McGeehan, John Edward (1998) Molecular characterisation of herpes simplex virus type 1 deoxyuridine triphosphatase. PhD thesis

http://theses.gla.ac.uk/6104/

MOLECULAR CHARACTERISATION OF HERPES SIMPLEX VIRUS TYPE 1 DEOXYURIDINE TRIPHOSPHATASE

by

JOHN EDWARD MCGEEHAN

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF SCIENCE AT THE UNIVERSITY OF GLASGOW

MRC VIROLOGY UNIT
IBLS DIVISION OF VIROLOGY
CHURCH STREET
GLASGOW
G11 5JR

DECEMBER, 1998

Contents

Chapter 3 - Methods

Chapter 4 - Results of Computer Modelling

Chapter 5 - Results of Experimental Work

Acknowledgements

Summary

Herpes simplex virus type 1 (HSV-1), like other members of the Herpesviridae, has the ability replicate in different cell types and cause latent infection. Replication of the viral genome requires a multitude of reactions that require protein catalysts. Herpesviruses in general encode many of their own replication enzymes in the viral genome allowing the efficient production of progeny virus. The reason why herpesviruses, and many other viruses, encode enzymes that essentially duplicate the function of cellular enzymes is presumably to increase the efficiency of virus replication in host tissues. This may be especially important to the herpesviruses during latent infection. HSV-1, for example, must replicate in both mucosal and neuronal tissues. For productive infection and spread to a new host the virus must reactivate from resting neurons where the availability of cellular replication machinery is limited.

To ensure the efficient transfer of genetic information to the next generation the fidelity of genome replication must be high. Many enzymes are involved in the replication of DNA although the number of these functions that are virally encoded varies dramatically between different virus families. DNA polymerase provides the main catalytic function for the replication of DNA and, despite its high processivity, retains the ability to faithfully replicate the DNA template with considerable fidelity. This accuracy is only possible if there is a balanced pool of the four deoxyribonucleotides necessary for DNA synthesis: dATP, dCTP, dGTP and dTTP. Disruption of the nucleotide pool can lead to increased mutation rates and DNA fragmentation. Transfer of intact genetic information is crucial and in order that these pools are maintained, many viruses encode their own version of specific nucleotide metabolism enzymes.

This thesis deals with the investigation of one of these enzymes, deoxyuridine triphosphatase (dUTPase). This highly specific enzyme converts dUTP to dUMP thus reducing the available pool of uracil for DNA misincorporation. Uracil can also occur in DNA by the deamination of cytosine which is potentially mutagenic. To counteract this, a repair mechanism has evolved mediated by uracil DNA glycosylase. This excision repair process involves a local strand break in the DNA backbone. Excessive uracil incorporation induces multiple rounds of excision repair and can lead to DNA fragmentation. dUTPase also acts to supply dUMP for the synthesis of dTTP. dUTPase is ubiquitous in nature and found in both prokaryotic and eukaryotic organisms. Many, but not all, viruses encode their own dUTPase. This is also true of the herpesviruses where members of the α- and γ-subfamilies encode a dUTPase whereas members of the β-subfamily do not. The dUTPase of HSV-1 was originally found to be non-essential in tissue culture. Only recently has interest in the herpesvirus dUTPase increased when experiments using the mouse model demonstrated that HSV-

1 mutants which lacked dUTPase activity had over a 1000-fold reduction in neurovirulence, neuroinvasiveness and reactivation from latency (R. B. Pyles, N. M. Sawtell & R. L. Thompson, *Journal of Virology* **66**:6706-6713, 1992). There have been several subsequent reports of dUTPase⁻ mutant viruses replicating with wild type kinetics in tissue culture but being severely impaired during growth *in vivo*. These include dUTPase⁻ mutants of FIV, EIAV, CAEV and visna virus.

Analysis of primary sequence data revealed a subset of open reading frames that were predicted to encode dUTPases based on five areas of local primary sequence conservation (Motifs 1-5) (D. J. McGeoch, *Nucleic Acids Research* **18**:4105-4110, 1990). This was subsequently confirmed by the characterisation of several of these predicted protein products. The differences in the primary sequence organisation of these motif regions allowed the description of two distinct dUTPase classes. The class I dUTPases are encoded by a diverse range of organisms and are characterised by a trimeric arrangement with subunit protein lengths approximating 150 amino acids. The class II dUTPases are specific to the herpesviruses and are characterised by a monomeric arrangement with a protein chain length approximately double that of their class I counterparts. It has been proposed that the class II dUTPases arose by the intragenic duplication of the class I open reading frame.

The first structure of a dUTPase was published with the crystallisation of the *E.coli* version (E. S. Cedergren-Zeppezauer, G. Larsson, P. O. Nyman, Z. Dauter & K. S. Wilson, *Nature* **355**:740-743, 1992). Subsequently structures of FIV, EIAV and the human dUTPase have now been published. To date there is no available structure of a class II dUTPase. In this thesis the class I structures were used as a basis to investigate the HSV-1 class II dUTPase in terms of structural and evolutionary relationships.

To allow a defined approach to functional analysis of the HSV-1 dUTPase a tertiary structural model was generated for the class II enzymes. Following intensive primary sequence analysis a method was devised from comparing class I and class II sequences directly. Secondary structure prediction programs were utilised to judge the basic structural similarities between the two classes allowing the proposition of several defined hypotheses. The available class I structural information was utilised in order to characterise highly conserved structural elements within the class I group. It was then possible to relate this data set to class I primary sequences and subsequently to the generation of a class II model. Various modelling techniques were used based on the constraints on the structural organisation that could achieve a functionally active monomer plus the set of hypotheses defined in the earlier work.

Mutagenic analysis of the HSV-1 dUTPase was then possible using the class II model as a reference. Several targets were investigated based on predicted functionally important regions. Analysis of these mutant enzymes was performed using purified recombinant HSV-1 dUTPase expressed from the T7 *E.coli* expression system. The results are discussed with regard to the evolution of structure and function in the class II enzymes.

Non-Standard Abbreviations

A	adenine
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
B.subtilis	*Bacillus subtilis*
BSA	bovine serum albumin
bp	base pair
C	cytosine or carboxy terminal end of protein
cpm	counts per minute
dAMP	2'-deoxyadenosine 5'-monophosphate
dATP	2'-deoxyadenosine 5'-triphosphate
dCMP	2'-deoxycytidine 5'-monophosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGMP	2'-deoxyguanosine 5'-monophosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTMP	2'-deoxythymidine 5'-monophosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
ddCTP	2',3'-dideoxycytidine 5'-triphosphate
ddGTP	2',3'-dideoxyguanosine 5'-triphosphate
ddTTP	2',3'-dideoxythymidine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
dUMP	2'-deoxyuridine 5'-monophosphate
dUTP	2'-deoxyuridine 5'-triphosphate
dUTPase	deoxyuridine triphosphatase
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DMSO	dimethylsulphoxide
ds	double stranded
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetic acid
EGTA	ethylene glycol-bis(β-aminoethyl ether)N,N,N',N'-tetraacetic acid
EtBr	ethidium bromide
FPLC	fast protein liquid chromatography
G	guanine
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
IPTG	isopropyl thiogalactoside

Kb	kilobase(s)
Kbp	kilobase pair(s)
KDa	kilodalton
MAb	monoclonal antibody
MCR	multiple cloning region
Mr	relative molecular mass
MW	molecular weight
N	unspecified nucleotide; amino terminal end of protein; asparagine
OD	optical density
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PMSF	phenylmethylsulphonyl fluoride
RNase A	ribonuclease A
rpm	revolutions per minute
RNA	ribonucleic acid
RT	room temperature
S.cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulphate
ss	single stranded
T	thymine
Taq	*Thermus aquaticus* DNA polymerase
TEMED	N,N,N',N' tetramethyl ethylenediamine
TK	thymidine kinase
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
ts	temperature-sensitive
μ	micro
UV	ultra-violet radiation
wt	wild type

Table of Amino Acids

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

Mutation Abbreviations

Each mutant dUTPase enzyme that was constructed was given a specific code. The first number indicates the motif involved (see section 1.4.8). The next letter/number/letter combination indicates the amino acid single letter code (above) and position of the wt residue followed by the single letter code of the mutant residue.

For example, mutant enzyme 3Y100A has the tyrosine residue, at position 100 in the motif 3 region, mutated to an alanine. An 'N' replacing the first number indicates a non-motif mutation. For example, NC76A has the cysteine mutated to an alanine at a locus outside the motif regions.

Virus Abbreviations

BHV-1	bovine herpesvirus 1
BHV-2	bovine herpesvirus 2
BHV-4	bovine herpesvirus 4
CAEV	caprine arthritis-encephalitis virus
CCV	channel catfish virus
EBV	Epstein-Barr virus
EHV-1	equine herpesvirus 1
EHV-2	equine herpesvirus 2
EHV-4	equine herpesvirus 4
EIAV	equine infectious anaemia virus
FIV	feline immunodeficiency virus
GCMV	guinea pig cytomegalovirus
HCMV	human cytomegalovirus
HHV-6	human herpesvirus 6
HHV-7	human herpesvirus 7
HIV	human immunodeficiency virus
HSV-1	herpes simplex virus type 1
HSV-2	herpes simplex virus type 2
HVS	herpesvirus saimiri
IAP-H18	hamster intracisternal A particle
KSHV	Kaposi's sarcoma associated herpesvirus (also referred to as human herpesvirus 8)
MCMV	murine cytomegalovirus
MDV	Marek's disease virus
MHV-68	murine herpesvirus 68
MMTV	mouse mammary tumour virus
MPMV	Mason-Pfizer monkey virus
PRV	pseudorabies virus
SHV-2	salmonid herpesvirus 2
SRV-1	simian type D retrovirus
VZV	varicella-zoster virus

List of Figures

Chapter 5 – Experimental Results

List of Tables

Chapter 1 -Introduction

Chapter 4 - Modelling Results

Chapter 5 – Experimental Results

Chapter 1 - Introduction

1.1 Overview

The research presented in this thesis concerns the herpes simplex virus type 1 (HSV-1) enzyme deoxyuridine triphosphatase (dUTPase). The introduction is designed to give an overview of the herpesviruses and summarise the current state of knowledge on dUTPases. The first chapters deal with background information on the more general aspects of the Herpesviridae including classification, role in human pathogenicity and replication. Further chapters are concerned specifically with the function of the dUTPase enzyme in a wide variety of organisms and in relation to potential antiviral targets and cancer therapy. The *Escherichia coli* enzyme, which forms the basis for the molecular modelling of the HSV-1 dUTPase, is dealt with in a separate chapter. The introduction concludes with details of the HSV-1 dUTPase and provides the starting point for the work presented in this thesis.

1.2 The Biology of the herpesviruses

1.2.1 Characteristics and classification of the Herpesviridae

The Herpesviridae is a large family with almost 100 members characterised to date. Herpesviruses infect a wide range of higher eukaryotic hosts and most animal species studied have yielded at least one member. The family is defined by the architecture of the virion which is composed of four morphologically distinct components termed the core, capsid, tegument and envelope (reviewed by Rixon, 1993). The virion core contains the virus genome ranging in size from 125 to 230kb and present as linear double-stranded DNA (Epstein, 1962). Initial studies based on electron microscopy described the DNA arrangement as a toroid structure around a central protein (Furlong *et al.*, 1972; Nazarian, 1974). Later studies indicate that the viral DNA is in a liquid crystalline form within the capsid (Booy *et al.*, 1991). The capsid is icosahedral in shape, approximately 100nm in diameter and consists of 162 capsomeres (Wildy *et al.*, 1960). The tegument is an amorphous proteinaceous region that surrounds the capsid and is in turn surrounded by the envelope (Roizman & Furlong, 1974). The envelope consists of a lipid bilayer derived from the host cell membranes and exhibits glycoprotein spikes on the surface (Spear & Roizman, 1972).

Historically the herpesviruses have been classified according to their biological properties and divided into three subfamilies, designated the α-herpesvirinae, β-herpesvirinae and γ-herpesvirinae (Roizman *et al.*, 1992). With few exceptions, this

classification corresponds with the genetic relationships determined by comparisons of amino acid sequences and gene organisation (McGeoch, 1989). Genomic analysis also provides a basis for assigning potential functions to homologous genes in related subfamilies (McGeoch & Davison, 1986; McGeoch *et al.*, 1993). There are now 14 complete herpesvirus sequences available (Table 1.1) which has allowed the determination of evolutionary relationships.

Virus	Size (bps)	Reference
Epstein-Barr virus	172 282	(Baer *et al.*, 1984)
Varicella-zoster virus	124 884	(Davison & Scott, 1986)
Herpes simplex virus 1	152 261	(McGeoch *et al.*, 1988a)
Human cytomegalovirus	229 354	(Chee *et al.*, 1990)
Equine herpesvirus 1	150 223	(Telford *et al.*, 1992)
Channel catfish virus	134 226	(Davison, 1992)
Herpesvirus saimiri	112 921	(Albrecht *et al.*, 1992)
Equine herpesvirus 2	184 427	(Telford *et al.*, 1995)
Human herpesvirus 6	159 321	(Gompels *et al.*, 1995)
Human herpesvirus 7	144 861	(Nicholas, 1996)
Human herpesvirus 8	140 500	(Russo *et al.*, 1996)
Murine cytomegalovirus	230 278	(Rawlinson *et al.*, 1996)
Murine herpesvirus 68	118 237	(Virgin *et al.*, 1997)
Herpes simplex virus 2	154 746	(Dolan *et al.*, 1998)

Table 1.1 Fully sequenced genomes of the herpesviruses.
The above table shows the herpesviruses that have been sequence with the respective genome size and author reference.

These sequenced genomes vary widely in a number of aspects including length (Table 1.1), gene content and overall arrangement which are discussed in Section 1.2.4. However, there is a subset of approximately 40 genes which are conserved between all the α-, β- and γ-herpesviruses implying a common ancestry (Davison & Taylor, 1987; McGeoch, 1989). In general, classification using genome analysis is in agreement with previous studies employing biological properties, although there are exceptions. EHV-2 and EHV-5 were originally classified as β-herpesviruses (Plummer *et al.*, 1969), however analysis of fragments from their genomes demonstrated their correct classification as distinct γ-herpesviruses (Telford *et al.*, 1993). Genomic analysis of MDV showed that it was more closely related to the α-herpesviruses rather than the γ-herpesviruses with which it was originally classified (Buckmaster *et al.*, 1988). Similarly, HHV-6, originally classified as a γ-herpesvirus on the basis of its biological

properties, has since been shown to be more related to the β-herpesviruses, in particular to HCMV (Gompels *et al.*, 1995; Lawrence *et al.*, 1990).

The sequencing of channel catfish virus (CCV) has yielded information on a herpesvirus from a lower vertebrate species. It was initially defined as a herpesvirus based on the morphological characteristics of its virion and as a member of the α-herpesvirinae based on its biological properties (Roizman, 1982). Analysis has shown that it does not possess the subset of genes that is common to the other herpesviruses. It has been suggested, based on this lack of similarity, that CCV may have evolved independently from the mammalian herpesviruses (Davison, 1992). Sequence data from salmon herpesvirus 2 (SalHV-2) support a relationship with CCV and it is likely that both fish viruses may have to be assigned to a separate subfamily (Bernard & Mercier, 1993). Recent investigation of the genome of salmonid herpesvirus 1 (SalHV-1) has shown that this virus also shares homology with CCV at the amino acid primary sequence level and is again unrelated to the mammalian herpesviruses supporting the classification of fish herpesviruses as a distinct subfamily (Davison, 1998).

