivatisation, though there is a possible weak site with thallium sulphate. The approach to derivatising this form of the protein will be to try replacing some of the sodium sulphate solution in the growth droplets with small quantities of heavy atom sulphates. A similar approach was previously attempted (Section **6.3.1**). In addition a new batch of protein will be diluted and stoichiometrically mixed slowly in aliquots at 4°C, with mercury, gold, and platinum containing compounds in the hope that single site

derivatives may be formed prior to crystallisation. The hope is that in a very dilute environment, with a slow enough addition of the metal, the reaction may favour the most reactive site on each molecule. There is no guarantee that the protein will still crystallise afterwards, or that any crystals formed will be isomorphous. Another experiment to be carried out is the iodination of tyrosine residues in these crystals (Sigler, 1970).

In addition, the production of selenomethionyl HSV1 UDGase was attempted, in order to attempt to grow isomorphous crystals for MAD phasing studies. The *E.coli* strain J51 was grown in M9 medium (Sambrook *et al.*, 1989), supplemented with recommended amino acids (Doublié and Carter, 1992), incorporating selenomethionine at 25 mg/ ml. The purification was done as normal, but although UDGase was present, some other *E.coli* proteins exhibited modified behaviour and were not lost in the initial ion-exchange steps. The preparation was not very pure, and was abandoned after several days. The HSV1 UDGase clone pBT1 has since been modified such that the N-terminal fusion of the translation product of the fused ORF is now only 11 residues instead of 40, negating the requirement for porcine enterokinase treatment. The protein thus produced has since been crystallised, under slightly modified conditions, and is isomorphous with the native protein crystals. B834(DE3). The methionine auxotrophic variant of BL21(DE3) has been obtained, and has been successfully transformed with the modified pBT1 plasmid (pBTS1). Production of selenomethionyl HSV1 UDGase will now be re-attempted, and purification should be possible in a single step by IMAC.

If enough phasing is obtained in each of the two crystal forms to positively identify some part of the secondary structure, then combination of the phases from both forms may be attempted. If there is a very large change in conformation of the enzyme (unlikely in the light of the CD analysis) this would be too difficult to achieve. Once the structure of the enzyme is solved and refined, other work is immediately possible. This is the elucidation of the reaction mechanism. This will be approached in a number of ways. Firstly, some carbocyclic deoxyribose containing oligonucleotides have been synthesised by Mike Blackburn and co-workers at Newcastle University. The uracil base is linked to the carbocyclic sugar, and it is hoped that this will inhibit the enzyme. This will be tested by end labelling the oligonucleotide, incubating it with HSV1 UDGase and then heating the reaction to 95°C in alkali to break abasic site backbones. The oligonucleotides will then be run out on sequencing gels to visualise the reaction products. Controls will be normal oligonucleotides with the uracil at the same position.

A similar approach is being used with oligonucleotides made at Edinburgh University by Katherine McAuley-Hecht and Tom Brown from a phosphoramidite supplied by Richard Walker and co-workers at Birmingham University. This phosphoramidite has uracil linked to a 4'-thiolated deoxyribose sugar. Again, it is hoped that this will inhibit HSV1 UDGase. If either of these types of oligonucleotide are successful inhibitors of the reaction, then co-crystallisation and molecular replacement from the solved structure will be used to see which residues in UDGase interact with uracil. In addition, the activity of the UDGase in crystals will be investigated. This will most likely be achieved by incubating a native crystal in the reaction mixture at pH 8.0 for varying lengths of time at 37°C. If it is found that the crystals are able to turn over the substrate, then time-resolved X-ray crystallography may be used to follow the reaction. This will be done by setting up co-crystals with bromouracil containing oligonucleotides, and taking Laue diffraction images after a pulse of 313 nm light, which causes the loss of bromine from uracil.

The co-crystallisation of specific HSV1 UDGase uracil-derivative inhibitors with the overexpressed HSV1 UDGase is currently being carried out in this laboratory with inhibitors made by George Wright at Massachusetts Medical School (Focher *et al.*, 1993). It is hoped that these inhibitors are bound under the crystallisation conditions, and will be visible in a solved crystal structure. The crystals appear to be identical to the native UDGase crystals, though X-ray analysis has yet to be carried out on them.

A PCR is to be carried out in order to amplify the *B.subtilis* bacteriophage PBS1 UDGase inhibitor protein (the PBS1 lysate was kindly supplied by Jeff Errington at Oxford University). This overexpressed protein will be purified by exploiting its affinity for UDGase, in order to attempt crystallisation of the complex. The structure of this very specific inter-protein complex will be of great interest, and will allow the exact mode of inhibition of UDGase to be visualised in atomic detail. This data will once again complement previously published biochemical studies of the complex (Winters and Williams, 1990).

Finally, many interesting mutagenesis experiments can be attempted once the structure of the HSV1 UDGase has been solved. The alteration of key residues in the base and DNA recognition sites can be attempted to try to change the specificity or efficiency of the enzyme.

REFERENCES

Aasland, R; Olsen, L. C; Spurr, N. K; Krokan, H. E. and Helland, D. E. (1990). Chromosomal Assignment of Human Uracil-DNA Glycosylase to Chromosome 12. *Genomics*, **7**, pp.139-141.

Adhya, S. and Gottesman, M. (1978). Control of Transcription Termination. *Ann. Rev. Biochem.*, **47**, pp.967-996.

Anderson, C. T. M. and Friedberg, E. C. (1980). The Presence of Nuclear and Mitochondrial Uracil-DNA Glycosylase in Extracts of Human KB Cells. *Nucl. Acids Res.*, **8**, pp.875-888.

Arnold, F. H. (1991). Metal-Affinity Separations: A New Dimension in Protein Processing. *Bio/Technology*, **9**, pp.151-156.

Barnes, W. M. (1977). Plasmid Detection and Sizing in Single Colony Lysates. *Science*, **195**, pp.393-394.

Bayley, P. (1980). Circular Dichroism and Optical Rotation. *An Introduction to Spectroscopy for Biochemists*, **Academic Press**, pp.148-234. Brown, S. B. Editor.

Bej, A. K; Mahbubani, M. H. and Atlas, R. M. (1991). Amplification of Nucleic Acids by Polymerase Chain Reaction (PCR) and Other Methods and their Applications. *Critical Reviews in Biochemistry and Molecular Biology*, **26**, pp.301-334.

Benesch, R. and Benesch, R. E. (1958). Thiolation of Proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **44**, pp.848-853.

Benisek, W. F. and Richards, F. M. (1968). Attachment of Metal-Chelating Functional Groups to Hen Egg White Lysozyme. *J. Biol. Chem.*, **243**, pp.4267-4271.

Bennett, S. E. and Mosbaugh, D. W. (1992). Characterization of the *Escherichia coli* Uracil-DNA Glycosylase·Inhibitor Protein Complex. *J. Biol. Chem.*, **267**, pp.22512-22521.

Bennett, S. E; Schimerlik, M. I. and Mosbaugh, D. W. (1993). Kinetics of the Uracil-DNA Glycosylase/Inhibitor Protein Association. *J. Biol. Chem.*, **268**, pp.26879-26885.

Asahara, H; Wistort, P. M; Bank, J. F; Bakerian, R. H. and Cunningham, R. P. (1989). Purification and Characterization of *Escherichia coli* Endonuclease III from the Cloned *n*th Gene. *Biochemistry*, **28**, pp.4444-4449.

Bensen, R. J. and Warner, H. R. (1987). Partial Purification and Characterization of Uracil-DNA Glycosylase Activity from Chloroplasts of *Zea mays* Seedlings. *Plant. Physiol.*, **84**, pp.1102-1106.