The wealth of sequence information from complete and partially sequenced members of the Herpesviridae has allowed phylogenetic analysis. It has been found that certain genes are highly conserved between evolutionary closely related members (McGeoch, 1990a). Gene sets within each subfamily, although closely related to each other, are substantially divergent from the other subfamilies. Using various gene sets as markers for evolutionary relatedness, it has been possible to construct phylogenetic trees and infer evolutionary timescales. This analysis closely groups the herpesviruses into the three recognised subfamilies α, β, and γ. These three subfamilies are clearly defined even when using diverse gene sets such as those from the viral DNA polymerase, DNA helicase and glycoprotein gB. In actual fact, these studies have allowed substantial resolution of evolutionary relatedness beyond the three subfamily categories (Figure 1.1). Sublineages within each subfamily reveal how individual members are related to one another with far greater resolution than the biological classification system alone.
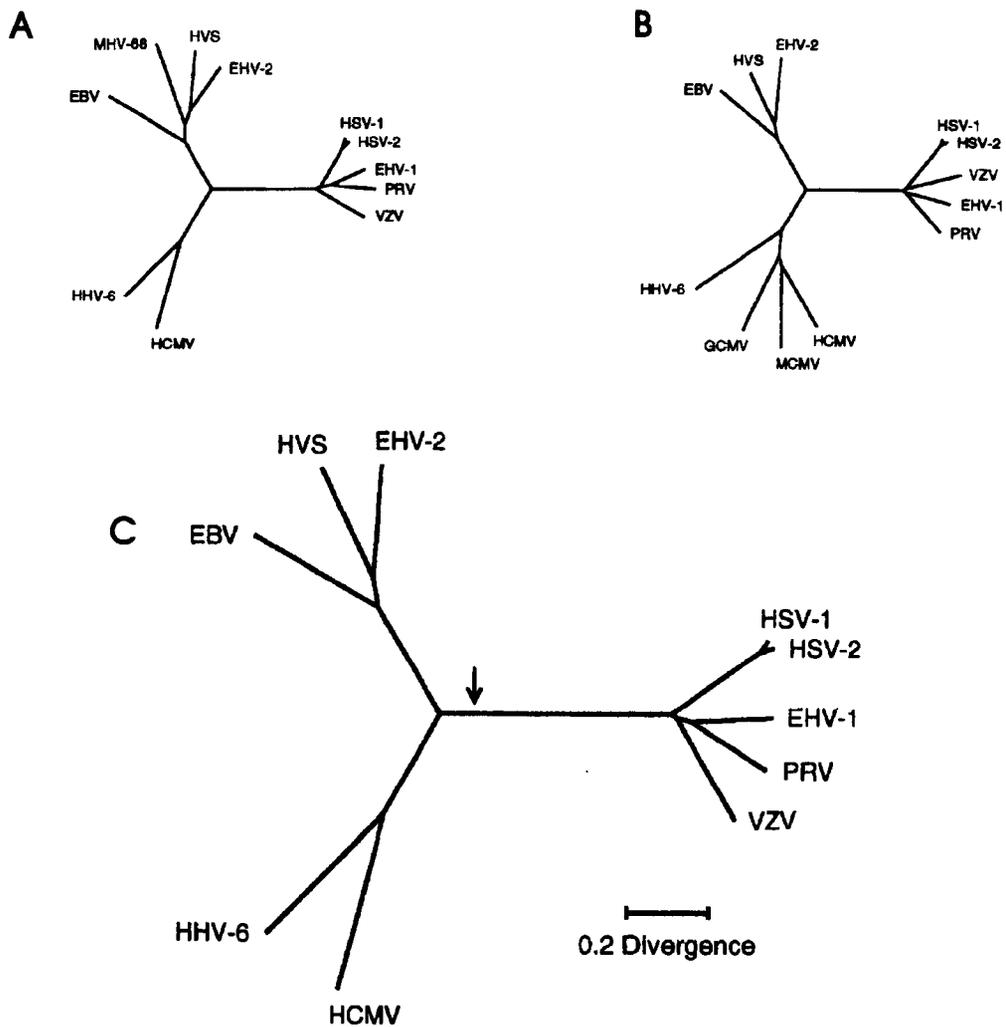
Figure 1.1 Phylogenetic distance trees from various gene sets.
Trees were derived from aligned amino acid sequence comparisons using the neighbour-joining distance method. (A) Tree derived from the DNA helicase gene set (UL5). (B) Tree derived from the DNA polymerase gene set (UL30). (C) Tree derived from a combined group of 10 members with 8 gene sets. The approximate root position is indicated by an arrow (see text below). Adapted from (McGeoch *et al.*, 1995).

Studies on the molecular phylogeny of the α-herpesviruses revealed that the branching pattern of their phylogenetic tree could be efficiently transposed onto their corresponding host tree (McGeoch & Cook, 1994). This strongly indicates that these viruses co-speciated with their host organisms. This finding makes it possible to calculate an evolutionary timescale for all the herpesviruses using the palaeontological data available from their hosts. The radiation of the three subfamilies has been suggested to occur approximately 180-200 million years ago with the major sublineages being generated about 60-80 million years ago (McGeoch *et al.*, 1995). These figures are based on the assumption that all members in the group have experienced a constant molecular clock and that there are no strong biasing effects

13

from base composition variations. The approximate root position is shown as an arrow on Figure 1.1(C). From this root position it can been seen that the initial radiation occurred between the α-herpesvirus and the precursor of the β and γ subfamilies.

Each of the three defined subfamily groupings exhibits characteristic biological properties. The members of the subfamily α-herpesviruses are characterised by variable host range, relatively short reproductive cycle (usually less than 24 hours), rapid spread and cytolytic infection of cells in culture. Latent infection is frequently established in sensory ganglia. Members of this subfamily include herpes simplex virus 1 (HSV-1), herpes simplex virus 2 (HSV-2), varicella-zoster virus (VZV), pseudorabies virus (PRV) and equine herpesviruses 1 and 4 (EHV-1 and -4).

In general the β-herpesviruses have a restricted host range with slow viral replication and lytic progression. The infected cells frequently become enlarged. Latent infections are usually established in secretory glands, lymphoreticular cells and the kidneys. Members of this subfamily include human cytomegalovirus (HCMV), murine cytomegalovirus (MCMV), human herpesvirus 6 (HHV-6) and human herpesvirus 7 (HHV-7).

The γ-herpesviruses also have a restricted host range and cannot infect experimental animals outside the family or order of their natural host. Replication occurs typically in T or B lymphocytes and latent infection is frequently established in lymphoid tissue. Members of this subfamily include Epstein-Barr virus (EBV), herpesvirus saimiri (HVS), equine herpesvirus 2 (EHV-2) and murine herpesvirus 68 (MHV-68).

1.2.2 Pathogenicity of the human herpesviruses

Herpesviruses cause a range of clinical disorders in humans and are characterised by the establishment of latent infection. The virus is generally spread by direct contact between infected individuals and uninfected individuals, particularly at mucosal tissues. Herpesvirus infection has received a higher profile in recent times due to increased numbers of immunocompromised individuals. This is mainly through drug therapy, as in transplant patients, or as a result of AIDS.

HSV-1 is the most studied herpesvirus and is widespread with over 90% of western population seropositive by the age of sixty (Nahmias et al., 1970). Infection of epithelial tissue generally causes vesicular lesions at mucosal membranes around the mouth, lips and nose. The distinction between the two serotypes HSV-1 and HSV-2 was historically on the basis that HSV-1 was responsible for the common 'cold-sore' and HSV-2 for genital lesions. However, it is noted that up to 50% of genital infections can be caused by HSV-1 (Kinghorn, 1993) and inversely up to 20% of oral infections

by HSV-2, therefore this distinction is no longer clear cut (Wiedbrauk & Johnston, 1993).

In general, HSV-1 establishes latency in the trigeminal ganglia and HSV-2 in the sacral ganglia. Both serotypes can periodically reactivate causing recurrent lytic infection and leading to lesions at peripheral sites (reviewed by Nash & Lohr, 1992). Herpes simplex viruses can also cause conjunctivitis, herpetic whitlow and ocular keratitis, particularly in immunocompromised individuals. Clinically, the most severe situation is infection of the central nervous system (CNS) causing acute necrotising encephalitis (Corey & Spear, 1986). Neonatal infection also has a high mortality rate and is usually due to a primary HSV-2 infection in the mother where no maternal antibody is present for protection (Sullivan-Bolyai et al., 1983).

VZV is the causative agent of chickenpox (varicella), a disease resulting from primary infection, generally in children. Following infection, the virus spreads through viraemia and ascends to the dorsal root ganglia of sensory nerves supplying the affected skin. Symptoms are in the form of a rash 14-15 days after infection and are often accompanied by fever. Reactivation of latent virus from neural ganglia in later life results in a localised vesicular condition known as shingles (herpes zoster). This condition is often accompanied by severe pain which can persist for months after the lesions have healed (Gelb, 1990). In contrast to HSV, VZV reactivation either does not occur in most individuals or is limited to a single recurrence of infection (Straus, 1989).

HCMV affects the majority of the population asymptomatically although in certain cases HCMV infection results in the development of a mononucleosis syndrome. Primary infection is characterised by cytomegaly (the enlargement and fusion of macrophages). In immunocompromised individuals HCMV can cause a wide range of diseases in many organs including the lungs, gastrointestinal tract and CNS (Britt & Alford, 1993; Gallant et al., 1992). HCMV respiratory infection in AIDS patients can lead to potentially fatal pneumonia (Meyers et al., 1986). Retinitis is the most common HCMV disease of the nervous system and affects up to 20-25% of long-lived AIDS patients. HCMV is also associated with neurological damage in neonates (Alford et al., 1990).

EBV is usually acquired during childhood when it is generally asymptomatic, however primary infection in later life results in infectious mononucleosis, commonly known as glandular fever (Niederman et al., 1968). EBV has a strong association with a variety of malignant diseases. Nasopharyngeal carcinoma (NPC) was first linked to EBV on the basis of epidemiological studies showing elevated levels of antiviral antibody titres in NPC patients (Old et al., 1966; Mansoor et al., 1997). There is also an association with Hodgkin's disease after the key finding that monoclonal viral genomes could be detected in tumour biopsies and that EBV DNA was localised to

malignant cells (Anagnostopoulos *et al.*, 1989; Weiss *et al.*, 1987). Burkitt's lymphoma has also been linked to EBV infection (Gunven *et al.*, 1970; Khanna *et al.*, 1997; Magrath *et al.*, 1992). EBV is associated with various T cell and B cell lymphomas (Jones *et al.*, 1988; Su *et al.*, 1991). In AIDS patients, EBV infection can cause oral hairy leukoplakia (Greenspan *et al.*, 1985).

HHV-6 is known to affect up to 90% of the population although comparatively little is know about the virus. Primary infection generally occurs in infants and can be asymptomatic or cause the skin rash roseola (exanthem subitum) (Yamanishi *et al.*, 1990). The significance of HHV-6 as a pathogen is unclear but its tropism for CD4+ lymphocytes has led to speculation that it may have an adjunct role in AIDS (Takahashi *et al.*, 1989; Levy *et al.*, 1990).

HHV-7 was isolated from peripheral blood lymphocytes (Frenkel *et al.*, 1990) and has been associated with roseola in infants (Ablashi *et al.*, 1995). HHV-7 shows limited hybridisation with probes derived from HHV-6 but has significant antigenic dissimilarity to allow seroepidemiological discrimination between the two viruses (Wyatt *et al.*, 1991).

An eighth human herpesvirus, related to HVS and now classified as a γ-herpesvirus, has been associated with the neoplasm, Kaposi's sarcoma (Chang *et al.*, 1994; Chang & Moore, 1996; Moore *et al.*, 1996). The virus has been given the trivial descriptive name KS-associated herpesvirus (KSHV) but is likely to be given the formal classification of human herpesvirus 8.

1.2.3 Herpesvirus replication

The lytic replication cycle of HSV-1 can be divided into attachment of the virus particle, entry into the host cell, DNA replication, and virion production. Initial events involve the attachment of the virus to the cell and fusion of the viral envelope to the plasma membrane. There have been ten surface glycoproteins identified in HSV-1, five of which are dispensable in cell culture. It is likely that HSV-1 can utilise more than one attachment pathway and there appears to be more than one viral protein-cell interaction. Studies have been complicated by the fact that *in vivo* HSV-1 infects two very different cell types, epithelial and neuronal. Unlike the continuous cell lines used for many of the attachment studies, both of these cell types are polarised *in vivo*, sorting membrane and secreted proteins to different surfaces (Dotti & Simons, 1990; Rodriguez-Boulan & Pendergast, 1980).

Heparan sulphate has been identified as a major component of HSV-1 cell surface binding although even on removal of the molecule from cells there is still some infectivity (Shieh *et al.*, 1992; Shieh & Spear, 1994). It is likely that heparan sulphate is one of a number of receptors or acts as a cofactor in glycoprotein binding. Following

attachment, entry into the cell is probably mediated by fusion of the viral envelope and the plasma membrane. Data from analysis of mutated viruses suggest that this process requires participation of several gene products (Brandimarti *et al.*, 1994). Infection results in an immediate drop in host DNA and protein synthesis (Read & Frenkel, 1983; Roizman & Roane, 1964). Following entry, viral capsids are transported to the nuclear pore where viral DNA is released into the nucleus (Batterson *et al.*, 1983). This process is thought to be mediated by the cellular cytoskeleton (Kristensson *et al.*, 1986).

HSV gene expression occurs in a coordinately regulated cascade with genes classified as immediate-early (IE), early (E) and late (L) (Clements *et al.*, 1977) or α, β and γ (Honess & Roizman, 1975). Genes are transcribed in the nucleus by the cellular RNA polymerase II with the participation of viral factors (Ben-Zeev & Becker, 1977; Costanzo *et al.*, 1977). The mRNAs that are produced are capped at the 5'-terminus, polyadenylated at the 3'-terminus and internally methylated (Bachenheimer & Roizman, 1972; Moss *et al.*, 1977; Silverstein *et al.*, 1973). Translation occurs on both free and bound polyribosomes producing the viral proteins. Modifications to these proteins can then occur such as cleavage, poly(ADP) ribosylation, phosphorylation, sulphation and glycosylation (reviewed by Roizman & Sears, 1990).

The HSV-1 tegument protein, Vmw65, is necessary for the transactivation of the five IE genes (Campbell *et al.*, 1984). In a cascade mechanism, at least three of these IE gene products are required for expression of E and L genes. The synthesis of the IE proteins reaches a peak at approximately 2 to 4 hours post infection although they continue to be produced at non-uniform rates until late on in infection (Ackermann *et al.*, 1984). The E polypeptides reach peak synthesis at about 5 to 7 hours post infection and include most of the enzymes involved in nucleotide metabolism (Honess & Roizman, 1975). Viral DNA synthesis begins shortly after the E genes are expressed. The L genes have been further classified into leaky-late (γ_1) and true-late (γ_2) (Holland *et al.*, 1980). The γ_1 genes are expressed at a low level before the onset of viral DNA replication, reaching maximal levels after replication. The γ_2 genes are expressed exclusively after the onset of DNA replication. The process of DNA replication is discussed further in Section 1.3.1.

The L genes code primarily for the structural proteins of the virion including the capsid. Much of the understanding of capsid assembly comes from the analysis of the three distinct capsid forms found in HSV-1 infected cells (Gibson & Roizman, 1972). Current classification defines the capsid types as A, empty, B, intermediate, containing a proteinaceous scaffolding and C, containing DNA. B capsids are known to be the progenitors of A and C capsids. It is likely that A capsids are formed as the result of abortive packaging of DNA. Capsids, which are made from seven viral proteins, are

assembled in the infected cell nucleus (reviewed by Rixon, 1993). The replicated concatemeric DNA is cleaved at the 'a' sequence (Section 1.2.4) into genome length molecules which are then packaged into the pre-formed capsids. Studies with L gene mutant viruses indicate that the processes of cleavage and packaging are tightly coupled (reviewed by Roizman & Sears, 1990).

The final events leading to the maturation and release of the virion are not well characterised. Two pathways relating to the formation of the tegument and envelope have been proposed (reviewed by Rixon, 1993). The capsids may bud through the inner nuclear membrane, fuse with the outer nuclear membrane and then enter the cytoplasm. This would allow the tegument to be assembled in the cytoplasm and the mature virion would be produced by budding through the plasma membrane. An alternative model has been described where the tegument is formed in the nucleus. This would require the capsid to bud through the inner nuclear membrane and enter the cytoplasm in a vacuole formed by the outer nuclear membrane. Fusion of the vacuole and plasma membrane would allow virion egress. Although the sites of tegumentation and envelopment have not been defined it is seems likely that the virion passes through Golgi derived cytoplasmic vesicles on the journey from the nucleus, through the cytoplasm and finally to the plasma membrane.

1.2.4 Structure and content of herpesvirus genomes

The herpesvirus genomes are diverse in a number of aspects. They vary in length (from about 125 to 240kbp), number of genes (from around 70 to 200) and base composition (from 31 to 75% G+C). The herpesvirus genomes also exhibit variation in the pattern of unique and repeated sequence elements. Figure 1.2 shows a diagrammatic representation of the human herpesvirus genome arrangements as described by McGeoch (1989).
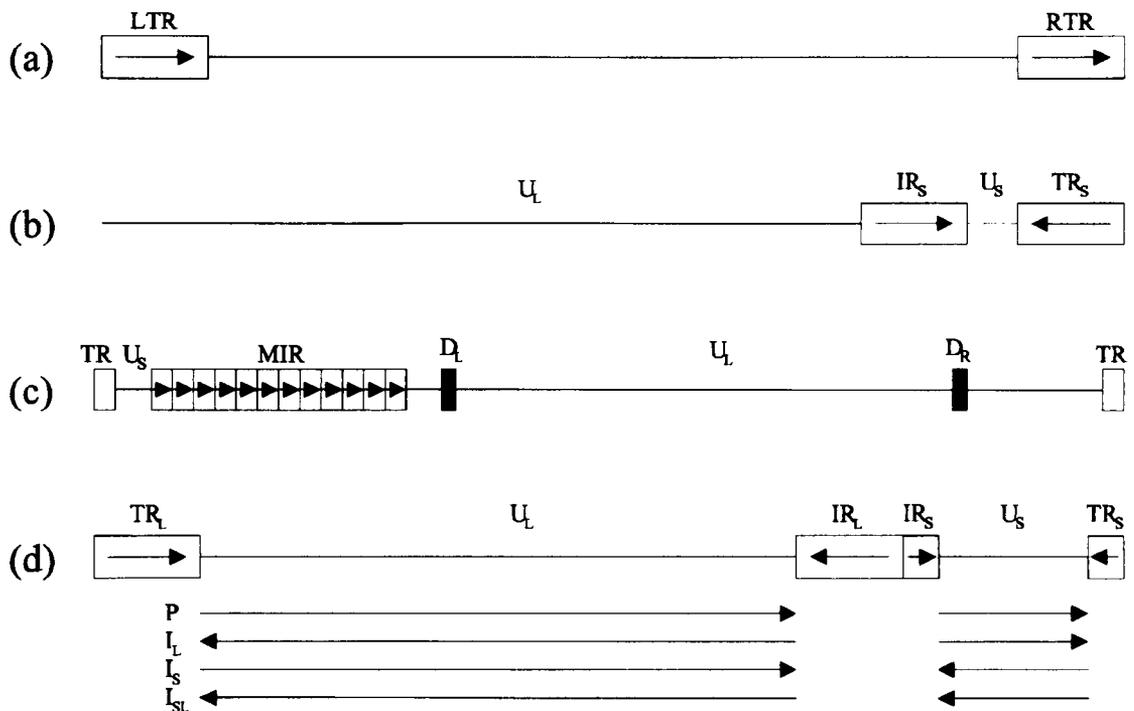
Figure 1.2 Structure of the genomes of the human herpesviruses (not to scale).
Unique regions are shown as horizontal lines and repeat regions as rectangles. The orientation of each repeat is shown by an arrow. The following abbreviations have been used: LTR (left terminal repeat), RTR (right terminal repeat), MIR (major internal repeat), D_L (direct repeat left), D_R (direct repeat right), TR_L (terminal repeat long), U_L (unique long), IR_L (internal repeat long), IR_S (internal repeat short), U_S (unique short) and TR_S (terminal repeat short). The orientations of the 4 possible isomers of the group (d) genome are represented below as long arrows. The four isomers are labelled P (prototype), I_L (inversion of L), I_S (inversion of S) and I_{LS} (inversion of both L and S). Diagram was adapted from McGeoch (1989).

The group (a) genome arrangement is the simplest consisting of a unique sequence flanked by large direct repeats. This type of genome structure has been described for HHV-6 and also for the non-human herpesviruses, CCV, EHV-2 and MCMV. The group (b) arrangement is characteristic of the VZV genome and consists of two unique regions with the U_L region flanked by inverted repeats. VZV has four sequence-orientation isomers (not shown), two of which are 20-fold more abundant than the other two (Davison, 1984). Other α-herpesvirus such as PRV and EHV-1 also have this arrangement. In these examples the two orientations of the U_S region are present in equimolar amounts but the U_L region is found completely (EHV-1) or predominately (VZV, PRV) in a single orientation. The group (c) arrangement represents the structure of the EBV genome where the DNA termini are formed by several direct repeats (TR). The genome also carries a set of major internal repeat elements (MIR) which vary in copy number. There are two other internal repeats, D_L and D_R, which are almost identical and lie in the same orientation.

The group (d) arrangement represents the genome structure of HSV-1 and HSV-2 and is also found in BHV-2 and HCMV. Two unique sequences, U_L and U_S are flanked by pairs of oppositely orientated repeat elements (TR_L and IR_L) and (IR_S and TR_S). The R_L repeats and the R_S repeats are distinct apart from a short (400bp) direct repeat element at the genome termini named the 'a' sequence. In HSV-1 this 'a' sequence is also found between the junction of the L and S segments in the opposite orientation to the genome termini copies (Wadsworth et al., 1975; Wagner & Summers, 1978). The recombination between the inverted repeats results in the formation of four sequence-orientation isomers. Each isomer is generated by inversion of the L and S components and are found in equimolar quantities in DNA preparations of HSV-1. The direction of the unique sequences for each isomer is shown in the diagram.

HSV-1 is the best studied of the herpesviruses and is the main focus of analysis in this thesis. The complete sequence of the HSV-1 strain $17syn^+$ has been determined and the total size is 152261bp with a G+C content of 68% (McGeoch, 1997; McGeoch et al., 1988a). The total sequence length varies due to small reiterations and the presence of a variable copy number of the 'a' sequence. So far approximately two-thirds of the ORFs have had functions assigned to their products (Table 1.2 and Figure 1.3).

The diagram in Figure 1.3 representing the HSV-1 genome was complied by McGeoch et al. (1988a). Since that time additional ORFs have been reported including UL26.5, UL49A, RL1 and US8A. UL26.5 encodes an internal capsid protein required for the formation of the capsid shell around the scaffold (Liu & Roizman, 1991; Kennard et al., 1995). UL49A (also referred to as UL49.5) encodes an envelope protein (Barker & Roizman, 1992; Barnett et al., 1992) and RL1 encodes the neurovirulence factor, ICP34.5, first discovered in HSV-1 strain F (Chou & Roizman, 1986; 1990) then confirmed in strain 17 (Dolan et al., 1992; McKie et al., 1994). US8A lies between US8 and US9 and appears to be a γ gene which is found in the nucleoli of HSV-1 infected cells (Georgopoulou et al., 1995). The position of these genes is given in Table 1.2.

There have been further potential ORFs reported including ORF-P, UL8.5, UL9.5 and UL43.5 although these are much less validated. ORF-P lies antisense to gene RL1 and although there have been experiments with over-producing virus mutants, the function is not clear (Lagunoff & Roizman, 1995; Lagunoff et al., 1996). UL8.5 and UL9.5 ORFs were reported during a mapping study of HSV-1 although no function has been attributed to either of the potential genes (Baradaran et al., 1994). UL43.5 has been reported lying antisense to gene UL43 and encoding a protein dispensable in cell culture which colocalises with capsid proteins (Ward et al., 1996).
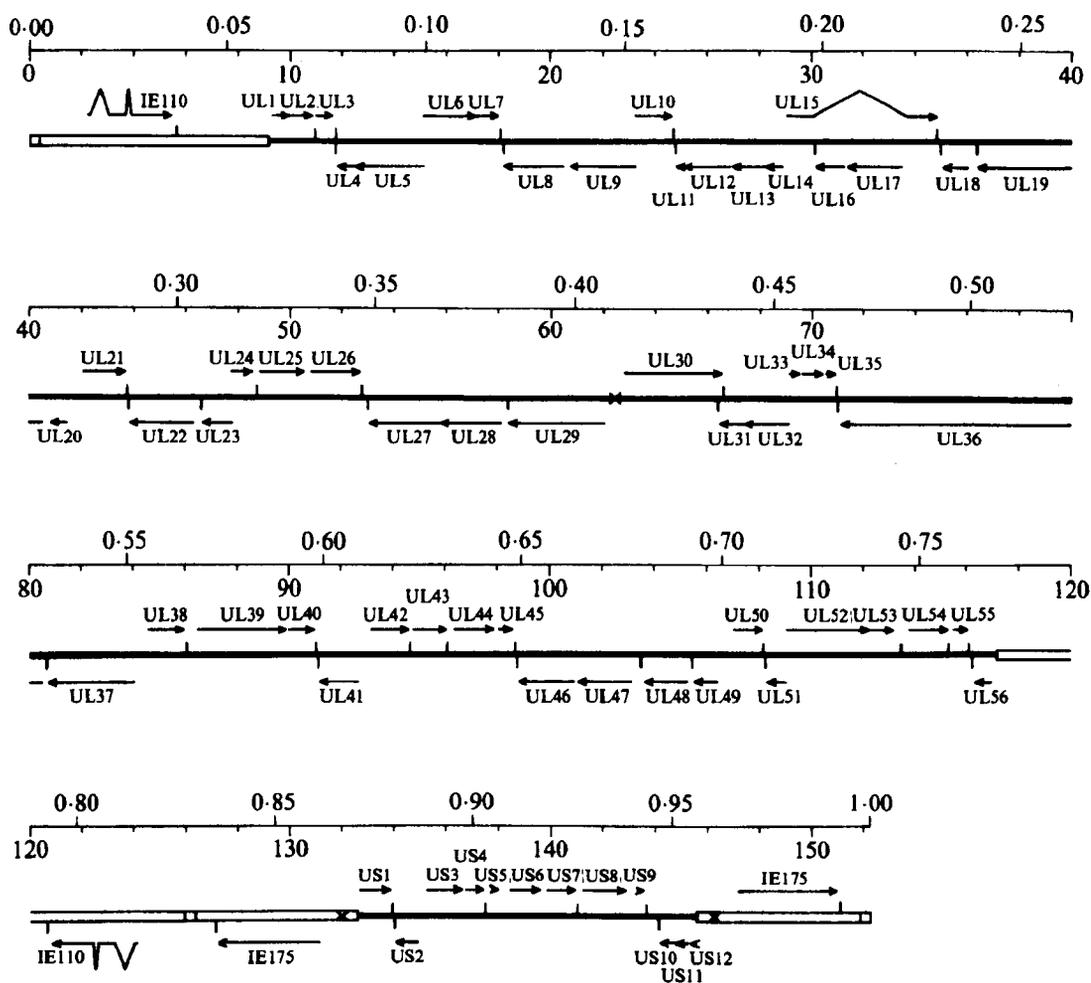
Figure 1.3 The layout of genes in the genome of HSV-1.

The HSV-1 genome is shown on four successive lines. Unique regions are represented by solid lines and major repeat elements as open boxes. The lower scale represents kilobases, numbered from the left terminus, and the upper scale represents fractional map units. The sizes and orientations of the proposed ORFs are shown by arrows. Overlaps of adjacent, similarly orientated ORFs are not shown explicitly. Locations of proposed transcriptional polyadenylation sites are indicated as short vertical bars. Locations of origins of DNA replication are shown as X. In the U_L region, on the first three lines, genes UL1 to UL56 are labelled. In the US region, on the bottom line, genes US1 to US12 are labelled. The locations of introns in the coding regions of gene UL15 and the two copies (TR_L and IR_L) of the IE110 gene are indicated. Copied from McGeoch *et al.* (1988a).

Table 1.2

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
RL1	513	1256	248	26194	NE	Neurovirulence factor (ICP34.5)
RL2			775	78452	NE	IE protein; modulator of cell state and gene expression
exon 1	2261	2317	19			(ICP0, Vmw110)
exon 2	3083	3749	222			
exon 3	3886	5486	534			
LAT C					NE	Latency-associated transcript; probably not protein-coding
(U$_L$ starts at 9213)						
UL1	9337	10008	224	24932	E	Glycoprotein L;complexes with glycoprotein H (UL22)
UL2	9884	10885	334	36326	NE	Uracil-DNA glycosylase
UL3	10957	11661	235	25607	NE	Function unknown
UL4 C	12422	11826	199	21516	NE	Function unknown
UL5 C	15131	12486	882	98710	E	Component of DNA helicase-primase complex; possesses helicase motifs
UL6	15130	17157	676	74087	E	Minor capsid protein
UL7	17135	18022	296	33057	E?	Function unknown
UL8 C	20476	18227	750	79921	E	Component of DNA helicase-primase complex
UL9 C	23259	20707	851	94246	E	Ori-binding protein essential for DNA replication
UL10	23204	24622	473	51389	NE	Virion surface glycoprotein M
UL11 C	25091	24804	96	10486	NE	Myristylated tegument protein; role in virion envelopment
UL12 C	26887	25010	626	67503	(E)	Deoxyribonuclease; role in maturation/packaging of DNA
UL13 C	28502	26949	518	57193	NE	Tegument protein; protein kinase
UL14 C	28915	28259	219	23454	E	Function unknown
UL15			735	80918	E	Role in DNA packaging; putative terminase component
exon 1	29020	30048	343			
exon 2	33635	34810	392			
UL16 C	31295	30177	373	40440	NE	Function unknown
UL17 C	33497	31389	703	74577	E	Function unknown
UL18 C	36051	35098	318	34268	E	Capsid protein (VP23); component of intercapsomeric triplex
UL19 C	40528	36407	1374	149075	E	Major capsid protein (VP5); forms hexons and pentons
UL20 C	41488	40823	222	24229	E/NE	Integral membrane protein; role in egress of nascent virions; host range phenotype; syn locus
UL21	42074	43678	535	57638	NE	Tegument protein
UL22 C	46382	43869	838	90361	E	Virion surface glycoprotein H; complexes with glycoprotein L (UL1); role in cell entry
UL23 C	47802	46675	376	40918	NE	Thymidine kinase
UL24	47737	48543	269	29474	NE	Function unknown; syn locus
UL25	48813	50552	580	62664	E	Capsid-associated tegument protein
UL26	50809	52713	635	62466	E	Protease, acts in virion maturation; N-terminal portion is capsid protein (VP24)
UL26.5	51727	52713	329	33758	(E)	Internal protein of immature capsids (VP22a); processed by UL26 protease

Table 1.2 continued.

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
UL27 C	55794	53083	904	100287	E	Virion surface glycoprotein B; role in cell entry; syn locus
UL28 C	58159	55805	785	85573	E	Role in DNA packaging
UL29 C	62053	58466	1196	128342	E	Single-stranded DNA-binding protein (ICP8)
(Centre of ori$_L$ is at 62475/62476)						
UL30	62807	66511	1235	136413	E	Catalytic subunit of replicative DNA polymerase; complexes with UL42 protein
UL31 C	67379	66462	306	33951	E	Function unknown
UL32 C	69162	67375	596	63946	E?	Function unknown
UL33	69161	69550	130	14436	E	Role in DNA packaging
UL34	69633	70457	275	29788	E?	Membrane-associated phosphoprotein; substrate for US3 protein kinase
UL35	70566	70901	112	12095	E?	Capsid protein (VP26); located on tips of hexons
UL36 C	80543	71052	3164	335841	E	Very large tegument protein
UL37 C	84084	80716	1123	120549	E?	Tegument protein
UL38	84531	85925	465	50260	E	Capsid protein (VP19C); component of intercapsomeric triplex
UL39	86444	89854	1137	124043	E/NE	Ribonucleotide reductase large subunit (ICP6, Vmw136, R1)
UL40	89926	90945	340	38017	E/NE	Ribonucleotide reductase small subunit (Vmw38, R2)
UL41 C	92637	91171	489	54914	NE	Tegument protein; host shut-off factor
UL42	93113	94576	488	51156	E	Subunit of replicative DNA polymerase; increases processivity; complexes with UL30 protein
UL43	94748	96049	434	44905	NE	Function unknown; probable integral membrane protein
UL44	96311	97843	511	54995	NE	Virion surface glycoprotein C; role in cell entry
UL45	98032	98547	172	18178	NE	Tegument/envelope protein
UL46 C	100952	98799	718	78239	NE	Tegument protein; modulates IE gene transactivation by UL48 protein
UL47 C	103116	101038	693	73812	NE	Tegument protein; modulates IE gene transactivation by UL48 protein
UL48 C	105079	103610	490	54342	E	Tegument protein; transactivates IE genes (VP16, Vmw65, α-TIF)
UL49 C	106391	105489	301	32252	NE?	Tegument protein
UL49A C	106993	106721	91	9201	NE?	Envelope protein disulphide-linked to tegument
UL50	107010	108122	371	39125	NE	Deoxyuridine triphosphatase
UL51 C	109011	108280	244	25468	(E)	Function unknown
UL52	109048	112221	1058	114416	E	Component of DNA helicase-primase complex
UL53	112179	113192	338	37570	(E)	Glycoprotein K
UL54	113734	115269	512	55249	E	IE protein; post-translational regulator of gene expression (ICP27, Vmw63)
UL55	115496	116053	186	20491	NE	Function unknown
UL56 C	116925	116224	234	25319	NE	Function unknown
(IR$_L$ starts at 117160)						
LAT					NE	Latency-associated transcript; probably not protein-coding

Table 1.2 continued.