Birnboim, H. C. and Doly, J. (1979). A Rapid alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. *Nucl. Acids Res.*, **7**, pp.1513-1523.

Blaisdell, P. and Warner, H. (1983). Partial Purification and Characterization of a Uracil-DNA Glycosylase from Wheat Germ. *J. Biol. Chem.*, **258**, pp.1603-1609.

Blake, C. C. F. (1968). The Preparation of Isomorphous Derivatives. *Advances in Protein Chemistry*, **3**, pp.59-120.

Blundell, T. L. and Johnson, L. N. (1976). *Protein Crystallography*, **Academic Press, New York.**

Bolivar, F; Rodriguez, R. L; Greene, P. J; Betlach, M. C; Heyneker, H. L. and Boyer, H. W. (1977). Construction and Characterization of New Cloning Vehicles - II. A Multipurpose Cloning System. *Gene*, **2**, pp.95-113.

Bones, A. M. (1993). Expression and Occurrence of Uracil-DNA Glycosylase in Higher Plants. *Physiologia Plantarum*, **88**, pp.682-688.

Boorstein, R. J; Levy, D. D. and Teebor, G. W. (1987). 5-Hydroxymethyluracil-DNA Glycosylase Activity may be a Differentiated Mammalian Function. *Mutation Res.*, **121**, pp.7-16.

Botta, M; Saladino, R; Gentile, G; Summa, V; Nicoletti, R; Verri, A; Focher, F. and Spadari, S. (1994). Researches on Antiviral Agents. 4[1]. Studies on the Chemistry of 6-Methyl-2-methoxy-4-O-acyloxy and 6-Methyl-2,4-di-O-acyloxypyrimidine Derivatives as New Acylation Reagents and Inhibitors of Uracil-DNA Glycosylases. *Tetrahedron*, **50**, pp.3603-3618.

Bradford, M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analyt. Biochem.*, **72**, pp.248-254.

Brynolf, K; Eliasson, R. and Reichard, P. (1978). Formation of Okazaki Fragments in Polyoma DNA Synthesis Caused by Misincorporation of Uracil. *Cell*, **13**, pp.573-580.

Boiteux, S; O'Connor, T. R. and Laval, J. (1987). Formamidopyrimidine-DNA Glycosylase of *Escherichia coli*: Cloning and Sequencing of the *fpg* structural gene and overproduction of the protein. *The EMBO J.*, **6**, pp.3177-3183.

Burgers, P. M. J. and Klein, M. B. (1986). Selection by Genetic Transformation of a *Saccharomyces cerevisiae*Mutant Defective for the Nuclear Uracil-DNA Glycosylase. *J. Bacteriol.*, **166**, pp.905-913.

Caradonna, S. and Cheng, Y. -C. (1980). Uracil-DNA Glycosylase - Purification and Properties of this Enzyme Isolated from Blast Cells of Acute Myelocytic Leukemia Patients. *J. Biol. Chem.*, **255**, 2293-2300.

Caradonna, S; Worrad, D. and Lirette, R. (1987). Isolation of a Herpes Simplex Virus cDNA Encoding the DNA Repair Enzyme Uracil-DNA Glycosylase. *J. Virol.*, **61**, pp.3040-3047.

Carmody, M. W. and Vary, C. P. H. (1993). Inhibition of DNA Hybridization Following Partial dUTP Substitution. *BioTechniques*, **15**, pp.692-699

Carter [Jr.], C. W; Baldwin, E.T. and Frick, L. (1988). Statistical Design of Experiments for Protein Crystal Growth and the Use of a Precrystallisation Assay. *J. Crystal Growth*, **90**, pp.60-73.

Carter [Jr.], C. W. (1992). Design of Crystallisation Experiments and Protocols. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.47-72. Ducruix, A. and Giegé, R. Editors.

CCP4. (1979). The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography, distributed from Daresbury Laboratory, Warrington WA4 4AD, UK

Centifanto-Fitzgerald, Y; Varnell, E. D. and Kaufman, H. E. (1982). Initial Herpes Simplex Virus Type 1 Infection Prevents Ganglionic Superinfection by Other Strains. *Infection and Immunity*, **35**, pp.1125-1132.

Chang, A. C. Y. and Cohen, S. N. (1978). Construction and Characterization of Amplifiable Multicopy DNA Cloning Vehicles Derived from the P15A Cryptic Miniplasmid. *J. Bacteriol.*, **134**, pp.1141-1156.

Chater, K. F; Hopwood, D. A; Kieser, T. and Thompson, C. J. (1981). Gene-Cloning in *Streptomyces. Curr. Topics in Microbiol. and Immunol.*, **96**, pp.69-95.

Chayen, N. E; Shaw Stewart, P. D. and Blow, D. M. (1992). Microbatch Crystallization Under Oil - A New Technique Allowing Many Small-Volume Crystallization Trials. *J. Crystal Growth*, **122**, pp.176-180.

Chen, J. -D. and Lacks, S. A. (1991). Role of Uracil-DNA Glycosylase in Mutation Avoidance by *Streptococcus pneumoniae*. *J. Bacteriol.*, **173**, pp.283-290.

Clark, J. M. (1988). Novel Non-Templated Nucleotide Addition Reactions Catalyzed by Procaryotic and eucaryotic DNA Polymerases. *Nucl. Acids Res.*, **16**, pp.9677-9686.

Cohen, S. N; Chang, A. C. Y. and Hsu, L. (1972). Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **69**, pp.2110-2114.

Colson, P. and Verly, W. G. (1983). Intracellular Localization of Rat-Liver Uracil-DNA Glycosylase. *Euro. J. Biochem.*, **134**, pp.415-420.

Cone, R; Duncan, J; Hamilton, L. and Friedberg, E. C. (1977). Partial Purification and Characterization of a Uracil-DNA N-Glycosidase from *Bacillus subtilis*. *Biochemistry*, **16**, pp.3194-3201.

Cook, M. L. and Stevens, J. G. (1973). Pathogenesis of Herpetic Neuritis and Ganglionitis in Mice: Evidence for Intra-Axonal Transport of Infection. *Infection and Immunity*, **7**, pp.272-288.

Cool, B. L. and Sirover, M. A. (1989). Immunocytochemical Localization of the Base Excision Repair Enzyme Uracil-DNA Glycosylase in Quiescent and Proliferating Normal Human Cells. *Cancer Res.*, **49**, pp.3029-3036.

Crosby, B; Prakash, L; Davis, H. and Hinkle, D. C. (1981). Purification and Characterization of a Uracil-DNA Glycosylase from the Yeast, *Saccharomyces cerevisiae*. *Nucl. Acids Res.*, **9**, pp.5797-5809.

Debenham, P. G; Webb, M. B. T; Jones, N. J. and Cox, R. (1987). Molecular Studies on the Nature of the Repair Defect in Ataxia-Telangiectasia and their Implications for Cellular Radiobiology. *The J. Cell Science*, **6**, (Supplement) pp.177-189.

Delort, A. -M; Neumann, J. M; Molko, D; Hervé, M; Téoule, R. and Tran Dinh, S. (1985). Influence of Uracil Defect on DNA Structure: ^{1}H NMR Investigation at 500 MHz. *Nucl. Acids Res.*, **13**, pp.3343-3355.

de Murcia, G. and de Murcia, J. M. (1994). Poly(ADP-Ribose) Polymerase: A Molecular Nick Sensor. *Trends in the Biochem. Sci.*, **19**, pp.172-176.