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
RL2			775	78452	NE	IE protein; modulator of cell state and gene expression (ICP0, Vmw110)
exon3 C	122485	120885	19			
exon2 C	123288	122622	222			
exon1 C	124110	124054	534			
RL1 C	125858	125115	248	26194	NE	Neurovirulence factor (ICP34.5)

(Left end of a' sequence is at 125972)
(Internal c sequence starts at 126373)

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
RS1 C	131128	127235	1298	132835	E	IE protein; transcriptional regulator (ICP4, Vmw175)

(Centre of oriS is at 131999)
(US starts at 132605)

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
US1	132644	133903	420	46521	E/NE	IE protein; function unknown (ICP22, Vmw68); host range phenotype
US2 C	134928	134056	291	32468	NE	Function unknown
US3	135222	136664	481	52831	NE	Protein kinase; phosphorylates UL34 protein
US4	136744	137457	238	25236	NE	Virion surface glycoprotein G
US5	137731	138006	92	9555	NE	Proposed glycoprotein J
US6	138419	139600	394	43344	E	Virion surface glycoprotein D; role in cell entry
US7	139785	140954	390	41366	NE	Virion surface glycoprotein I; complexed with glycoprotein E (US8) in Fc receptor
US8	141243	142892	550	59090	NE	Virion surface glycoprotein E; complexed with glycoprotein I (US7) in Fc receptor
US8A	142744	143220	159	16801	NE	Function unknown
US9	143313	143582	90	10026	NE	Tegument protein
US10 C	145095	144160	312	34053	NE	Virion protein
US11 C	145246	144764	161	17756	NE	Virion protein; ribosome-associated in infected cell
US12 C	145577	145314	88	9792	NE	IE protein; inhibitor of peptide transport by TAP and of antigen presentation(ICP47, Vmw12)

(TRS starts at 145585)
(Centre of oriS is at 146235)

Gene	Start	Stop	Codons	Mr	Status	Function of Protein
RS1	147105	150998	1298	132835	E	IE protein; transcriptional regulator (ICP4, Vmw175)

(Last nucleotide is 152261)

Table 1.2 Features of the genes of herpes simplex virus type 1 (strain 17)

Locations of protein-coding regions are given from the first residue of the translation initiation codon to the last residue of the last coding codon, omitting the stop codon. Leftward oriented genes are marked C. Bold genes are conserved in the three herpesvirus subfamilies. The status of each gene in cell culture is indicated: E = essential, NE = non-essential, E? = probably essential, NE? = probably non-essential, (E) = a mutant is viable, but very disabled, E/NE = non-essential under certain conditions, ? = unknown. E genes are those for which mutants have been constructed that require complementing cell lines or those for which attempts to generate insertion mutants using the cosmid system have yielded only wild type. E? genes are those where attempts to derive a mutant by plasmid recombination have failed. NE? genes are those where data are available for alphaherpesviruses other than HSV-1. IE = immediate early.

This listing was compiled by A.J. Davison with help from C. MacLean and D.J. McGeoch. The table was updated by A. Dolan in March 1997.

From Figure 1.3 it can be seen that the genes are arranged about equally on the two coding strands. There are a number of overlaps between coding regions in different reading frames. Genes are densely packed (89% of U_L codes for protein), and in many cases transcriptional control elements overlap with the coding regions of adjacent genes (Wagner, 1985). The majority of genes have their own promoters and are expressed as single exons although groups of adjacent genes in the same orientation may share a single polyadenylation site downstream of the most 3' gene (Wagner, 1985). Genes contained in the repeat elements, such as IE110 and IE175, are present in two copies in the genome (Perry *et al.*, 1986; Rixon & Clements, 1982). The HSV-1 genome has few spliced genes which include IE110 (Perry *et al.*, 1986), US1 and US12 (Rixon & Clements, 1982) and UL15 (Costa *et al.*, 1985; McGeoch *et al.*, 1988a).

1.3   Replication of the herpesvirus genome

1.3.1   HSV DNA replication

Herpesviruses and other double stranded DNA viruses such as poxviruses, adenoviruses, papovaviruses, and iridoviruses, all exploit replication mechanisms which are utilised by the host cell. This is in contrast to the RNA viruses whose replication requires processes not available in the uninfected cell. Despite this, most DNA viruses encode much of their own replication machinery even though many of these viruses (herpes-, adenoviruses) replicate in the nucleus where the cellular processes are localised. The dsDNA viruses have a wide range of genome sizes from about 5kbp in polyomaviruses up to 300kbp in poxviruses. The number of virally encoded proteins bears a direct relation to the dependence the virus has on the host cellular functions for successful replication. The herpesviruses are at the larger end of the dsDNA virus scale in terms of genome size, and encode a substantial number of their own functions (Table 1.2).

There are two mechanisms by which the infecting virus can provide the functions required for its replication. The virus can utilise the host enzymes and impart a degree of regulation over them. However, even if there is a host enzyme supplying the basic function required, its properties may not be optimally adapted to the process of viral replication. Alternatively the virus can encode its own set of enzymes. A combination of regulating host enzymes and inducing viral enzymes ensures that the virus carries only those proteins which are advantageous for its own replication. The set of virally encoded enzymes is not constant throughout the Herpesviridae as shown in Table 1.3. We are principally concerned with the genes involved in DNA replication

and those required for nucleotide provision. Each of the enzymes listed in Table 1.3 are dealt with in relation to HSV-1 and the homologues in the other subfamilies.

Enzyme/protein	HSV-1 (α)	VZV(α)	HCMV (β)	EBV (γ)
Replication:				
DNA polymerase (catalytic)	UL30	28	UL54	BALF5
(accessory)	UL42	16	UL44	BMRF1
helicase/primase complex (I)	UL5	55	UL105	BBLF4
(II)	UL8	52?	UL102?	BBLF3?
(III)	UL52	6	UL70	BSLF1
DNase	UL12	48	UL98?	BGLF5
DNA repair:				
uracil DNA glycosylase	UL2	59	UL114?	BKRF3?
Nucleotide provision:				
thymidine kinase	UL23	36	none	BXLF1
ribonucleotide reductase (L)	UL39	19	UL45	BORF2
ribonucleotide reductase (S)	UL40	18	none	BaRF1
deoxyuridine triphosphatase	UL50	8	UL72?	BLLF2
thymidylate synthase	none	13	none	none

Table 1.3 Homologues of virally encoded enzymes between various herpesviruses.
The HSV-1 homologues in VZV, HCMV and EBV which show a low level of similarity are marked '?'. Details of each enzyme are given in the text below. Table was adapted from J.M. Morrison (1991).

1.3.1.1 DNA replication

HSV-1 DNA replication occurs by a rolling circle mechanism and head-to-tail concatemers accumulate in the nucleus (Roizman & Sears, 1990). There are two identical origins of replication, ori_{S1} and ori_{S2}, which map to the R_S inverted repeat of the S segment of the genome (Stow, 1982; Stow & McMonagle, 1983) (Figure 1.2). A third origin of replication, ori_L, maps to the U_L segment of the genome. Replication of the genome is dependent on seven major gene products as determined by the plasmid replication assay of Challberg (Challberg, 1986; McGeoch *et al.*, 1988b; Wu *et al.*, 1988). The replication enzymes of HSV-1 are generally the most highly characterised of all the Herpesviridae members.

DNA polymerase

HSV-1 DNA polymerase was one of the first animal virus induced enzymes to be discovered (Keir & Gold, 1963) and has been studied extensively particularly in relation to antiviral therapy (see Section 1.4.5). The DNA polymerase is made up of a

catalytic subunit, the UL30 product, and an accessory subunit, the UL42 product (reviewed by Challberg, 1991).

Sequence analysis of the UL30 gene predicted a protein product of 1235 residues and approximately 136kDa (Gibbs *et al.*, 1985; Quinn & McGeoch, 1985). The DNA polymerase activity from the UL30 ORF was found to be sufficient for catalysis when analysed by various methods (Dorsky & Crumpacker, 1988; Haffey *et al.*, 1988). The UL30 protein also has a proof-reading 3'-5' exonuclease activity (Knopf, 1979) and a 5'-3' exonuclease activity. The later function can act as an RNase H presumably removing RNA primers (Crute *et al.*, 1989). The product of the UL42 gene, a dsDNA binding protein, acts as an accessory function and has been reported to stimulate the DNA polymerase activity (Gallo *et al.*, 1988; Gottlieb *et al.*, 1990).

DNA polymerase activity as been described in other α-herpesviruses (including PRV, EHV-1 and VZV), in the β-herpesvirus, HCMV, and also in the γ-herpesvirus EBV (reviewed by Morrison, 1991).

Helicase/primase complex

The helicase unwinds the dsDNA in an ATP-dependent reaction and the primase catalyses the synthesis of the oligoribonucleotide which initiates DNA replication on the lagging strand. The primase and helicase complex of HSV-1 is made up from the products of UL5, UL8 and UL52 (Crute *et al.*, 1989). The presence of the UL8 protein appears to be unnecessary for either the helicase or ATPase activities although it may have a role in the nuclear localisation of the complex and the binding of HSV-1 DNA polymerase (Barnard *et al.*, 1997; Calder & Stow, 1990; Marsden *et al.*, 1997). Specific binding to the origin is accomplished by the product of UL9 and the UL29 gene product binds single stranded DNA (Elias *et al.*, 1986; Powell *et al.*, 1981). Homologues of the HSV-1 helicase/primase complex have been found in VZV although the UL8 homologue, gene 52, has homology of only 28% (Davison & Scott, 1986). HCMV homologues for UL5 and UL52 have been assigned to gene 105 and gene 70 respectively (Chee *et al.*, 1990). The homologue of HSV-1 UL8, gene 102, is based on genome position only although it is similar in length. A similar situation exists with EBV where there are strong homologues of HSV-1 UL5 and UL52 but only a positional homologue of UL8 (McGeoch, 1989).

DNase

The role of DNase in the viral life cycle was not well understood until recently. The enzyme is also referred to as an alkaline exonuclease due to its optimal activity at a high pH. The DNase of HSV-1 can hydrolyse both ssDNA and dsDNA and acts mainly on the 3'-terminus but can also act on the 5'-terminus. It was originally proposed that

the hydrolysis of cellular DNA helped provide the virus with a source of preformed nucleotides or had a role in the cleavage of viral DNA concatemers (Hoffmann, 1981).

Subsequent analysis has shown that DNase has a role in HSV-1 capsid egress. An HSV-1 DNase deletion mutant was found to induce almost wild type levels of viral DNA but was deficient in the production of infectious virions (Shao *et al.*, 1993). HSV-1 DNase has been shown to be essential *in vivo* and it has been proposed that it plays a role in the processing of complex DNA intermediates (Goldstein & Weller, 1998). The HSV-1 DNase is encoded on gene UL12 producing a protein of 626 amino acids and 67.5kDa (McGeoch *et al.*, 1986b).

A UL12 homologue with 29% sequence homology, gene 48, is induced in VZV (Cheng *et al.*, 1980). The UL98 gene of HCMV has only low sequence similarity to HSV-1 but has recently been shown to allow functional complementation of the HSV-1 UL12*lacZ* deletion mutant in tissue culture (Gao *et al.*, 1998). EBV has been shown to induce a DNase which is similar to the HSV-1 enzyme even though the EBV gene BGLF5 exhibits a low degree of sequence similarity to HSV-1 UL12 (Baylis *et al.*, 1989; Cheng *et al.*, 1980; McGeoch *et al.*, 1986b).

1.3.1.2 DNA repair

Uracil DNA glycosylase (UDGase)

UDGase catalyses the removal of uracil from DNA. Its role in viral replication is described in Section 1.4.1. HSV-1 gene UL2 encodes a UDGase which has been shown to be dispensable for growth in cell culture (Worrad & Caradonna, 1988; Mullaney *et al.*, 1989). VZV gene 59 has a 39% sequence homology with the HSV-1 gene. No UDGase activity has been reported for HCMV or EBV although gene 114 and gene BKRF3 show limited homology to the HSV-1 enzyme respectively (McGeoch, 1989).

1.3.1.3 Nucleotide provision

Successful replication of the viral DNA in the host cell requires an additional group of virally encoded proteins involved in nucleotide metabolism. The substrates for the DNA polymerase reaction are the four deoxyribonucleoside triphosphates (dATP, dGTP, dCTP and dTTP) which must be supplied in sufficient quantity to allow efficient DNA synthesis, cell growth and metabolism (Reichard, 1988). In normal cellular DNA replication, small imbalances in the levels of each dNTP can lead to DNA mutation and fragmentation. Furthermore, cellular levels of dNTPs are cell-cycle regulated and only reach their maximal level during the S phase. The provision of virally encoded nucleotide metabolism enzymes may be necessary to allow

independence from the cell-cycle (Morrison, 1991). The role of each of the following enzymes in the *de novo* production of dNTPs can be seen in Figure 1.4.

<u>Thymidine kinase (TK)</u>

The TK of HSV-1 has been studied extensively due to its function as a gene transfer marker and its role in antiviral therapy (see Section 1.4.5). HSV-1 TK, encoded on gene UL23, phosphorylates deoxycytidine and thymidylate in addition to thymidine (Jamieson & Subak-Sharpe, 1974). This enzyme is not essential for virus growth in tissue culture but virus mutants in the enzyme show significantly reduced pathogenicity (Field & Wildy, 1978). VZV gene 36 encodes an active TK with 28% sequence homology to the HSV-1 enzyme (Cheng *et al.*, 1980). There appears to be no sequence homologue for the TK gene in HCMV although it may encode another type of nucleoside kinase and also appears to induce the cellular TK (Chee *et al.*, 1989). It has been found that HCMV UL97 encodes a protein capable of phosphorylating ganciclovir, a function performed by TK in HSV-1 (Littler *et al.*, 1992). In the γ-herpesvirus, EBV-encoded TK activity has been demonstrated in transformed mammalian cell lines (Littler *et al.*, 1986).

<u>Ribonucleotide reductase (RR)</u>

RR catalyses the reduction of all four ribonucleoside diphosphates to their corresponding deoxyribonucleoside diphosphates. The HSV-1 enzyme is composed of two pairs of subunits (large and small) which are encoded by the neighbouring genes UL39 and UL40 (Cohen, 1972; Frame *et al.*, 1985; Ingemarsson & Lankinen, 1987; McLauchlan & Clements, 1983). Inhibition of the enzyme by peptide disruption of the subunit interaction has focused attention on the possibility of antiviral therapy (Dutia *et al.*, 1986; McClements *et al.*, 1988). The synthetic nonapeptide which was found to cause inhibition of HSV-1 RR (equivalent to the nine C-terminal residues of the small subunit) was also found to cause similar inhibition of EHV-1 RR (Conner *et al.*, 1993; Telford *et al.*, 1990). The two VZV genes, 18 and 19, show homology to the HSV-1 genes UL40 and UL39 respectively. HCMV gene 45 has homology to the large subunit gene UL39 but no convincing homologue of UL40. It has been suggested that the cellular version of the small subunit may be utilised by HCMV to produce an active RR (Chee *et al.*, 1989). Genes BaRF1 and BORF2 show homology to HSV-1 UL40 and UL39 respectively.

<u>Deoxyuridine triphosphatase (dUTPase)</u>

HSV-1 specific dUTPase activity was discovered by Wohlrab and Francke (1980) and later mapped to gene UL50 by Preston and Fisher (1984). The enzyme

catalyses the hydrolysis of dUTP to dUMP and inorganic phosphate. The herpesvirus dUTPases are discussed separately in Section 1.4.8. It is noted here that the β-herpesvirus plus HHV6 and HHV7 do not contain a convincing homologue of the dUTPase enzyme (see Results Section 4.7).