Derewenda, U; Derewenda, Z; Dodson, E. J; Dodson, G. G; Reynolds, C. D; Smith, G. D; Sparks, C. and Swenson, D. (1989). Phenol Stabilizes More Helix in a New Symmetrical Zinc Insulin Hexamer. *Nature*, **338**, pp.594-596.

Deshmane, S. L. and Fraser, N. W. (1989). During Latency, Herpes Simplex Virus Type 1 DNA is Associated with Nucleosomes in a Chromatin Structure. *J. Virol.*, **63**, pp.943-947.

Devereux, J; Haeberli, P. and Smithies, O. (1984). A Comprehensive Set of Sequence Analysis Programs for the VAX. *Nucl. Acids Res.*, **12**, pp.387-395.

Dianov, G; Sedgwick, B; Daly, G; Olsson, M; Lovett, S. and Lindahl, T. (1994). Release of 5'-Terminal Deoxyribose-Phosphate Residues from Incised Abasic Sites in DNA by the *Escherichia coli* RecJ Protein. *Nucl. Acids Res.*, **22**, pp.993-998.

Di Donato, A; de Nigris, M; Russo, N; Di Biase, S. and D'Alessio, G. (1993). A Method for Synthesizing Genes and cDNAs by the Polymerase Chain Reaction. *Analyt. Biochem.*, **212**, pp.291-293.

Domena, J. D. and Mosbaugh, D. W. (1985). Purification of Nuclear and Mitochondrial Uracil-DNA Glycosylase from Rat Liver. Identification of Two Distinct Subcellular Forms. *Biochemistry*, **24**, pp.7320-7328.

Domena, J. D; Timmer, R. T.; Dicharry, S. A. and Mosbaugh, D. W. (1988). Purification and Properties of MitochondrialUracil-DNA Glycosylase from Rat Liver. *Biochemistry*, **27**, pp.6742-6751.

Doublié, S. and Carter [Jr.], C. W. (1992). Preparation of Selenomethionyl Protein Crystals. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.47-72. Ducruix, A. and Giegé, R. Editors.

Downes, C. S; Anderson, J. R, and Johnson, R. T. (1993). Fine Tuning of DNA Repair in Transcribed Genes: Mechanisms, Prevalence and Consequences. *BioEssays*, **15**, pp.209-216.

Dube, D. K; Kunkel, T. A; Seal, G. and Loeb, L. A. (1979). Distinctive Properties of Mammalian DNA Polymerases. *Biochimica et Biophysica Acta*, **561**, pp.369-382.

Ducruix, A. and Giegé, R. (1992). Methods of Crystallization. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.47-72. Ducruix, A. and Giegé, R. Editors.

Dudley, B; Hammond, A. and Deutsch, W. A. (1992). The Presence of Uracil-DNA Glycosylase in Insects is Dependent upon Developmental Complexity. *J. Biol. Chem.*, **267**, pp.11964-11967.

Duncan, B. K; Rockstroh, P. A. and Warner, H. R. (1978). *Escherichia coli* K-12 Mutants Deficient in Uracil-DNA Glycosylase. *J. Bacteriol.*, **134**, pp.1039-1045.

Duncan, B. K. (1981). DNA Glycosylases. *The Enzymes*, **XIV**, pp.565-586.

Dunning, A. M; Talmud, P. and Humphries, S. E. (1988). Errors in the Polymerase Chain Reaction. *Nucl. Acids Res.*, **16**, p.10393.

Eftedal, I; Guddal, P. H; Slupphaug, G; Volden, G. and Krokan, H. E. (1993). Concensus Sequences for Good and Poor Removal of Uracil from Double Stranded DNA by Uracil-DNA Glycosylase. *Nucl. Acids Res.*, **21**, pp.2095-2101.

Ellis, K. J. and Morrison, J. F. (1982). Buffers of Constant Ionic Strength for Studying pH-Dependent Processes. *Methods Enzymol.*, **87**, pp.405-426.

Focher, F; Mazzarello, P; Verri, A; Hübscher, U. and Spadari, S. (1990). Activity Profiles of Enzymes that Control the Uracil Incorporation into DNA During Neuronal Development. *Mutation Res.*, **237**, pp.65-73.

Focher, F; Verri, A; Verzeletti, S; Mazzarello. P. and Spadari, S. (1992*a*). Uracil in Ori_S of Herpes Simplex 1 Alters its Specific Recognition By Origin Binding Protein (OBP): Does Virus Induced Uracil-DNA Glycosylase Play a Key Role in Viral Reactivation and Replication? *Chromosoma*, **102**, (Supplement) pp.S67-S71.

Focher, F; Mazzarello, P; Verri, A; Biamonti, G. and Spadari, S. (1992*b*). Enzymatic and DNA/Protein Interaction Studies Indicate That Uracil in Neuronal DNA Could Contribute to Nerve Cell Aging. *Biology of Aging*, **Springer-Verlag**, pp.8-16. Zwilling, R. and Balduini, C. Editors.

Focher, F; Verri, A; Spadari, S; Manservigi, R; Gambino, J. and Wright, G. E. (1993). Herpes simplex Virus Type 1 Uracil-DNA Glycosylase: Isolation and Selective Inhibition by Novel Uracil Derivatives. *Biochem. J.*, **292**, pp.883-889.

Fox, M. S; Radicella, J. P. and Yamamoto, K. (1994). Some Features of Base Pair Mismatch Repair and its Role in the Formation of genetic Recombinants. *Experientia*, **50**, pp.253-260.

Frederico, L. A; Kunkel, T. A. and Ramsey Shaw, B. (1993). Cytosine Deamination in Mismatched Base Pairs. *Biochemistry*, **32**, pp.6523-6530.

Freifelder, D. (1982). *Physical Biochemistry - Applications to Biochemistry and Molecular Biology* (Second Edition), **W. H. Freeman and Company**.

Friedberg, E. C; Ganesan, A. K. and Minton, K. (1975). N-Glycosidase Activity in Extracts of *Bacillus subtilis* and its Inhibition After Infection with Bacteriophage PBS2. *J. Virol.*, **16**, pp.315-321.

Friedberg, E. C. (1985). *DNA Repair*, **W. H. Freeman and Company**. Friedberg, E. C. Author.

Gallwitz, U; King, L. and Perham, R. N. (1974). Preparation of an Isomorphous Heavy-Atom Derivative of Tobacco Mosaic Virus by Chemical Modification with 4-Sulpho-phenylisothiocyanate. *J. Mol. Biol.*, **87**, pp.257-264.

Gibbs, R. A. (1990). DNA Amplification by the Polymerase Chain Reaction. *Analytical Chemistry*, **62**, pp.1202-1214.

Giegé, R; Dock, A. C; Kern, D; Lorber, B; Thierry, J. C. and Moras, D. (1986). The Role of Purification in the Crystallization of proteins and Nucleic acids. *J. Crystal Growth*, **76**, pp.554-561.

Gilliland, G. L. (1988). A Biological Macromolecule Crystallization Database: A Basis for a Crystallization Strategy. *J. Crystal Growth*, **90**, pp.51-59.

Glusker, J. P. and Trueblood, K. N. (1985). *Crystal Structure Analysis - A Primer* (Second Edition), **Oxford University Press.**

Gottesman, S. (1990). Minimizing Proteolysis in *Escherichia coli*: Genetic Solutions. *Methods Enzymol.*, **185**, pp.119-129.

Grodberg, J. and Dunn, J. J. (1988). *ompT* Encodes the *Escherichia coli* Outer Membrane Protease That Cleaves T7 RNA Polymerase During Purification. *J. Bacteriol.*, **170**, pp.1245-1253.

Grunstein, M. and Hogness, D. S. (1975). Colony Hybridization: A Method for the Isolation of Cloned DNAs That Contain a Specific Gene. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, pp.3961-3965.

Guyer, R. B; Nonnemaker, J. M. and Deering, R. A. (1986). Uracil-DNA Glycosylase Activity from *Dictyostelium discoideum*. *Biochimica et Biophysica Acta*, **868**, pp.262-264.

Hädener, A; Matzinger, P. K; Malashkevich, V. N; Louie, G. V; Wood, S. P; Oliver, P; Alefounder, P. R; Pitt, A. R; Abell, C. and Battersby, A. R. (1993). Purification, Characterization, Crystallisation and X-ray Analysis of Selenomethionine-Labelled Hydroxymethylbilane Synthase from *Escherichia coli*. *Euro. J. Biochem.*, **211**, pp.615-624.

Hajdu, J; Acharya, K. R; Stuart, D. I; McLaughlin, P. J; Barford, D; Oikonomakos, N. G; Klein, H. and Johnson, L. N. (1987).Catalysis in the Crystal: Synchrotron Radiation Studies with Glycogen Phosphorylase *b*. *The EMBO J.*, **6**, pp.539-546.

Hanahan, D. (1983). Studies on Transformation of *Escherichia coli* with Plasmids. *J. Mol. Biol.*, **166**, pp.557-580.

Hanawalt, P. C; Cooper, P. K; Ganesan, A. K. and Smith, C. A. (1979). DNA Repair in Bacteria and Mammalian Cells. *Ann. Rev. Biochem.*, **48**, pp.783-836.

Hatahet, Z; Kow, Y. W; Purmal, A. A; Cunningham, R. P. and Wallace, S. S. (1994). New Substrates for Old Enzymes - 5-hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are Substrates for *Escherichia coli* Endonuclease III and Formamidopyrimidine DNA N-Glycosylase While 5-hydroxy-2'-deoxyuridine is a Substrate for Uracil DNA N-Glycosylase. *J. Biol. Chem.*, **269**, pp.18814-18820.

Hemdan, E. S; Zhang, Y. -J.; Sulkowski, E. and Porath, J. (1989). Surface Topography of Histidine Residues: A Facile Probe by Immobilised Metal Ion Affinity Chromatography. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, pp.1811-1815.

Hendrickson, W. A; Pähler, A; Smith, J. L; Satow, Y; Merritt, E. A. and Phizackerley, R. P. (1989). Crystal Structure of Core Streptavidin Determined from Multiwavelength Anomalous Diffraction of Synchrotron Radiation. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, pp.2190-2194.

Hendrickson, W. A; Horton, J. R. and LeMaster, D. M. (1990). Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle for Direct Determination of Three-Dimensional Structure. *The EMBO J.*, **9**, pp. 1665-1672.

Hernandez, P. and Gutierrez, C. (1987). Uracil-DNA Glycosylase Activity is Modulated by the Proliferation Rate and is Lost Upon Differentiation of *Allium cepa* Root Cells. *Biochimica et Biophysica Acta*, **908**, pp.293-297.

Higley, M. and Lloyd, R. S. (1993). Processivity of Uracil-DNA Glycosylase. *Mutation Res.*, **294**, pp.109-116.

Hollstein, M. C; Brooks, P; Linn, S. and Ames, B. N. (1984). Hydroxymethyluracil-DNA Glycosylase in Mammalian Cells. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, pp.4003-4007.

Honess, R. W. and Watson, D. H. (1977). Unity and Diversity in the Herpesviruses. *J. Gen. Virol.*, **37**, pp.15-37.

Honess, R. W. (1984). Herpes Simplex and 'The Herpes Complex': Diverse Observations and A Unifying Hypothesis. *J. Gen. Virol.*, **65**, pp.2077-2107.

Howell, P. L. and Smith, G. D. (1992). Identification of Heavy-Atom Derivatives by Normal Probability Methods. *J. Appl. Cryst.*, **25**, pp.81-86.

Ingraham, H. A; Dickey, L. and Goulian, M. (1986). DNA Fragmentation and Cytotoxicity from Increased Cellular Deoxyuridylate. *Biochemistry*, **25**, pp.3225-3230.

Inouye, M; Arnheim, N. and Sternglanz, R. (1978). Bacteriophage T7 Lysozyme Is an *N*-Acetylmuramyl-L-alanine Amidase. *J. Biol. Chem.*, **248**, pp.7247-7252.

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

Ivarie, R. (1987). Thymine Methyls and DNA-Protein Interactions. *Nucl. Acids Res.*, **15**, pp.9975-9983.

Jancarik, J. and Kim, S. -H. (1991). Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins. *Journal of Applied Crystallography*, **24**, pp.409-411.

Jiricny, J. (1994). Colon Cancer and DNA Repair: Have Mismatches Met Their Match? *Trends in Genet.*, **10**, pp.164-168.

Jones, D. T; Taylor, W. R. and Thornton, J. M. (1992). A New Approach to Protein Fold Recognition. *Nature*, **358**, pp.86-89.

Jones, D. and Thornton, J. (1993). Protein Fold Recognition. *Journal of Computer Aided Molecular Design*, **7**, pp.439-456.

Kaboev, O. K; Luchkina, L. A; Akhmedov, A. T. and Bekker, M. L. (1981). Uracil-DNA Glycosylase from *Bacillus stearothermophilus*. *FEBS LETTERS*, **132**, pp.337-340.

Kaboev, O. K; Luchkina, L. A and Kuziakina, T. I. (1985). Uracil-DNA Glycosylase of Thermophilic *Thermothrix thiopara*. *J. Bacteriol.*, **164**, pp.421-424.

Kabsch, W. (1988). Automatic-Indexing of Rotation Diffraction Patterns. *J. Appl. Cryst.*, **21**, pp.67-71.

Kabsch, W. (1993). Automatic Processing of Rotation Diffraction Data From Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Cryst.*, **26**, pp.795-800.

Karlovsky, P. (1990). Misuse of PCR. *Trends in the Biochem. Sci.*, **15**, p. 419.

Karran, P; Cone, R. and Friedberg, E. C. (1981). Specificity of the Bacteriophage PBS2 Induced Inhibitor of Uracil-DNA Glycosylase. *Biochemistry*, **20**, pp.6092-6096.

Klimasauskas, S; Kumar, S; Roberts, R. J. and Cheng, X. D. (1994). *Hha* I Methyltransferase Flips its Target Base Out of the DNA Helix. *Cell*, **76**, pp.357-369.

Kristensson, K; Lycke, E; Röyttä, M; Svennerholm, B. and Vahlne, A. (1986). Neuritic Transport of Herpes Simplex Virus in Rat Sensory Neurons *in vitro*. Effects of Substances Interacting with Microtubular Function and Axonal Flow [Nocodazole, Taxol and Erythro-9-3-(2-hydroxynonyl)adenine]. *J. Gen. Virol.*, **67**, pp.2023-2028.

Krokan, H. and Wittwer, C. U. (1981). Uracil-DNA Glycosylase from HeLa Cells: General Properties, Substrate Specificity and Effect of Uracil Analogs. *Nucl. Acids Res.*, **9**, pp.2599-2613.

Kunz, B. A. and Kohalmi, S. E. (1991). Modulation of Mutagenesis by Deoxyribonucleotide Levels. *Ann. Rev. Genet.*, **25**, pp.339-359.

Kuo, C. -F; McRee, D. E; Fisher, C. L; O'Handley, S. F; Cunningham, R. P. and Tainer, J. A. (1992). Atomic Structure of the DNA Repair [4Fe-4S] Enzyme Endonuclease III. *Science*, **258**, pp.434-440.