Thymidylate synthase (TS)

In the generation of thymine, dUMP is methylated to dTMP by the enzyme TS (see Figure 1.4). The enzyme TS was discovered in HVS based on 70% homology with the human gene (Honess *et al.*, 1986). HSV-1, HCMV and EBV do not possess a TS however a copy was found during sequencing of the VZV genome (gene 13) and this was later shown to be active (Davison & Scott, 1986; Thompson *et al.*, 1987). It has been suggested that TS is only required by viruses with high A+T genomes (Honess *et al.*, 1986). This is backed up by the relatively high A+T genomes of HVS and VZV and the discovery of a TS in herpesvirus ateles (Richter *et al.*, 1988).

Dihydrofolate reductase (DHFR)

Previous to 1988, possession of a DHFR had only been described in the T-even and T5 bacteriophages and not for any mammalian virus. Subsequently a virally encoded DHFR was described for HVS and herpesvirus ateles and more recently for HHV8 (Trimble *et al.*, 1988; Nicholas *et al.*, 1998). The coding of a DHFR in HVS may be again related to the possession of an A+T rich genome although the reason for coding DHFR in HHV8 is still unclear.

1.3.2 Nucleotide metabolism

There are two pathways utilised by the mammalian cell for the production of RNA and DNA precursors (NTPs and dNTPs), the *de novo* and the salvage pathways. The *de novo* pathway is the main route where nucleotides are built up stepwise from small metabolic precursors such as ribose and amino acids. In the salvage pathway, free bases and nucleosides released from nucleic acid breakdown are recycled. Nucleosides can also be taken up from extracellular fluids by active transport across the cell membrane (Reichard, 1988). Specific kinases phosphorylate these nucleosides to yield nucleotides.

We are principally concerned with pyrimidine metabolism and the role of dUTPase in this pathway. DNA differs from RNA in both the sugar moiety and the set of bases utilised. In DNA, uracil is replaced with thymine. This has direct consequences for nucleotide provision since dTTP has a more complex synthesis pathway compared to dATP, dGTP and dCTP (Figure 1.4). The *de novo* synthesis of NTP's and dNTP's is linked by the enzyme ribonucleotide reductase which acts on

ribonucleotides at the diphosphate level to produce deoxyribonucleotides (Reichard, 1993).

ADP $\xrightarrow{\quad 1 \quad}$ dADP $\xleftarrow{\quad 6 \quad}$ dATP

GDP $\xrightarrow{\quad 1 \quad}$ dGDP $\xleftarrow{\quad 6 \quad}$ dGTP

CDP $\xrightarrow{\quad 1 \quad}$ dCDP $\xleftarrow{\quad 6 \quad}$ dCTP

dCMP

2

dUMP $\xrightarrow{\quad 3 \quad}$ dTMP $\xleftarrow{\quad 5 \quad}$ dTDP $\xleftarrow{\quad 6 \quad}$ dTTP

7

UDP $\xrightarrow{\quad 1 \quad}$ dUDP $\xrightarrow{\quad 6 \quad}$ dUTP
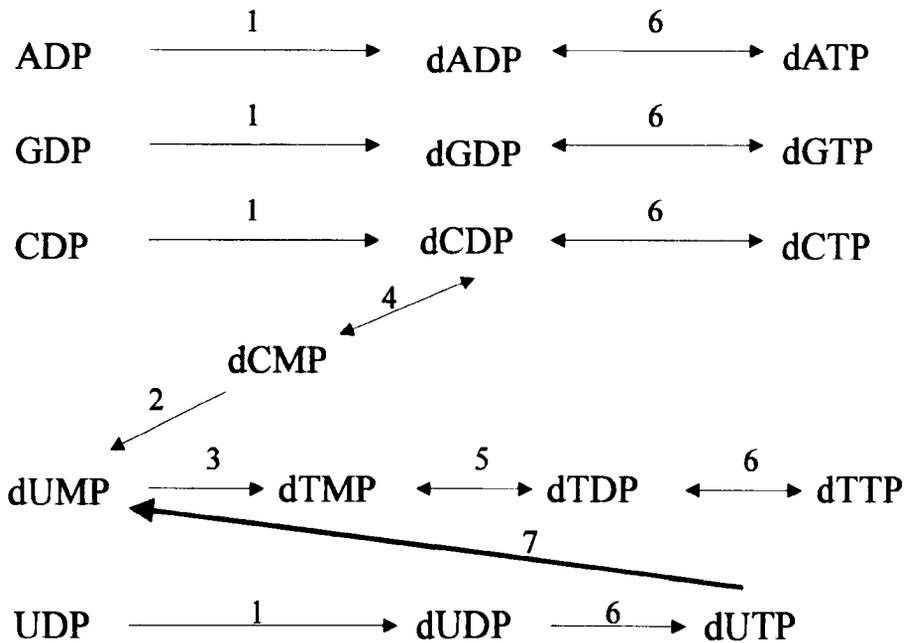
Figure 1.4 The later stages of the *de novo* pathway of pyrimidine deoxyribonucleotide synthesis in mammals.
The enzymes are labelled as follows: (1) ribonucleotide reductase, (2) dCMP deaminase, (3) thymidylate (dTMP) synthase, (4) (d)CMP kinase, (5) thymidylate kinase, (6) nucleoside diphosphate kinase, (7) dUTPase. This diagram was adapted from (Reichard, 1988).

The synthesis of dTTP occurs at great cost to the cell with enzymes encoded exclusively to deal with the replacement of uracil with thymine in DNA. A supply of dUMP is maintained by the action of two enzymes, dCMP deaminase and dUTPase. In mammals, the deamination of dCMP provides over 80% of the dUMP substrate for thymidylate synthase and the remaining amount arises from the activity of dUTPase (Nicander & Reichard, 1985; Reichard & Nicander, 1985). In contrast, *E.coli* produces about 75% of dUMP for dTMP synthesis by the action of dUTPase. Unlike mammalian cells, *E.coli* possesses a deaminase which generates dUTP by the deamination of dCTP (Danielsen *et al.*, 1992).

Thymidylate synthase (TS) links dNTP synthesis to folate metabolism and is dependent on the activity of dihydrofolate reductase (Reichard & Nicander, 1985). TS catalyses the methylation of dUMP to dTMP. The methyl donor in this reaction is N^5, N^{10}-methylenetetrahydrofolate which is oxidised to dihydrofolate. This molecule is recycled back to tetrahydrofolate by the enzyme dihydrofolate reductase using NADPH as the reductant. Phosphorylation of dTMP to dTDP is performed by thymidylate kinase which also phosphorylates dTDP, dUMP and dUDP (Pearl & Savva, 1996). The

final reaction in this pathway is carried out by the non-specific nucleoside diphosphate kinase which phosphorylates the dNDP's to their corresponding dNTP's.

The maintenance of dNTP levels by these enzymes is crucial to the genetic stability of the cell. The results of an imbalance can have widespread effects including mutation, recombination, enhanced sensitivity to mutagens and carcinogens, chromosome breakage, exchange or loss, and oncogenic transformation (reviewed by Kunz *et al.*, 1994).

1.4   The enzyme deoxyuridine triphosphatase (dUTPase)

1.4.1   The function of dUTPase

Deoxyuridine triphosphatase (EC 3.6.1.23) is also referred to as deoxyuridine triphosphate nucleotidohydrolase and dUTP pyrophosphatase in the literature. The position of dUTPase in the *de novo* pathway of nucleotide synthesis is shown in Figure 1.4. The enzyme plays a crucial role in the regulation of dNTP synthesis by catalysing the hydrolysis of dUTP to form dUMP and pyrophosphate (Figure 1.5) (Bertani *et al.*, 1961).
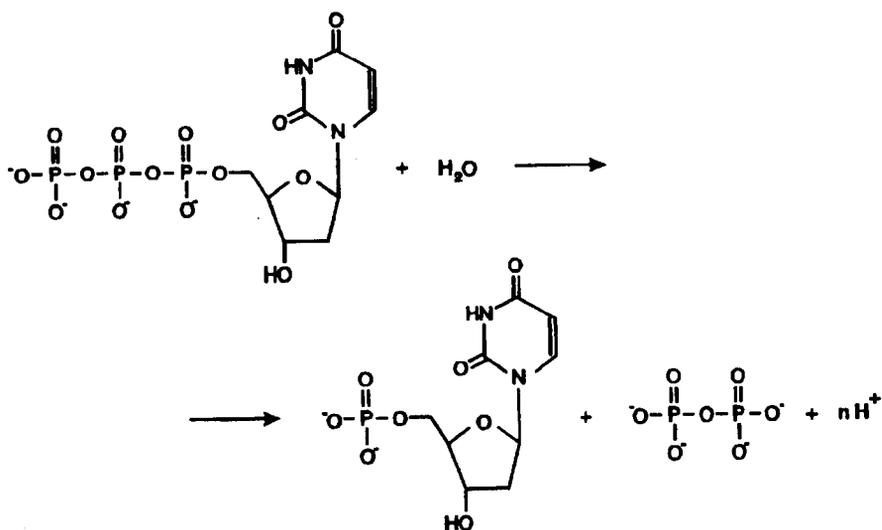
Figure 1.5 The dUTPase reaction.
This diagram was adapted from (Bergman, 1997). The number of protons released in the reaction 'n' is dependent on pH and Mg^{2+} concentration (Rogers *et al.*, 1997).

The reaction mechanism of the *E.coli* dUTPase has been investigated and the substrate for the reaction appears to be a complex of dUTP with a divalent metal ion, preferably Mg^{2+} (Hoffmann *et al.*, 1987). The number of protons released in the

reaction has been shown to be determined by the concentration of metal ions and the pH of the reaction buffer (Rogers *et al.*, 1997). The enzyme has a high specificity for its substrates and discriminates efficiently between the base, sugar and phosphate moieties. This is a necessary capability for the enzyme since the hydrolysis of other nucleotides required for DNA and RNA synthesis would be highly detrimental. This is exemplified in the *E.coli* enzyme by discrimination against other highly structurally related molecules such as dUDP, UTP, dCTP and dTTP where phosphate, sugar and base are discriminated against with remarkable efficiency (Björnberg & Nyman, 1996; Larsson *et al.*, 1996a). All dUTPases characterised to date show a high degree of specificity for dUTP although there are subtle differences between organisms (Björnberg & Nyman, 1996). This may be a reflection of the evolutionary adaptation of the enzyme in different environments driven by a trade off between specificity and catalytic efficiency.

The reactions involved in the provision of nucleotides must be controlled to ensure a balance is maintained. Disruption in this balance has direct consequences for DNA replication and the continued growth of the cell (Reichard, 1988). An increase in the ratio of dUTP compared to dUMP leads to large quantities of uracil being incorporated into the newly synthesised DNA chain. DNA polymerase, in contrast to dUTPase, does not distinguish between dUTP and dTTP (Shlomai & Kornberg, 1978; Focher *et al.*, 1990). Incorporation of uracil into DNA in place of thymine is not in itself mutagenic since both base pair with adenine. However, uracil also arises in DNA by the spontaneous deamination of cytosine which, if not corrected, will result in a point mutation in newly replicating DNA. Replacement of uracil for thymine in DNA can also disrupt sequence-specific DNA recognition by gene-regulatory proteins (Verri *et al.* 1990).
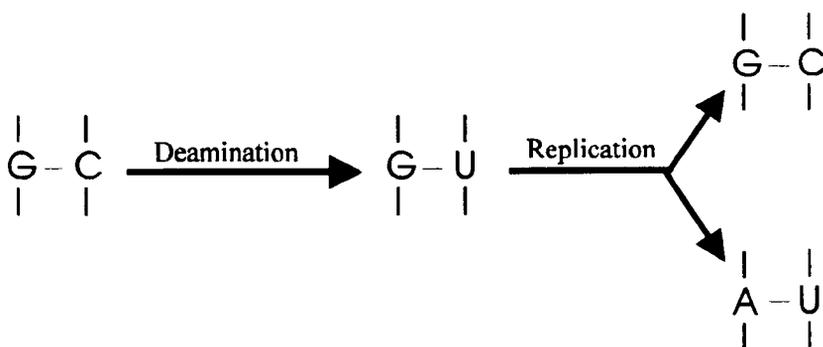
Figure 1.6 Effect of unrepaired deamination of cytosine in replicating DNA
The deamination of the cytosine residue in the G-C base pair results in an A-U base pair after replication. Further rounds of replication will result in the transition to an A-T base pair.

Cytosine is the most sensitive of the bases to nucleophilic attack even in physiological conditions and deamination is estimated to create about 100 uracil residues per day in an average sized mammalian genome (Lindahl, 1993). To cope with this potentially mutagenic situation an excision repair mechanism has evolved, utilising the enzyme uracil DNA glycosylase (UDGase) (Lindahl, 1982). The enzyme hydrolyses the glycosidic bond between the uracil and deoxyribose moiety in DNA creating an abasic site. An AP endonuclease (where AP represents an *ap*urinic or *ap*yrimidinic site) recognises the abasic site and nicks the phosphodiester backbone at an adjacent position. The residual deoxyribose phosphate is removed by an exonuclease and DNA polymerase I inserts the complementary base on the undamaged strand. DNA ligase completes the repair by rejoining the nicked backbone (reviewed by Friedberg *et al.*, 1995).

The instability of cytosine may help to explain why thymine, and not uracil, is the standard constituent of DNA. UDGase specifically removes uracil and not thymine from DNA by discriminating against a single methyl group. The methyl group on thymine may therefore act as a tag to distinguish it from deaminated cytosine. In this context, it is likely that thymine is utilised in DNA to increase fidelity. RNA in comparison is a short lived genetic molecule where uracil, the less energetically expensive building block, is sufficient for information transfer.

Excessive uracil incorporation as a consequence of an elevated dUTP/dTTP ratio results in excessive levels of UDGase mediated repair. Since this process involves a transient strand break in the DNA backbone, multiple repairs in close proximity can result in DNA fragmentation (Tye *et al.*, 1977; Ingraham *et al.*, 1986). It has been suggested that the misincorporation of uracil into DNA has direct consequences for neuronal ageing (Mazzarello *et al.*, 1990). Inhibitors of TS, such as fluorouracil, create a dramatic reduction in the dTTP pool and the resulting cytotoxic effect has been called 'thymineless death'. It has been proposed that the cytotoxic effect of these inhibitors is due to extensive uracil incorporation and excision repair (Ingraham *et al.*, 1986).

In summary dUTPase has two functions;
1. As evident from the nucleotide pathway in Figure 1.2, dUTPase is a necessary component in the *de novo* synthesis of dTTP by supplying the substrate, dUMP, for dTMP synthase.
2. The misincorporation of dUTP results in over stimulation of a natural repair mechanism which is highly detrimental to the cell. dUTPase must act to limit the amount of dUTP available to DNA polymerase and maintain the critical balance between dUTP and dTTP.

A further consideration concerns the physical organisation of dUTPase relative to other nucleotide metabolism enzymes (Wheeler *et al.*, 1992). In the bacteriophage T4 a non-covalently linked dNTP synthesising complex with a mass of 1300kDa was isolated from infected *E.coli* cells. This complex contained at least ten enzymes, two of which were encoded by the host cell and the rest by the viral genome, and exhibited dUTPase activity (Mathews, 1993). It has been further suggested that this complex is in turn linked to the replication machinery, allowing dNTPs to be delivered directly to the replication fork (Mathews, 1993). The complex may be based on a common affinity of all proteins in the complex for the bacteriophage encoded gene 32 protein (Wheeler *et al.*, 1996). The perceived potential of such a complex is that physical connection of enzymes carrying out sequential steps in a metabolic sequence could allow significant kinetic advantage. Similar complexes have not been reported in other systems although the weak interactions involved would make this a difficult area of study. It would be interesting to discover if any eukaryotic viruses had evolved such a mechanism for the formation of complexes with both host and virally encoded enzymes. Such organisation would help to account for the variety of nucleotide metabolism enzymes which are specifically virally encoded since physical interaction with a host enzyme may overcome the requirement to encode a viral copy.