Ladd, M. F. C. and Palmer, R. A. (1993). *Structure Determination by X-ray Crystallography* (Third Edition), **Plenum Press.**

Laemmli, U. K. (1970). Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. *Nature*, **227**, pp.680-685.

Leblanc, J. -P; Martin, B; Cadet, J. and Laval, J. (1982). Uracil-DNA Glycosylase - Purification and Properties of Uracil-DNA Glycosylase from *Micrococccus luteus*. *J. Biol. Chem.*, **257**, pp.3477-3483.

Lindahl, T. (1974). An N-Glycosidase from *Escherichia coli* That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, pp.3649-3653.

Lindahl, T. and Nyberg, B. (1974). Heat-Induced Deamination of Cytosine Residues in Deoxyribonucleic Acid. *Biochemistry*, **13**, pp.3405-3410.

Lindahl, T; Ljungquist, S; Siegert, W; Nyberg, B. and Sperens, B. (1977). DNA N-Glycosidases - Properties of Uracil-DNA Glycosidase from *Escherichia coli*. *J. Biol. Chem.*, **252**, pp.3286-3294.

Lindahl, T. (1979). DNA Glycosylases, Endonucleases for Apurinic/ Apyrimidinic Sites, and Base Excision-Repair. *Prog. Nucl. Acid Res. Mol. Biol.*, **22**, pp.135-192.

Lindahl, T. (1982). DNA Repair Enzymes. *Ann. Rev. Biochem.*, **51**, pp.61-87.

Lindahl, T; Demple, B. and Robins, P. (1982). Suicide-Inactivation of the *Escherichia coli* O^6-Methylguanine-DNA Methyltransferase. *The EMBO J.*, **1**, pp.1359-1363.

Lindahl, T. (1990). Repair of Intrinsic DNA Lesions. *Mutation Res.*, **238**, pp.305-311.

Lindahl, T. (1993). Instability and Decay of the Primary Structure of DNA. *Nature*, **362**, pp.709-715.

Littler, E. and Powell, K. L. (1992). Herpesviruses. *Encyclopedia of Microbiology*, **2**, pp.381-391.

Lorber, B. and Giegé, R. (1992). Preparation and Handling of Biological Macromolecules for Crystallization. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.47-72. Ducruix, A. and Giegé, R. Editors.

Lu, Y. H. and Nègre, S. (1993). Use of Glycerol for Enhanced Efficiency and Specificity of PCR Amplification. *Trends in Genet.*, **9**, p.297.

Lycke, E; Kristensson, K; Svennerholm, B; Vahlne, A. and Ziegler, R. (1984). Uptake and Transport of Herpes Simplex Virus in Neurites of Rat Dorsal Root Ganglia Cells in Culture. *J. Gen. Virol.*, **65**, pp.55-64.

Maizels, N. M. (1973). The Nucleotide Sequence of the Lactose Messenger Ribonucleic Acid Transcribed from the UV5 Promoter Mutant of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **70**, pp.3585-3589.

Makino, F. and Munakata, N. (1977). Isolation and Characterization of a *Bacillus subtilis* Mutant with a Defective N-Glycosidase Activity for Uracil-Containing Deoxyribonucleic Acid. *J. Bacteriol.*, **131**, pp.438-445.

Masters, C. I; Moseley, B. E. B. and Minton, K. W. (1991). AP endonuclease and Uracil-DNA Glycosylase Activities in *Deinococcus radiodurans*. *Mutation Res.*, **254**, pp.263-272.

Mauro, D. J; De Riel, J. K; Tallarida, R. J. and Sirover, M. A. (1993). Mechanisms of Excision of 5-Fluorouracil by Uracil-DNA Glycosylase in Normal Human Cells. *Molecular Pharmacology*, **43**, pp.854-857.

Mazzarello, P; Focher, F; Verri, A. and Spadari, S. (1990). Misincorporation of Uracil into DNA as Possible Contributor to Aging and Abiotrophy of Nerve Cells. *Internat. J. Neuroscience*, **50**, pp. 169-174.

Mazzarello, P; Poloni, M; Spadari, S. and Focher, F. (1992). DNA Repair Mechanisms in Neurological Diseases: Facts and Hypotheses. *Journal of the Neurological Sciences*, **112**, pp.4-14.

McAllister, W. T. and Morris, C. (1981). Utilization of Bacteriophage T7 Late Promoters in Recombinant Plasmids During Infection. *J. Mol. Biol.*, **153**, pp.527-544.

McGeogh, D. J; Dolan, A; Donald, S. and Rixon, F. J. (1985). Sequence Determination and Genetic Content of the Short Unique Region in the Genome of Herpes Simplex Virus Type 1. *J. Mol. Biol.*, **181**, pp.1-13.

McGeogh, D. J; Dolan, A; Donald, S. and Brauer, D. H. K. (1986). Complete DNA Sequence of the Short Repeat Region in the Genome of Herpes Simplex virus Type 1. *Nucl. Acids Res.*, **14**, pp.1727-1745.

McGeogh, D. J. (1987). The Genome of Herpes Simplex Virus: Structure, Replication and Evolution. *J. Cell Science*, **7**, (Supplement) pp.67-94.

McGeogh, D. J; Dalrymple, M. A; Davison, A. J; Dolan, A; Frame, M. C; McNab, D; Perry, L. J; Scott, J. E. and Taylor, P. (1988). The Complete DNA Sequence of the Long Unique Region in the Genome of Herpes Simplex Virus Type 1. *J. Gen. Virol.*, **69**, pp.1531-1574.

McGeogh, D. J; Cunningham, C; McIntyre, G. and Dolan, A. (1991). Comparative Sequence Analysis of the Long Repeat Regions and Adjoining Parts of the Long Unique Regions in the Genomes of Herpes Simplex viruses Types 1 and 2. *J. Gen. Virol.*, **72**, pp.3057-3075.

McPherson, A. (1982). *The Preparation and Analysis of Protein Crystals*, **John Wiley and Sons, New York.**

McPherson, A; Koszelak, S; Axelrod, H; Day, J; Williams, R; Robinson, L; McGrath, M. and Cascio, D. (1986a). An Experiment Regarding Crystallization of Soluble Proteins in the Presence of β-octyl Glucoside. *J. Biol. Chem.*, **261**, pp.1969-1975.

McPherson, A; Koszelak, S; Axelrod, H; Day, J; Robinson, L; McGrath, M; Williams, R. and Cascio, D. (1986b). The Effects of Neutral Detergents on the Crystallization of Soluble Proteins. *J. Crystal Growth*, **76**, pp.547-553.

McPherson, A. (1990). Current Approaches to Macromolecular Crystallization. *Euro. J. Biochem.*, **189**, pp.1-23.

Mellerick, D. M. and Fraser, N. W. (1987). Physical State of the Latent Herpes Simplex Virus Genome in a Mouse Model System: Evidence Suggesting an Episomal State. *Virology*, **158**, pp.265-275.

Meyer-Siegler, K; Mauro, D. J; Seal, G; Wurzer, J; de Riel, J. K. and Sirover, M. A. (1991). A Human Nuclear Uracil-DNA Glycosylase is the 37-kDa Subunit of Glyceraldehyde-3-Phosphate Dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, pp.8460-8464.

Meyer-Siegler, K; Rahman-Mansur, N; Wurzer, J. C. and Sirover, M. A. (1992). Proliferative Dependent Regulation of the Glyceraldehyde-3-Phosphate Dehydrogenase/Uracil-DNA Glycosylase Gene in Human Cells. *Carcinogenesis*, **13**, pp.2127-2132.

Millns, A. K; Carpenter, M. S. and DeLange, A. M. (1994). The Vaccinia Virus-Encoded Uracil-DNA Glycosylase Has an Essential Role in Viral DNA Replication. *Virology*, **198**, pp.504-513.

Modrich, P. (1991). Mechanisms and Biological Effects of Mismatch Repair. *Ann. Rev. Genet.*, **25**, pp.229-253.

Moffatt, B. A. and Studier, F. W. (1987). T7 Lysozyme Inhibits Transcription by T7 RNA Polymerase. *Cell*, **49**, pp.221-227.

Moore, M. H; Gulbis, J. M; Dodson, E. J; Demple, B. and Moody, P. C. E. (1994). Crystal Structure of a Suicidal DNA Repair Protein: The Ada O^6-methylguanine-DNA Methyltransferase from *E.coli*. *The EMBO J.*, **13**, pp.1495-1501.

Morgan, A. R. and Chlebek, J. (1987). Quantitative Assays for Uracil-DNA Glycosylase of High Sensitivity. *Biochemistry and Cell Biology*, **66**, pp.157-160.

Morgan, A. R. and Chlebek, J. (1989). Uracil-DNA Glycosylase in Insects - Drosophila and the Locust. *J. Biol. Chem.*, **264**, pp.9911-9914.

Morgenegg, G; Winkler, G. C; Hübscher, U; Heizmann, C. W; Mous, J. and Kuenzle, C. C. (1986). Glyceraldehyde-3-Phosphate Dehydrogenase is a Nonhistone Protein and a Possible Activator of Transcription in Neurons. *Journal of Neurochemistry*, **47**, 54-62.

Morikawa, K; Matsumoto, O; Tsujimoto, M; Katayanagi, K; Ariyoshi, M; Doi, T; Ikehara, M; Inaoka, T. and Ohtsuka, E. (1992). *Science*, **256**, pp.523-526.

Mosbaugh, D. W. (1988). Enzymology of Uracil-DNA Repair in Mammalian Cells. *Reviews in Biochemical Toxicology*, **9**, pp.69-130. Hodgson, E; Bend, J. R. and Philpot, R. M. Editors.

Mowbray, S. L. and Petsko, G. A. (1983). The Introduction of Specific Sites for Metal Binding in a Crystalline Protein. *J. Biol. Chem.*, **258**, pp.5634-5637.

Mullaney, J; Moss, H. W. McL. and McGeogh, D. J. (1989). Gene UL2 of Herpes Simplex Virus Type 1 Encodes a Uracil-DNA Glycosylase. *J. Gen. Virol.*, **70**, pp.449-454.

Muller, S. J. and Caradonna, S. (1991). Isolation and Characterization of a Human cDNA Encoding Uracil-DNA Glycosylase. *Biochimica et Biophysica Acta*, **1088**, pp.197-207.

Muller, S. J. and Caradonna, S. (1993). Cell Cycle Regulation of a Human Cyclin-Like Gene Encoding Uracil-DNA Glycosylase. *J. Biol. Chem.*, **268**, pp.1310-1319.

Mullis, K; Faloona, F; Scharf, S; Saiki, R; Horn, G. and Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symp. Quant. Biol.*, **LI**, pp.263-273.

Mullis, K. and Faloona, F. (1987). Specific Synthesis of DNA *in vitro* Via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol.*, **155**, pp.335-350.

Myers, L. C. and Verdine, G. L. (1994). DNA Repair Proteins. *Current Opinion in Structural Biology*, **4**, pp.51-59.

Neddermann, P. and Jiricny, J. (1994). Efficient Removal of Uracil from G·U Mispairs by the Mismatch -Specific Thymine DNA Glycosylase from HeLa Cells. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, pp.1642-1646.

Nakabeppu, Y; Kondo H. and Sekiguchi, M. (1984).Cloning and Characterization of the *AlkA* Gene of *Escherichia coli* that Encodes 3-Methyladenine DNA Glycosylase-II *J. Biol. Chem.*, **259**, pp.3723-3729.

Nicotera, T. M. (1991). Molecular and Biochemical Aspects of Bloom's Syndrome. *Cancer Genetics and Cytogenetics*, **53**, pp.1-13.

Norrander, J; Kempe, T. and Messing, J. (1983). Construction of Improved M13 Vectors Using Oligodeoxynucleotide-Directed Mutagenesis. *Gene*, **26**, pp.101-106.

Olsen, L. C; Aasland, R; Wittwer, C. U; Krokan, H. E. and Helland, D. E. (1989). Molecular Cloning of Human Uracil-DNA Glycosylase, a Highly Conserved DNA Repair Enzyme. *The EMBO J.*, **8**, pp.3121-3125.

Park, H. -W; Sancar, A. and Deisenhofer, J. (1993). Crystallisation and Preliminary Crystallographic Analysis of *Escherichia coli* DNA Photolyase. *J. Mol. Biol.*, **231**, pp.1122-1125.

Perucho, M; Salas, J. and Salas, M. L. (1977). Identification of the Mammalian DNA-Binding Protein P8 as Glyceraldehyde-3-Phosphate Dehydrogenase. *Euro. J. Biochem.*, **81**, pp.557-562.

Petsko, G. A. (1985). Preparation of Isomorphous Heavy-Atom derivatives. *Methods Enzymol.*, **114**, pp.147-167.

Prodromou, C. and Pearl, L. H. (1992). Recursive PCR - A Novel Technique for Total Gene Synthesis. *Protein Engineering*, **5**, pp.827-829.

Pyles, R. B. and Thompson, R. L. (1994). Evidence that the Herpes Simplex Virus Type 1 Uracil DNA Glycosylase is Required for Efficient Viral Replication and Latency in the Murine Nervous System. *J. Virol.*, **68**, pp.4963-4972.

Riès-Kautt, M. M. and Ducruix, A. F. (1991). Crystallization of Basic Proteins by Ion Pairing. *J. Crystal Growth*, **110**, pp.20-25.

Riès-Kautt, M. M. and Ducruix, A. F. (1992). Phase Diagrams. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.195-218. Ducruix, A. and Giegé, R. Editors.

Riley, M. and Perham, R. N. (1973). The Reaction of Protein Amino Groups with Methyl 5-Iodopyridine-2-carboximidate - A Possible General Method of Preparing Isomorphous Heavy-Atom Derivatives of Proteins. *Biochem. J.*, **131**, pp.625-635.

Robert, M. C. and Lefaucheux, F. (1988). Crystal Growth in Gels: Principle and Applications. *J. Crystal Growth*, **90**, pp.358-367.

Robert, M. C; Provost, K. and Lefaucheux, F. (1992). Crystallization in Gels and Related Methods. *Crystallisation of Nucleic Acids and Proteins - A Practical Approach*, **IRL PRESS**, pp.127-144. Ducruix, A. and Giegé, R. Editors.

Robinson, M; Lilley, R; Little, S; Emtage, J. S; Yarranton, G; Stephens, P; Millican, A; Eaton, M. and Humphreys, G. (1984). Codon Usage Can Affect Efficiency of Translation of Genes in *Escherichia coli*. *Nucl. Acids Res.*, **12**, pp.6663-6671.

Roizman, B. and Sears, A. E. (1987). An Inquiry into the Mechanisms of Herpes Simplex virus Latency. *Ann. Rev. Microbiol.*, **41**, pp.543-571.

Roizman, B. (1991). Herpesviridae: A Brief Introduction. *Fundamental Virology*, **Raven Press Ltd., New York**, (2nd Edition) pp.849-895. Fields, B. N; Knipe, D. M *et al.* Editors.

Roizman, B. and Sears, A. E. (1991). Herpes Simplex Viruses and Their Replication. *Fundamental Virology*, **Raven Press Ltd., New York**, (2nd Edition) pp.849-895. Fields, B. N; Knipe, D. M *et al.* Editors.