1.4.2   Distribution and control of cellular dUTPases

The enzyme dUTPase is ubiquitous in nature and many representatives have been substantially characterised. In prokaryotes examples include *E.coli* (Larsson *et al.*, 1996a) and *Bacillus subtilis* (Dunham & Price, 1974). In eukaryotes examples include *Saccharomyces cerevisiae* (Gadsden *et al.*, 1993), *Drosophila melanogaster* (Nation *et al.*, 1989), rat (Hokari & Sakagishi, 1987), human (Williams & Cheng, 1979; Ladner *et al.*, 1996b), *Allium cepa* (onion) (Pardo & Gutierrez, 1990) and *Lycopersicon esculentum* (tomato) (Pri-Hadash *et al.*, 1992). Many viral dUTPases have also been characterised and are discussed separately in Section 1.4.3. With such a large range of organisms encoding an active dUTPase it has been suggested that this enzyme is indispensable to all cells (McIntosh *et al.*, 1992). The structural and functional relationships between dUTPases from diverse organisms provide an insight into the evolution of a cellular nucleotide metabolism enzyme. Representatives from distantly related organisms, such as those from *E.coli* and human cells, show remarkable structural similarity in terms of their trimeric arrangement and active site cavities (Mol *et al.*, 1996).

The herpesvirus dUTPase proteins represent a distinct subset of the dUTPase enzymes as a whole. In order to easily distinguish between these two groups the

standard trimeric dUTPases (representing all those mentioned above) are referred to as class I dUTPases while the herpesvirus monomeric enzymes are referred to as class II dUTPases. This distinction addresses the structural differences between these two groups. It should be noted however that the dUTPase of the herpesvirus, CCV (and probably other fish herpesviruses) are clearly members of the class I group. The class II dUTPases are approximately double the chain length of their class I counterparts and are active as monomers as opposed to trimers (Caradonna & Adamkiewicz, 1984). The class II dUTPases are discussed separately in Section 1.4.8 and are the main subject of Chapter 4. Active dUTPases have also been discovered in Leishmania and Trypanosoma species which appear to show no sequence similarity to either the class I or class II enzymes. It appears likely that these examples represent an additional distinct subset of dUTPases (Camacho *et al.*, 1997). Different organisms not only present different structural versions of the same enzyme but they also utilise the enzyme in subtly different ways.

dUTPases have been found to be essential in *E.coli* (El-Hajj *et al.*, 1988) and yeast (Gadsden *et al.*, 1993) although a viable double mutant of *E.coli* has been constructed which lacks both dUTPase and UDGase activities (Warner *et al.*, 1981). This mutant accumulates large amounts of uracil in its DNA but is presumed to survive due to the removal of the uracil extracting repair mechanism (Hochhauser & Weiss, 1978).

A conditional temperature sensitive dUTPase mutant (*dut* Ts) has been constructed in *E.coli* (El-Hajj *et al.*, 1992). It was found that phenotypic revertants of *dut* Ts restored viability without restoring the enzymatic activity of dUTPase. This secondary mutation was designated *dus* for *dut* suppressor. Further studies mapped, cloned, and identified the *dus* locus which was found to encode a dCTP deaminase (Wang & Weiss, 1992). As mentioned in Section 1.3.2., in *E.coli*, the major source of dUMP for dTTP synthesis is from the action of dUTPase (as opposed to the deamination of dCMP). The pathway utilised by *E.coli* involves the following intermediates: dCTP→dUTP→dUMP→dTMP. Mutation of the dCTP deaminase gene is likely to suppress the lethality of the dUTPase mutation by reducing the formation of dUTP.

A multiple *E.coli* mutant with additional mutations to *dut* and *dus* including *ung* (UDGase), *deoA* (thymidine/deoxyuridine phosphorylase) and *thyA* (TS) has been created (El-Hajj *et al.*, 1992). It has been proposed that mutation in *ung* would prevent uracil excision from DNA, mutation in *deoA* would allow more efficient use of exogenous deoxyuridine and mutation in *thyA* would prevent synthesis of dTMP from any remaining dUMP. This multiple mutant displayed up to 93 to 96% substitution of uracil for thymine in new DNA. Growth of this mutant ceased after cellular DNA had

increased 1.6 to 1.9 fold and the cell mass had increased 1.7-2.7 fold suggesting an overall failure of macromolecular synthesis.

From this evidence it may be suggested that dUTPase is a necessary enzyme which must be actively reducing the dUTP/dTTP ratio and supplying dUMP for dTTP synthesis at all times. However, it has been shown that the cell may regulate dUTPase activity to a high degree. In the root meristem cells of the plant *Allium cepa*, dUTPase activity has been shown to correlate closely with cellular proliferation (Pardo & Gutierrez, 1990). Activity is higher at optimal growth conditions and decreases as cells begin to differentiate. There is strong evidence for cell-cycle regulation in *Allium cepa* cells with a large increase in dUTPase activity at the G_1/S boundary and continuing throughout the S phase. No activity could be detected at any other point in the cell cycle indicating a strong cellular control mechanism. A similar cell cycle dependent activity has been shown in Chinese hamster cell temperature sensitive mutants (Duker & Grant, 1980).

In *Drosophila melanogaster*, cellular control of dUTPase activity has been attributed to a developmentally expressed protein inhibitor (Nation *et al.*, 1989). dUTPase has been purified from *Drosophila* embryos and has been shown to be active only at early times in first-instar larvae (Giroir & Deutsch, 1987). A heat stable protein with a subunit molecular mass of 61kDa has been partially purified from the embryo and has been shown to be an active inhibitor of the embryonic dUTPase (Nation *et al.*, 1989). This direct method of cellular control may be indicative of another utilisation of dUTPase. Controlled reduction in cellular dUTPase activity may result in uracil incorporation into DNA and UDGase mediated strand breakage. In the developing embryo this strategy could be used to increase DNA degradation during the histolysation process in the pupae where internal organs are dissolved before new growth of imaginal tissue.

The role of dUTPase in the proliferation and maturation of human T cells has been investigated and it appears that the human enzyme is also dependent on the cell cycle (Strahler *et al.*, 1993). There appears to be at least two forms of the human dUTPase, nuclear and mitochondrial. The mitochondrial form possesses an extended region at the amino terminus compared to the nuclear form (Ladner *et al.*, 1996b). The nuclear form can be phosphorylated and a cyclin-dependent kinase phosphorylation site has been identified (Ladner *et al.*, 1996a). It has been suggested that the phosphorylation of the human dUTPase results in regulation of enzyme activity (Lirette & Caradonna, 1990). However, disruption of the phosphorylation site prevents phosphorylation but has no significant effect on enzyme activity *in vivo* (Ladner *et al.*, 1996a). Other possible roles for the phosphorylation of the human enzyme are multimerisation and cellular localisation (Ladner *et al.*, 1996a). Whether the control

mechanism is translational, post-translational or mediated by a *Drosophila*-like inhibitor protein, it endows the human cell with regulation of dUTPase activity.

1.4.3 The distribution of viral dUTPases

Many, but not all viruses, encode a dUTPase. The discovery that a wide variety of viruses genomes encode a dUTPase was originally based on a number of sequence analysis studies. McClure *et al.* (1987) identified a supposedly protease-like gene segment specified in two distantly related groups, the lentiviruses and oncoviruses, although it was not present in all retroviruses. The gene segment was originally identified in the polymerase region of the lentiviruses, visna virus and equine infectious anaemia virus (EIAV). Homologous sequences were identified in a simian retrovirus type 1 (SRV-1) and its close relative, hamster intracisternal A particle (IAP-H18) although these gene segments were located adjacent to the protease gene. On the basis of position and low level sequence homology, it was proposed that these gene segments originated by tandem duplication and subsequent divergence of the retroviral protease coding sequence (McClure *et al.*, 1987; McClure *et al.*, 1988). No function was ascribed to these polypeptides which were termed 'protease-like' domains. Subsequent studies found related genes in the poxviruses, vaccinia virus and orf virus (Slabaugh *et al.*, 1989; Mercer *et al.*, 1989). These related genes, termed 'pseudoproteases' were found to comprise an independent ORF and possess transcriptional control signals, rather than as a subunit of a larger polypeptide as in retroviruses.

McGeoch (1990b) following on from these studies, correctly identified these so called 'protease-like' or 'pseudoprotease' genes as dUTPases. The gene segments identified from retroviruses and poxviruses were compared to the dUTPase sequences from *E.coli* and the herpesviruses, HSV-1, VZV and EBV. All sequences in the alignment were found to contain five highly conserved, short amino acid motifs. This work is discussed in detail in Section 1.4.8. This finding has allowed the subsequent identification of dUTPases from other viruses based on the conservation of these five motif regions. The viral dUTPases characterised to date include members of the Herpesviridae, Retroviridae, and Poxviridae (Table 1.4).

Family	Virus	Reference
Herpesvirus	Herpes simplex virus type 1 (HSV-1)	(Wohlrab & Francke, 1980)
	Herpes simplex virus type 2 (HSV-2)	(Wohlrab et al., 1982)
	Epstein-Barr virus (EBV)	(Williams et al., 1985)
	Bovine herpesvirus 1 (BHV-)	(Liang et al., 1993)
	Pseudorabies virus (PRV)	(Jons & Mettenleiter, 1996)
	Varicella-zoster virus (VZV)	(Ross et al, 1997)
Poxvirus	Vaccinia virus	(Broyles, 1993)
Retrovirus	Mason-Pfizer monkey virus (MPMV)	(Elder et al., 1992)
	Simian type D retrovirus (SRV-1)	(Elder et al., 1992)
	Feline immunodeficiency virus(FIV)	(Wagaman et al., 1993)
	Equine infectious anaemia virus (EIAV)	(Threadgill et al., 1993)
	Mouse mammary tumour virus (MMTV)	(Koppe et al., 1994)
	Caprine arthritis-encephalitis virus (CAEV)	(Turelli et al., 1996)
	Visna virus	(Turelli et al., 1996)

Table 1.4 Viral dUTPases which have been confirmed to be functional

From Table 1.4 is evident that most of the current research in viral dUTPases is focused on members of the Herpesviridae and Retroviridae which are both discussed in later sections. Sequence homologues of dUTPase have also been found in the Adenoviridae, and the unclassified African swine fever virus (previously in the Iridoviridae family) but these have yet to be characterised. Many viruses which possess the dUTPase enzyme are sure to be identified in the near future from these families and others.

1.4.4 The role of dUTPase in viral life cycles

In this section a variety of viral life cycles are reviewed which exhibit different nucleotide metabolism pathways. These life cycles range from bacteriophages which utilise non-standard bases to those which encode a multiple array of enzymes to allow reversal of normal cellular functions. The retroviruses are also described here in relation to the possession of a dUTPase whereas the herpesviruses are described separately in Section 1.4.8.

The replication mechanisms of many organisms differ with regard to nucleotide metabolism but none more so than bacteriophages of E.coli and B.subtilis. The φe bacteriophages of B.subtilis contain 5-hydroxymethyluracil (HMU) in place of thymine and possess dTTPase and dUTPase activities. Partial purification suggests that both activities are attributable to one, dual function enzyme. This dTTPase-dUTPase is thought to be responsible for excluding both uracil and thymine from phage DNA and

providing dUMP, the substrate for the φe-induced dUMP hydroxymethylase (Price & Warner, 1969).

The T-even phages of *E.coli* replace cytosine with 5-hydroxymethylcytosine (HMC) in their DNA and encode a hydroxymethylase (Adams *et al.*, 1981). They encode a dCTPase-dUTPase whose dCTPase activity prevents the incorporation of cytosine and produces dCMP, the substrate for dCMP hydroxymethylase. A kinase is encoded to phosphorylate the HMC to the diphosphate level and a host enzyme converts the diphosphate to the triphosphate. The return for encoding these extra enzymes is the advantage of promoting degradation of the host DNA by deoxyribonuclease. Specificity for the host DNA is attributed to the lack of cytosine clusters in the phage DNA due to replacement with HMC (Stryer, 1988). Another phage encoded enzyme catalyses the glycosylation of HMC molecules which appears to further facilitate survival of the viral genome by reducing degradation by cellular enzymes (Cohen, 1968).

The two examples above reveal how subtle changes to the chemical constituents of the viral DNA can be advantageous by allowing the replication and survival of viral DNA in preference to host DNA. Another replication mechanism variation is that of the PBS-2 phage of *B.subtilis* which has DNA containing uracil instead of thymine. To enable this phage to replicate in the host cell it must encode a number of proteins including a dTTPase (Price & Fogt, 1973). To allow accumulation of dUTP for incorporation into newly replicating DNA it encodes a dUTPase inhibitor protein (Price & Frato, 1975). PBS-2 must also encode a UDGase inhibitor to prevent purposely incorporated uracil molecules being excised (Wang & Mosbaugh, 1989). The UDGase inhibitor has been crystallised and characterised kinetically revealing that its mechanism of action is based on nucleotide mimicry (Bennet *et al.*, 1993; Savva & Pearl, 1995). The purification and crystallisation of the dUTPase inhibitor protein to reveal its mechanism of action would be extremely interesting. The question as to what advantage can be gained from such a complex replication cycle remains unclear. Since the uracil rich DNA will be resistant to many restriction endonucleases this may have incurred an evolutionary advantage for this genome.

Given the importance of dUTPase in general cellular metabolism it is not surprising that many viruses encode their own copy of the enzyme. It is interesting however that possession of a virally encoded dUTPase varies between viruses, even those closely related. One of the most striking examples of this variation is found in the Retroviridae where the lentivirus FIV encodes a dUTPase whereas its human counterpart HIV does not. Existence of this gene in such a small, 10kb genome is indicative of its functional importance. It has been found that only distinct subsets of retroviruses encode a dUTPase gene (Elder *et al.*, 1992) probably attributable to their

distinct life cycles. Retroviruses possess a single-stranded, RNA genome in the region of 8-10kb in length. The viral RNA is converted to DNA by virus encoded reverse transcriptase. Integrase, another virally encoded enzyme, performs integration of the viral DNA into the host cell chromosome. The retroviral genome is made up of three major genes, *gag, pol* and *env*. The *gag* region encodes structural proteins, *pol* encodes reverse transcriptase and integrase and *env* encodes a transmembrane glycoprotein which allows attachment and entry into the host cell. In the lentiviruses, the dUTPase gene is located in the *pol* gene between the reverse transcriptase and the integrase (McClure *et al.*, 1987). In the type D retroviruses, MMTV, the dUTPase is expressed as an uncleaved transframe protein with the nucleocapsid protein (Bergman *et al.*, 1994; Koppe *et al.*, 1994).

Many of the retroviral dUTPase genes have been analysed by mutation studies and it is apparent that the replication of these dUTPase-deficient (dUTPase⁻) viruses is highly dependent on the state of the host cell. In many cases replication was close to wild type levels in certain dividing host cells. Studies with FIV dUTPase⁻ mutants *in vitro* have shown that replication occurred with close to wild type levels in actively dividing host cells but was greatly reduced in non-dividing cells such as primary macrophages (Wagaman *et al.*, 1993). A similar situation is seen with *in vitro* studies on EIAV (Threadgill *et al.*, 1993), CAEV and visna virus (Turelli *et al.*, 1996) where again there was reduced replication in primary macrophages but not in actively dividing cells. Studies with these mutants *in vivo* show that dUTPase does have an essential role in virus replication. Using FIV dUTPase⁻ mutants *in vivo*, several key discoveries were made: 1) mutants infect their host with similar kinetics to wt, 2) mutants elicit a similar humoral antibody response to wt and 3) the virus burden is reduced in the mutants. Furthermore, it was found that the mutation rate of mutant FIV's integrated in the DNA of primary macrophages after 9 months was five fold greater than the mutation rate of mutant FIV's integrated in the DNA of T-lymphocytes. Comparative mutation rates of the wt virus between cell types were the same (Lerner *et al.*, 1995). This is a clear indication that a virally encoded dUTPase becomes advantageous in non-dividing cells. This is backed up with similar *in vivo* data from studies with CAEV showing an increased mutation rate in dUTPase⁻ mutants (Turelli *et al.*, 1997). dUTPase⁻ mutants of EIAV also shows increased incorporation of uracil into viral DNA (Steagall *et al.*, 1995) and replication deficiencies in macrophages (Lichtenstein *et al.*, 1995).

From these studies it is clear that dUTPase has a role *in vivo* in non-dividing cells where levels of host dUTPase activity may be low. The fact that HIV does not carry a dUTPase while other lentiviruses do may simply be because HIV does not encounter significantly high levels of dUTP during infection of the natural host.

However, since the FIV dUTPase is actually packaged into the virion (Elder *et al.,* 1992) it has been suggested that HIV may have developed a mechanism for the incorporation of the host dUTPase within its own mature virion (McIntosh *et al.,* 1992). Based on the discovery that human endogenous retrovirus-like elements (HERVs) also carry a gene with similarity to dUTPase (Cordonnier *et al.,* 1995) another possibility exists.   If these viruses encode a functional copy of the dUTPase gene it may be possible for HIV to utilise this host-based source of the enzyme during replication (McIntosh *et al.,* 1992).