Rosenberg, A. H; Lade, B. N; Chui, D. -S; Lin, S. -W; Dunn, J. J. and Studier, F. W. (1987). Vectors for Selective Expression of Cloned DNAs by T7 RNA Polymerase. *Gene*, **56**, pp.125-135.

Sakumi, K. and Sekiguchi, M. (1990). Structures and Functions of DNA Glycosylases. *Mutation Res.*, **236**, pp.161-172.

Salvesen, G. and Nagase, H. (1989). Inhibition of Proteolytic Enzymes. *Proteolytic Enzymes - A Practical Approach*, **IRL PRESS**, pp.83-124. Beynon, R. J. and Bond, J. S. Editors.

Sambrook, J; Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (Second Edition), **Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.**

Sancar, A. and Sancar, G. B. (1988). DNA Repair Enzymes. *Ann. Rev. Biochem.*, **57**, pp.29-67.

Sanger, F; Nicklen, S. and Coulson, A. R. (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **74**, pp.5463-5467.

Saridakis, E. G; Shaw Stewart, P. D., and Lloyd, L. F. and Blow, D. M. (1994). Phase Diagram and Dilution Experiments in the Crystallization of Carboxypeptidase G_2. *Acta Cryst. - Section D*, **50**, pp.293-297.

Savva, R. and Pearl L. H. (1993). Crystallization and Preliminary X-ray Analysis of the Uracil-DNA Glycosylase DNA Repair Enzyme from Herpes Simplex Virus Type 1. *J. Mol. Biol.*, **234**, pp.910-912.

Schrag, J. D; Venkataram Prasad, B. V; Rixon, F. J. and Chiu, W. (1989). Three-Dimensional Structure of the HSV1 Nucleocapsid. *Cell*, **56**, pp.651-660.

Seal, G. and Sirover, M. A. (1986). Physical Association of the Human Base-Excision Repair Enzyme Uracil-DNA Glycosylase With the 70,000-Dalton Catalytic Subunit of DNA Polymerase α. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, pp.7608-7612.

Seal, G; Arenaz, P. and Sirover, M.A. (1987). Purification and Properties of the Human Placental Uracil-DNA Glycosylase. *Biochimica et Biophysica Acta*, **925**, pp.226-233.

Seal, G; Brech, K; Karp, S. J; Cool, B. L. and Sirover, M. A. (1988). Immunological Lesions in Human Uracil-DNA Glycosylase: Association with Bloom Syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, pp.2339-2343.

Seal, G; Henderson, E. E. and Sirover, M. A. (1990). Immunological Alteration of the Bloom's Syndrome Uracil-DNA Glycosylase in Epstein-Barr Virus-Transformed Human Lymphoblastoid Cells. *Mutation Res.*, **243**, pp.241-248.

Seal, G; Tallarida, R. J. and Sirover, M. A. (1991). Purification and Propertiesof the Uracil-DNA Glycosylase from Bloom's Syndrome. *Biochimica et Biophysica Acta*, **1097**, pp.299-308.

Sedgwick, B. (1987). Molecular Signal for Induction of the Adaptive Response to Alkylation Damage in *Escherichia coli*. *J. Cell Science*, **6**, (Supplement) pp.215-223.

Sedwick, D. W; Kutler, M. and Brown, O. E. (1981). Antifolate-Induced Misincorporation of Deoxyuridine Monophosphate into DNA: Inhibition of High Molecular Weight DNA Synthesis in Human Lymphoblastoid Cells. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, pp.917-921.

Selby, C. P. and Sancar, A. (1993). Molecular Mechanism of Transcription-Repair Coupling. *Science*, **260**, pp.53-58.

Shapiro, A. L; Viñuela, E. and Maizel, J. B. (1967). Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in SDS-Polyacrylamide Gels. *Biochem. Biophys. Res. Comm.*, **28**, pp.815-820.

Shapiro, R. and Klein, R. S. (1966). The Deamination of Cytidine and Cytosine by Acidic Buffer Solutions. Mutagenic Implications. *Biochemistry*, **5**, pp.2358-2362.

Shaw Stewart, P. D. and Khimasia, M. (1994). Predispensed Gradient Matrices - A New Rapid Method of Finding Crystallization Conditions. *Acta Cryst. - Section D*, **50**, pp.441-442.

Sigler, P. B. (1970). Iodination of a Single Tyrosine in Crystals of α-Chymotrypsin. *Biochemistry*, **9**, pp.3609-3617.

Silverstone, A. E; Arditti, R. R. and Magasanik, B. (1970). Catabolite-Insensitive Revertants of *Lac* Promoter Mutants. *Proc. Natl. Acad. Sci. U.S.A.*, **66**, pp.773-779.

Skopes, R. K. (1987). *Protein Purification: Principles and Practice* (Second Edition), Springer-Verlag, New York.

Slupphaug, G; Olsen, L. C; Helland, D; Aasland, R. and Krokan, H. E. (1991). Cell Cycle Regulation and *in vitro* Hybrid arrest Analysis of the Major Human Uracil-DNA Glycosylase. *Nucl. Acids Res.*, **19**, pp.5131-5137.

Slupphaug, G; Markussen, F. -H; Olsen, L. C; Aasland, R; Aarsæther, N; Bakke, O; Krokan, H. E. and Helland, D. E. (1993). Nuclear and Mitochondrial Forms of Human Uracil-DNA Glycosylase are Encoded by the Same Gene. *Nucl. Acids Res.*, **21**, pp.2579-2584.

Stuart, D. T; Upton, C; Higman, M. A; Niles, E. G. and McFadden, G. (1993). A Poxvirus-Encoded Uracil-DNA Glycosylase Is Essential for Virus Viability. *J. Virol.*, **67**, pp.2503-2512.

Studier, F. W. and Moffatt, B. A. (1986). Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J. Mol. Biol.*, **189**, pp.113-130.

Studier, F. W; Rosenberg, A. H; Dunn, J. J. and Dubendorff, J. W. (1990). Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.*, **185**, pp.60-89.

Stura, E. A. and Wilson, I. A. (1990). Analytical and Production Seeding Techniques. *Methods - A Companion To Methods in Enzymology*, **1**, pp.38-49.

Talpaert-Borlé, M; Clerici, L. and Campagnari, F. (1979). Isolation and Characterization of a Uracil-DNA Glycosylase from Calf Thymus. *J. Biol. Chem.*, **254**, pp.6387-6391.

Talpaert-Borlé, M; Campagnari, F. and Creissen, M. (1982). Properties of Purified Uracil-DNA Glycosylase from Calf Thymus. *J. Biol. Chem.*, **257**, pp.1208-1214.

Tanaka, K. and Wood, R. D. (1994). Xeroderma Pigmentosum and Nucleotide Excision Repair of DNA. *Trends in the Biochem. Sci.*, **19**, pp.83-86.

Tessman, I; Kennedy, M. A. and Liu, S. -K. (1994). Unusual Kinetics of Uracil Formation in Single and Double-Stranded DNA by Deamination of Cytosine in Cyclobutane Pyrimidine Dimers. *J. Mol. Biol.*, **235**, pp.807-812.

Timasheff, S. N. and Arakawa, T. (1988). Mechanism of Protein Precipitation and Stabilization by Co-Solvents. *J. Crystal Growth*, **90**, pp.39-46.

Steinum, A-L. and Seeburg, E. (1986). Nucleotide Sequence of the *tag* gene from *Escherichia coli*. *Nucl. Acids Res.*, **14**, pp. 3763-3772.