1.4.5   dUTPase as an antiviral target

There are relatively few successful antiviral therapies available at present. The main reason for this is that because viruses replicate in cells and employ much of the host's biosynthetic machinery, highly specific viral targets are hard to find. Many of the current antivirals suffer from low therapeutic indexes due to non-specific disruption of host cell components. The main areas of research into potential antiviral targets are historically nucleotide metabolism and DNA replication.   As discussed previously, many viruses encode their own enzymes to allow efficient replication of viral DNA in the host cell. Identification and characterisation of these enzymes has allowed some success in the development of viral specific inhibitors. There are very few antiviral drugs licensed for use and most of them are nucleoside analogues (Table 1.5). The other antiviral drugs work in a variety of ways.

Amantadine targets the process of endosome acidification and thus prevents release of the viral genome into the cytoplasm. Rimantadine has a similar mechanism and is less toxic than amantadine.   Phosphonoformic acid, an analogue of pyrophosphate, inhibits DNA polymerase and viral reverse transcriptase (Blackburn & Gait, 1995).

Drug	Virus
Nucleoside analogues	
5-Iodo-2'-deoxyuridine	HSV
5-Trifluoromethyl-2'-deoxyuridine	HSV
Adenine arabinoside	HSV
5-Ethyl-2'deoxyuridine	HSV (in Germany)
5-Iodo-2'-deoxycytidine	HSV (in France)
Acyclovir (ACV, guanosine analogue)	HSV, VZV
(E)-5-(2-Bromovinyl)-2'-deoxyuridine	HSV, VZV (in Germany)
3'-Azido-2',3'-dideoxythymidine (AZT)	HIV
2',3'-Dideoxyinosine	HIV
2',3'-Dideoxycytidine	HIV
Ganciclovir (guanosine analogue)	HCMV
Ribavirin (purine nucleoside analogue)	Respiratory syncytial virus
Other	
Amantadine	Influenza A
Phosphonoformic acid	HCMV
Rimantadine	Influenza A (in Russia)

Table 1.5 Drugs approved for use against viral infection in humans in 1993 (Blackburn & Gait, 1995)

The most successful antiviral drugs in recent years include acyclovir (ACV) for herpesviruses and 3'-azido-2',3'-dideoxythymidine (AZT) for HIV-1. Referring to Table 1.5 it can been seen that many of the licensed drugs are active against herpesvirus. The reason for this is twofold: Firstly, many of the herpesviruses enzymes have been substantially characterised allowing greater scope for directed drug development and secondly, many of the active drugs against herpesviruses are nucleoside analogues which are a by-product of extensive anti-cancer drug programs.

The variation between the viral and host enzyme provides a basis for designing specific, non-toxic drugs. The majority of the herpesvirus antivirals rely on the virally encoded enzyme, thymidine kinase (TK). TK selectively phosphorylates nucleoside analogues, such as ACV, allowing conversion to the triphosphate state which is the active form. Since this reaction is not performed by the cellular host enzymes it provides a basis for specificity. Direct inhibition of the viral DNA polymerase has also been possible although the differences from the cellular version are limited (reviewed by Cohen, 1992).
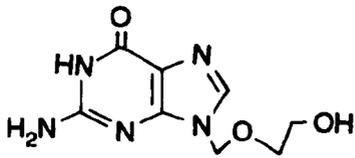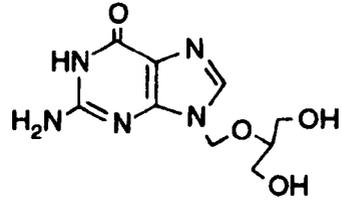
Figure 1.7 Structure of Acyclovir (ACV) and Ganciclovir (GCV)

ACV has the highest therapeutic index of any antiviral drug. Unlike many of the antivirals, the mechanism of action of acyclovir is well characterised. The structure of ACV is based on a deoxyguanosine which lacks part of the sugar ring (Figure 1.7). ACV is activated by the virally encoded TK which can phosphorylate a variety of nucleoside analogues. The mechanism of action is shown in Figure 1.8.
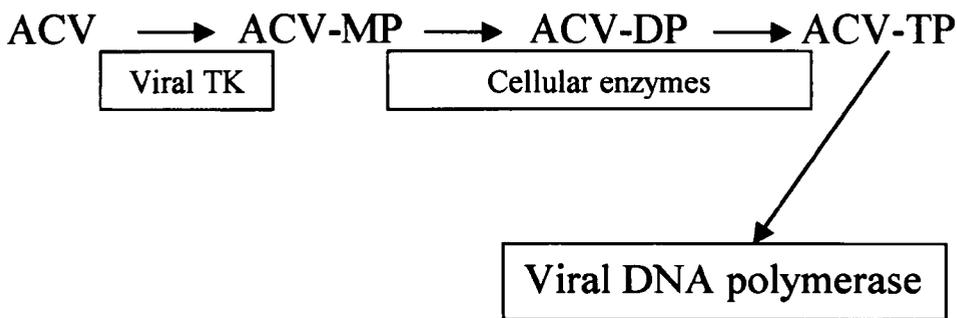
Figure 1.8 The mechanism of Acyclovir action (Cohen, 1992)

ACV is phosphorylated by the virally encoded TK converting it to ACV-monophosphate. Subsequent conversion to ACV-diphosphate and -triphosphate is performed by cellular enzymes. The triphosphate form of ACV acts as a competitive inhibitor with dGTP and also as a substrate for the viral DNA polymerase which incorporates the analogue into the growing DNA chain. Because there is no 3'-hydroxyl on the acyclo-GTP, the viral polymerase cannot add another dNTP. Furthermore, for reasons that are not yet clear, the polymerase becomes inactivated and viral DNA replication ceases. Non-infected mammalian cells do not phosphorylate ACV efficiently, imparting a high degree of specificity for HSV infected cells (Cohen, 1992). Ganciclovir has a similar structure to ACV although by possessing the C-3' moiety on the sugar ring it is closer to the natural substrate. This may account for the

fact this is more toxic than ACV. Ganciclovir has the same basic mechanism of action as ACV and is commonly used to treat HCMV infection.

The first report of ACV resistance in an immunocompromised HSV-infected patient was in 1982 (Crumpacker *et al.*, 1982). Cases of resistance have been particularly common in patients suffering from AIDS or undergoing transplants. It is thought that the increased viral burden, characteristic of severely immunocompromised patients, increases the likelihood of resistance (Chatis & Crumpacker, 1992). There are three mechanisms of resistance including the selection of TK deficient mutants which lack the enzyme (Schmipper *et al.*, 1980) and selection of mutants which produce an TK with a substrate specificity excluding ACV (Darby *et al.*, 1981). Alternatively a mutated polymerase is selected for which is capable of elongating DNA in the presence of high concentrations of ACV (Knopf *et al.*, 1981). Ganciclovir suffers from a similar set of resistance mechanisms (Chatis & Crumpacker, 1992).

It is clear that there is a very limited number of selective antivirals available to treat herpesvirus infections and there are resistance problems with those in common use. There are, however, ongoing studies with other approaches such as disruption of viral or viral-cellular protein interactions and inhibition of proteases to block viral assembly. Antisense oligonucleotides are being developed to inhibit viral gene expression and work is continuing in the production of soluble receptors which block herpesvirus attachment and entry.

It is possible to disrupt specific protein subunit interactions by peptide inhibition (reviewed by Marsden, 1992). The peptide is engineered to correspond to one of the two subunits, whether it is of viral or cellular origin. The best studied peptide inhibition of a viral protein subunit interaction is that of HSV ribonucleotide reductase (RR) (Dutia *et al.*, 1986). RR is a functionally important enzyme in viral replication and is discussed in Section 1.3.1. The HSV RR is a heterodimer composed of two molecules of a large subunit and two molecules of a small subunit (Ingemarsson & Lankinen, 1987). A nonapeptide was synthesised, corresponding to the carboxy-terminal region of the large subunit. This nonapeptide was able to specifically inhibit the viral RR by competing with the binding site on the large subunit to which the small subunit normally associates (Dutia *et al.*, 1986). The specificity for the viral enzyme appears to be from the difference in amino acid sequence at this binding region between the viral and mammalian RRs. The approach has also had some success in other herpesvirus protein interaction such as that between the polymerase (UL30) and the dsDNA binding protein (UL42) (Marsden *et al.*, 1994).

Although these studies have proved to be useful *in vitro*, there are obstacles to overcome before they become clinically useful, such as delivery of the peptide into the target cell and degradation by cellular enzymes. Antisense oligonucleotides have

similar problems to peptides in terms of clinical use. The strategy involves the design of oligonucleotides that are complementary to a unique viral sequence. These molecules then hybridise to the viral DNA or RNA and interfere with the target sequence (reviewed by Bischofberger & Wagner, 1992).

A relatively new target for a herpesvirus drug is the virally encoded dUTPase. Evidence suggesting an important role for the enzyme plus structural studies on related dUTPases has generated further interest. The herpesvirus encoded dUTPases are substantially different from the mammalian versions yielding the potential of selective inhibition of the viral enzyme. Elucidation of the active site region of the herpesvirus dUTPases may allow analogues to be developed that act specifically against the viral enzyme. HSV and EBV dUTPases have now been accepted as serious contenders for novel drug design (Williams, 1988; Sommer *et al.*, 1996). Sequence analysis studies have identified a conserved motif present in the herpesvirus dUTPases but not in other dUTPases (McGeoch, unpublished work). This is discussed fully in the Results (Section 4.7). Clearly, if a function can be attributed to this conserved region of the viral enzyme, it may be possible to design drugs targeted at this area. To adequately design selective inhibitors for the herpesvirus dUTPase it is first necessary to characterise the enzyme thoroughly both structurally and functionally. This thesis deals with the preliminary results of this characterisation.

A substrate analogue has been developed which is a potent inhibitor of both the *E.coli* and the HSV-1 dUTPases (Bergman *et al.*, 1997; Persson *et al.*, 1996). The dUDP analogue, 2'-deoxyuridine 5'-(α,β-imido)diphosphate (dUPNPP) was chemically synthesised and the active triphosphate was prepared enzymatically using the enzyme pyruvate kinase and phosphoenolpyruvate as a phosphate donor. This method was also used to phosphorylate the imidodiphosphate analogue of 2'-deoxythymidine to 2'-deoxythymidine 5'-(α,β-imido)triphosphate (dTPNPP). Replacement of the α,β-bridging oxygen in dUTP with an imido group results in a non-hydrolysable substrate analogue (Figure 1.9).
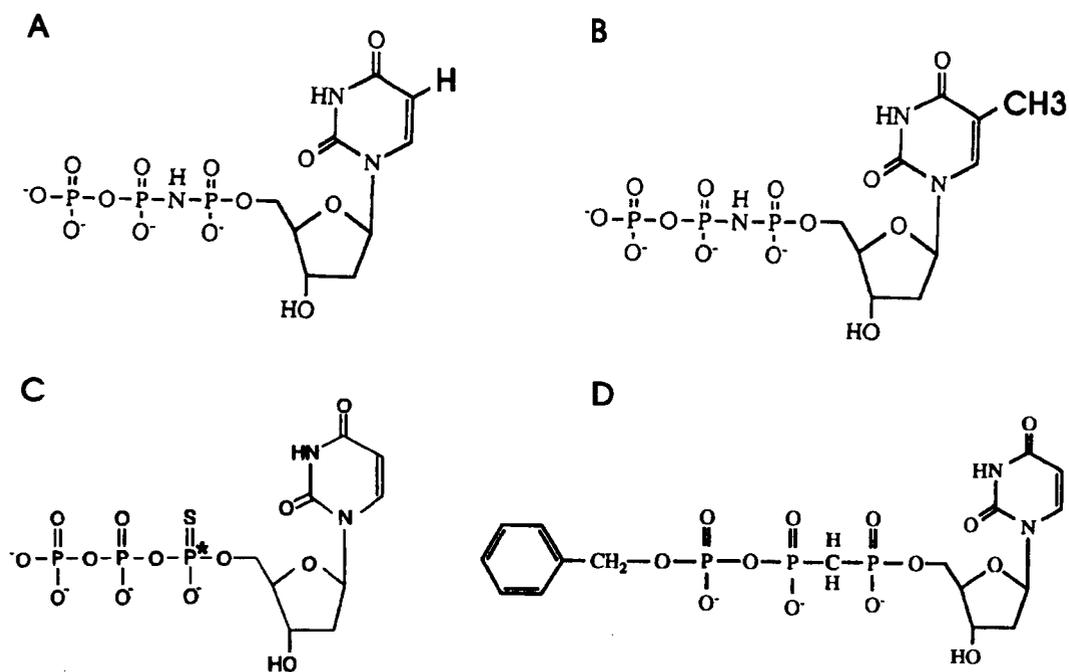
Figure 1.9 Structure of pyrimidine analogues
The chemical structures of the following compunds are shown: **(A)** dUPNPP, **(B)** dTPNPP, **(C)** dUTPαS and **(D)** BM-dUTP.

Both dUPNPP and dTPNPP compounds were developed for use in crystallisation experiments but also demonstrate that inhibitors can be readily produced. The structure of analogue benzyl-(2'-deoxy-5'-uridinyl-hydroxyphosphinyl-methylenyl-hydroxyphosphinyl-oxo)-phosphonate (BM-dUTP) is also shown in Figure 1.9. This substrate analogue has been synthesised and found to be a useful inhibitor of dUTPase in cancer cell lines (see Section 1.4.6). The molecule dUTPαS has two enantiomers of which only one is hydrolysed by dUTPase in the presence of Mg^{2+}. The non-hydrolysable enantiomer can be hydrolysed upon the addition of Co^{2+} thus providing evidence for the participation of the α-phosphate with a metal ion in the reaction mechanism (Bergman, 1997).

1.4.6 Human dUTPase and chemotherapy

It is clear that dUTPase plays a key role in many viral replication cycles. It is also clear that dUTPase is an essential enzyme to mammalian cells specifically those cells which are actively growing or dividing. The variation in requirement for cellular dUTPase appears to be dependent on the cell state, with actively dividing cells producing much larger quantities of the enzyme than resting cells (Nation *et al.*, 1989).

This makes dUTPase a useful target for cancer therapy since its inactivation may selectively inhibit rapidly dividing cells (McIntosh *et al.*, 1992).

It was widely considered that cell death due to inhibition of thymidylate synthase (TS) was a direct result of low levels of intracellular dTTP. It was not, however, understood why such a transient nucleotide imbalance could result in irreversible cell damage and death. It has been found that induction of cell death by inhibition of TS involves an increase of intracellular dUTP resulting in misincorporation into DNA and fragmentation due to excision repair (Curtin *et al.*, 1991). This enzyme is a key target for cancer chemotherapy and several inhibitory drugs have been developed including methotrexate and fluorodeoxyuridine (FdUrd). FdUrd gives a significant benefit to cancer patients but a large percentage of tumours exhibit intrinsic or acquired resistance. Many of the mechanisms of resistance of the group of fluoropyrimidine drugs have been identified as mutations of the cellular TS resulting in reduced inhibition. Further studies with FdUrd have implicated dUTPase in the resistance mechanism of the drug (Canman *et al.*, 1993). It appears that in response to high dUTP levels induced by FdUrd the cell can increase dUTPase activity to compensate (Lirette & Caradonna, 1990). This increase in dUTPase activity allows the intracellular dUTP levels to be reduced thereby lowering the potentially cytotoxic misincorporation of uracil into DNA (Beck *et al.*, 1984). This is backed up by older studies which show that inhibition of dUTPase with nucleoside analogues increases the cytotoxicity of methotrexate in cell culture (Beck *et al.*, 1985; 1986).

The role of dUTPase in resistance to FdUrd induced cytotoxicity has been confirmed by expressing *E.coli* dUTPase in a mammalian cell line prior to drug treatment (Canman *et al.*, 1994). Cells expressing the recombinant *E.coli* enzyme showed a 4-5 fold increase in overall dUTPase activity. These cells were found to be protected against the action of FdUrd compared to cells expressing standard levels of dUTPase activity (Canman *et al.*, 1994). Although the *in vivo* situation has not been investigated, these results support the theory that the action of anti-tumour drugs such as FdUrd may be enhanced by combining them with dUTPase inhibitors thus maximising uracil incorporation into DNA and the consequent strand breakage. In terms of resistance to drugs such as FdUrd, dUTPase inhibitors may become extremely useful. Studies on the human enzyme indicate that there are at least two distinct versions of the enzyme of which one may be activated by phosphorylation (Section 1.4.2). Further characterisation of the human dUTPase will allow greater potential to regulate the enzyme in tumour cells and may provide a degree of specificity.