Timmer, W. C. and Villalobos, J. M. (1993). The Polymerase Chain Reaction. *Journal of Chemical Education*, **70**, pp.273-280.

Tomilin, N. V. and Aprelikova, O. N. (1989). Uracil-DNA Glycosylases and DNA Uracil Repair. *International Review of Cytology*, **114**, pp.125-179.

Twigg, A. J. and Sherratt, D. (1980). Trans-Complementable Copy-Number Mutants of Plasmid ColE1. *Nature*, **283**, pp.216-218.

Upton, C.; Stuart, D. T. and McFadden, G. (1993). Identification of a Poxvirus Gene Encoding a Uracil-DNA Glycosylase. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, pp.4518-4522.

Varshney, U. and van de Sande, J. H. (1991). Specificities and Kinetics of Uracil Excision from Uracil-Containing DNA Oligomers by *Escherichia coli* Uracil-DNA Glycosylase. *Biochemistry*, **30**, pp.4055-4061.

Verri, A; Mazzarello, P; Biamonti, G; Spadari, S. and Focher, F. (1990). The Specific Binding of Nuclear Protein(s) to the cAMP Rsponsive Binding Element (CRE) Sequence (TGACGTCA) is Reduced by the Misincorporation of U and Increased by the Deamination of C. *Nucl. Acids Res.*, **18**, pp.5775-5780.

Verri, A; Mazzarello, P; Spadari, S. and Focher, F. (1992). Uracil-DNA Glycosylases Preferentially Excise Mispaired Uracil. *Biochem. J.*, **287**, pp.1007-1010.

Vilpo, J. A. and Vilpo, L. M. (1989). Normal Uracil-DNA Glycosylase Activity in Bloom's Syndrome Cells. *Mutation Res.*, **210**, pp.59-62.

Vollberg, T. M; Seal, G. and Sirover, M. A. (1987). Monoclonal Antibodies Detect Conformational Abnormality of Uracil-DNA Glycosylase in Bloom's Syndrome Cells. *Carcinogenesis*, **8**, pp.1725-1729.

Vollberg, T. M; Meyer-Siegler, K; Cool, B. L. and Sirover, M. A. (1989). Isolation and Characterization of the Human Uracil-DNA Glycosylase Gene. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, pp.8693-8697.

Walker, J. E; Saraste, M; Runswick, M. J. and Gay, N. J. (1982). Distantly Related Sequences in the Alpha-Subunits and Beta-Subunits of ATP Synthase, Myosin, Kinases, and other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. *The EMBO J.*, **1**, pp.945-951.

Walker, N; Marcinowski, S and Hillen, H./ Mächtle, W./ Jones, Y. and Stuart, D. (1990). Crystallization of Human Tumour Necrosis Factor. *J. Crystal Growth*, **100**, pp.168-170.

Wallace, S. S. (1988). AP Endonucleases and DNA Glycosylases That Recognise Oxidative DNA Damage. *Environmental and Molecular Mutagenesis*, **12**, pp.431-477.

Wang, Z. and Mosbaugh, D. W. (1988). Uracil-DNA Glycosylase Inhibitor of Bacteriophage PBS2: Cloning and Effects of Expression of the Inhibitor Gene in *Escherichia coli*. *J. Bacteriol.*, **170**, pp.1082-1091.

Wang, Z. and Mosbaugh, D. W. (1989). Uracil-DNA Glycosylase Inhibitor Gene of Bacteriophage PBS2 Encodes a Binding Protein Specific for Uracil-DNA Glycosylase. *J. Biol. Chem.*, **264**, pp.1163-1171.

Wang, Z; Smith, D. G. and Mosbaugh, D. W. (1991). Overproduction and Characterization of the Uracil-DNA Glycosylase Inhibitor of Bacteriophage PBS2. *Gene*, **99**, pp.31-37.

Warner, H. R; Johnson, L. K. and Snustad, D. P. (1980). Early Events After Infection of *Escherichia coli* by Bacteriophage T5. *J. Virol.*, **33**, pp.535-538.

Weber, K. and Osborn, M. (1969). The Reliability of Molecular weight Determined by Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. *J. Biol. Chem.*, **244**, pp.4406-4412.

Weber, P. C. (1991). Physical Principles of Protein Crystallization. *Advances in Protein Chemistry*, **41**, pp.1-36.

Weeda, G; Hoeijimakers, H. J. and Bootsma D. (1993). Genes Controlling Nucleotide Excision Repair in Eukaryotic Cells. *BioEssays*, **15**, pp.249-258.

Weng, Y. and Sirover, M. A. (1993). Developmental regulation of the Base Excision Repair Enzyme Uracil-DNA Glycosylase in the Rat. *Mutation Res.*, **293**, pp.133-141.

Williams, M. V. and Pollack, J. D. (1990). A Mollicute (Mycoplasma) DNA Repair Enzyme: Purification and Characterization of Uracil-DNA Glycosylase. *J. Bacteriol.*, **172**, pp.2979-2985.

Williamson, M. P. (1994). The Structure and Function of Proline-Rich Regions in Proteins. *Biochem. J.*, **297**, pp.249-260.

Winters, T. A. and Williams, M. V. (1990). Use of the PBS2 Uracil-DNA Glycosylase Inhibitor to Differentiate the Uracil-DNA Glycosylase Activities Encoded by Herpes Simplex virus Types 1 and 2. *Journal of Virological Methods*, **29**, pp.233-242.

Winters, T. A. and Williams, M. V. (1993). Purification and Characterization of the Herpes Simplex Virus Type 2-Encoded Uracil-DNA Glycosylase. *Virology*, **195**, pp.315-326.

Wist, E; Unhjem, O. and Krokan, H. (1978). Accumulation of Small Fragments of DNA in Isolated HeLa Cell Nuclei Due to Transient Incorporation of dUMP. *Biochim. Biophys. Acta*, **520**, pp.253-270.

Wood, S. P. (1990). Purification for Crystallography. *Protein Purification Applications - A Practical Approach*, **IRL PRESS**, pp. 45-58. Harris, E. L. V. and Angal, S. Editors.

Wyckoff, H. W; Doscher, M; Tsernoglou, D; Inagami, T; Johnson, L. N; Hardman, K. D; Allewell, N. M; Kelly, D. M. and Richards, F. M. (1967). Design of a Diffractometer and Flow Cell System for X-ray Analysis of Crystalline Proteins with applications to the Crystal Chemistry of Ribonuclease-S. *J. Mol. Biol.*, **27**, pp.563-578.

Yamagata, Y; Odawara, K; Tomita, K; Nakabeppu, Y. and Sekiguchi, M. (1988). Crystallisation and Preliminary X-ray Diffraction Studies of 3-methyladenine-DNA Glycosylase II from *Escherichia coli*. *J. Mol. Biol.*, **204**, pp.1055-1056.

Yamamoto, Y. and Fujiwara, Y. (1986). Abnormal Regulation of Uracil-DNA Glycosylase Induction During Cell Cycle and Cell Passage in Bloom's Syndrome Fibroblasts. *Carcinogenesis*, **7**, pp.305-310.

Yang, A. -S. and Honig, B. (1993). On the pH Dependence of Protein Stability. *J. Mol. Biol.*, **231**, pp.459-474.

Yanisch-Perron, C; Vieira, J. and Messing, J. (1985). Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. *Gene*, **33**, pp.103-119.

Yamagata, Y; Odawara, K; Tomita, K; Nakabeppu, Y. and Sekiguchi, M. (1988). Crystallisation and Preliminary X-ray Diffraction Studies of 3-methyladenine-DNA Glycosylase II from *Escherichia coli*. *J. Mol. Biol.*, **204**, pp.1055-1056.