1.4.7   The dUTPase of *Escherichia coli*

The *E.coli* enzyme is the most highly characterised dUTPase to date in terms of both structure and function.   After large scale purification was successful, *E.coli* dUTPase was the first to be crystallised (Hoffmann *et al.*, 1987; Cedergren-Zeppezauer *et al.*, 1992).   The *E.coli* enzyme is used as a working model for a typical class I dUTPase and as a basis for the molecular modelling of the class II HSV-1 dUTPase in this thesis.   There are now four class I enzymes for which a crystal structure has been solved: *E.coli* (Cedergren-Zeppezauer *et al.*, 1992), FIV (Prasad *et al.*, 1996), EIAV (Persson *et al.*, 1997) and a human (Larsson *et al.*, 1996b; Mol *et al.*, 1996) dUTPase. The crystal structure of the *E.coli* enzyme with bound dUDP has also been solved allowing identification of functionally important residues within the structure (Larsson *et al.*, 1996c).

The *E.coli* dUTPase was originally described as a tetramer (Shlomai & Kornberg, 1978) but in the course of crystallographic analysis was found to be a homo-trimer (Cedergren-Zeppezauer *et al.*, 1992).   The structure of the *E.coli* dUTPase is presented in Figure 1.10(a) with each of the three subunits coloured individually.   Each subunit comprises a globular region with an extended tail which is formed by the C-terminal region of the protein.   The extended arm of each subunit crosses the adjacent subunit.   In the *E.coli* crystal structure the last 16 residues of this arm are not visible in the electron density map even when dUDP is bound and it has been proposed that the arm is flexible (Larsson *et al.*, 1996c).   Subsequent studies using non-hydrolyzable substrate analogues have demonstrated that the *E.coli* arm closes over the active site following the binding of 2'-deoxyuridine 5'-(α,β-imido)triphosphate (diphosphate shown in Figure 1.9A) in complex with Mg^{2+} and then opens after catalytic cleavage (Vertessy *et al.*, 1998).

Primary sequence comparisons of the class I and class II dUTPases revealed five distinct regions of local sequence similarity which have been named motifs 1-5 (McGeoch, 1990b).   These conserved regions are dealt with in detail in Chapter 4. Analysis of the *E.coli* dUTPase-dUDP co-crystal demonstrated that at least four of these motifs regions condense to form an active site pocket (Figure 1.10 b and c).   The trimer has three such active sites, with each one positioned at the interface region between adjacent subunits (Larsson *et al.*, 1996c).   Motif 5 is positioned at the end of the C-terminal tail and is not visible even in the dUDP co-crystal structure.
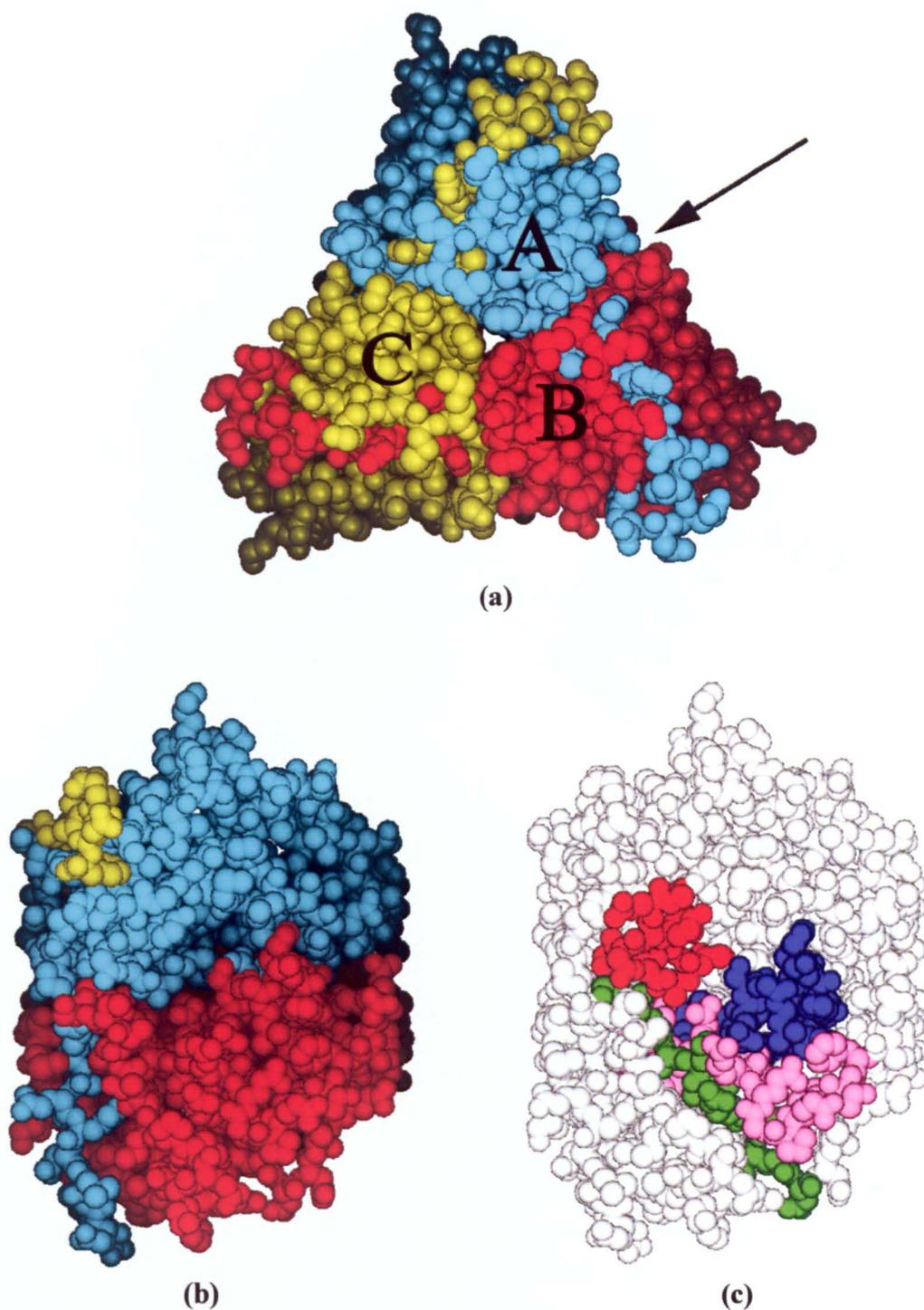
Figure 1.10 Structure of the *E.coli* trimer (top) and active site region (bottom).
(a) The *E.coli* trimer. Each of the three identical subunits is coloured: 'A', cyan, 'B', red and 'C', yellow. There are three active sites in the trimer positioned at the interface between adjacent subunits. An arrow indicates the position of one of the three active sites. **(b)** Side view of the trimer looking in the direction of the arrow in (a). **(c)** Active site region. View as in (b) with all three subunits coloured white and the motif regions highlighted. Subunit 'A' contributes motif 3 (red). Subunit 'B' contributes motifs 1 (green), 2 (blue) and 4 (pink). Motif 5 is not visible in this structure but is proposed to be contributed by subunit 'C' at the end of the C-terminal arm. The last visible residues, 134-136, of the 'C' subunit (yellow) can be seen approaching the active site area in figure (b).
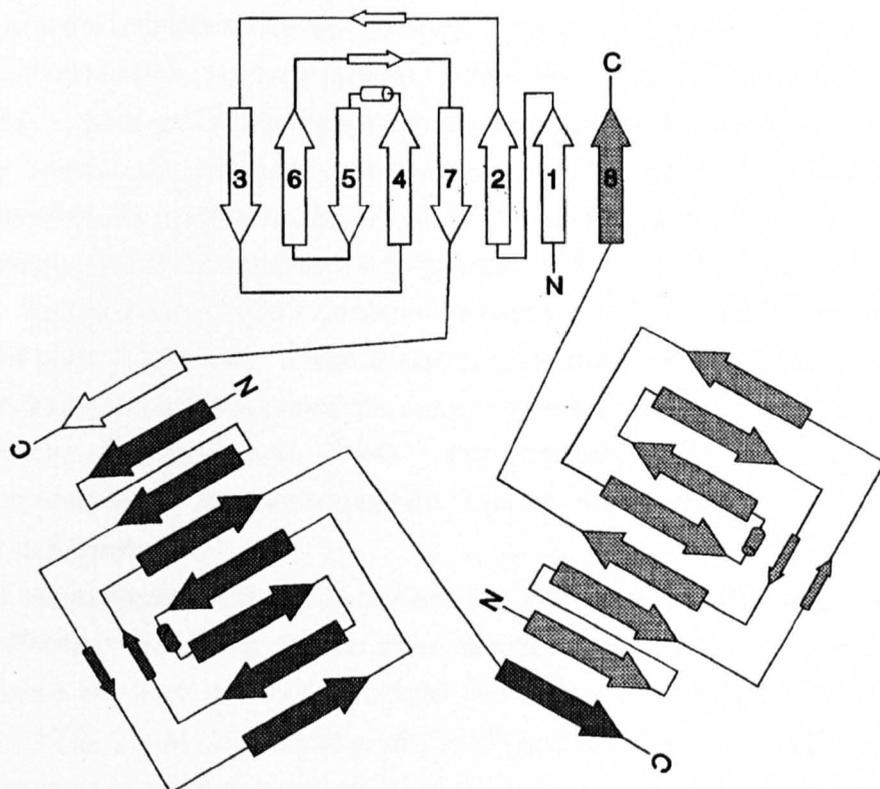
Figure 1.11 The secondary structure of the *E.coli* trimer (Bergman, 1997).

Each subunit is composed of 10 β-strands and one short α-helix (see Section 4.4.2 for a detailed description). The 7 main β-strands from one subunit plus an eighth extended β-strand from a neighbouring subunit form a secondary structure resembling the 'jelly-roll' often found in viral coat proteins. Three such folds condense to form the trimer depicted in Figure 1.11. The last visible residue of the enzyme, Phe-136, lies at the end of the arm and points in the direction of the active site. The structure of the human dUTPase trimer reveals that motif 5 also participates in the active site area and it is likely that a similar situation exists in the *E.coli* enzyme (Mol *et al.*, 1996). It has been suggested that this region becomes ordered only upon binding a nucleoside triphosphate in complex with a metal ion (Larsson *et al.*, 1996c). Motif 5 is rich in glycine residues and has similarity to phosphate binding sequences in other nucleotide binding proteins (Moller & Amons, 1985; Bossemeyer, 1994).

One of the most interesting discoveries is that with the participation of motif 5, each active site is made up from motifs from all three subunits. This can be visualised in Figure 1.10 where subunit 'A' contributes motif 3, subunit 'B' contributes motifs 1, 2 and 4, and subunit 'C' is likely to contribute motif 5 at the end of the arm structure. This has important implications in the modelling of the HSV-1 enzyme and will be discussed in some detail in the Results Chapter 4.

The most important residues involved in the specificity for dUTP are contained in motif 3. This region may be a general uridine binding motif since homologues have been found in both dCTP deaminases and pseudouridine synthases (Koonin, 1996). There are, however, no structural data to determine if these homologous sequences are in fact involved in uridine binding. Motif 3 has a β-hairpin loop structure with discrimination against ribose achieved by a tyrosine (Tyr-93) which lies at the corner of this loop. This residue is highly conserved between dUTPases and is also found in the majority of class II enzymes. There is chemical evidence that acetylation or nitration of this residue in *E.coli* inactivates the enzyme (Vertessy *et al.*, 1996; Vertessy *et al.*, 1994; Vertessy & Zeppezauer, 1994). The structure of the active site and the mechanism employed for discrimination against sugar, base and phosphate are discussed in Section 4.2.

Kinetic studies with the *E.coli* enzyme and other dUTPases have resulted in widely different values even for the most standard measurements. The K_M for the *E.coli* enzyme has been reported as 1.5μM (Bertani *et al.*, 1961), 12μM (Shlomai & Kornberg, 1978), 22μM (Hoffmann *et al.*, 1987) and recently as 0.2μM (Rogers *et al.*, 1997). The most recent measurement, carried out by stopped-flow analysis, has shown that by excluding Mg^{2+} from the reaction, an almost 100-fold higher K_M was obtained and metal-free dUTP was shown to be an inhibitor of the reaction (Larsson *et al.*, 1996a). It is possible that Mg^{2+} concentration was not taken into account in earlier experiments. It is also possible that the discontinuous measurement of dUTPase activity (as measured by reaction product quantification using TLC separation) is not accurate enough to measure K_M in the submicromolar range.

The stopped-flow technique measures the release of protons and subsequent pH change in a weakly buffered solution containing a pH indicator dye. This system allows real-time measurement of the hydrolysis reaction and was utilised for a detailed kinetic analysis of the *E.coli* enzyme (Larsson *et al.*, 1996a). It was found that the enzyme was highly specific for its substrate dUTP, with the next best substrate, dCTP, hydrolysed 10^5 times less efficiently (by comparison of specificity constants, k_{cat}/K_M). This vast difference is mainly attributable to a higher K_M for dCTP. A catalytic mechanism has been interpreted from these data and involves the magnesium binding to the α-phosphate, rate-limiting hydrolysis by an activated water molecule and fast ordered desorption of the products. The turnover (k_{cat}) for the *E.coli* dUTPase, in the range of 6 to 9 s^{-1}, is considered to be slow and it has been suggested that this is a consequence of the high specificity the enzyme has for dUTP.

1.4.8 The dUTPases of herpesviruses

The dUTPase of HSV-1 is the focal enzyme in this thesis. It is encoded by gene UL50, a rightward oriented ORF of 371 codons. The position of the gene in the HSV-1 genome was correctly identified by (Preston & Fisher, 1984) although there were earlier reports describing a different location (Williams & Parris, 1987; Wohlrab *et al.*, 1982). It lies head to head with UL49A and tail to tail with UL51. The close proximity of UL50 and UL49A means that they overlap their respective promoter regions.

Fisher & Preston (1986) generated the HSV-1 mutant dut⁻1218 which contains an insertion of 12 base pairs at the KpnI site within the UL50 ORF (see Section 5.3.2 & 5.3.3). Inactivation of the HSV-1 dUTPase gene had no effect on growth of the virus in tissue culture and demonstrated that the virally encoded dUTPase was non-essential *in vitro* (Fisher & Preston, 1986; Barker & Roizman, 1990). This is not surprising since the natural cellular hosts of HSV-1 include nondividing cells such as neurons. It has been reported that infection with HSV-1 down regulates the host cell dUTPase whereas infection with a dUTPase⁻ mutant does not (Lirette & Caradonna, 1990). In the latter case it is likely that the virus can rely on the host dUTPase in cell culture. Again, this study was carried out with actively dividing HeLa cells and therefore does not represent the true picture *in vivo*.

Later studies using the mouse model indicated that infection with HSV-1 dUTPase⁻ mutants results in a marked reduction in neurovirulence, neuroinvasiveness and reactivation from latency (Pyles *et al.*, 1992). This gives a better perspective of the role of the enzyme in neural tissue. The HSV-1 insertion mutant, dut⁻1218 (Fisher & Preston, 1986) and a deletion mutant, 17B1, were tested *in vivo* (Pyles *et al.*, 1992). Neuroinvasiveness, as measured by footpad inoculation, was over 1,000 fold less than wt HSV-1. A complication results from the deletion of the UL49A promoter in mutant 17B1. This mutant was constructed with a deletion between base pairs 107028 and 107956 (UL50 represents bp 107010 to 108123) which includes the promoter region for UL49A. Analysis of this mutant demonstrated that UL49A was not transcribed. This may have had an effect on invasion and replication in the CNS since UL49A has been reported to be a virion membrane protein (Barnett *et al.*, 1992; Jons *et al.*, 1996). Furthermore, the restoration of the dUTPase gene in mutant 17B1 did not restore neuroinvasiveness in some isolates suggesting that a second mutation may have been inadvertently selected. The insertion mutant dut⁻1218 was also shown to be over 1,000 fold less neuroinvasive following footpad inoculation as compared to the wt 17 *syn*⁺ virus. Restoration of the dUTPase gene to this mutant resulted in fully wt viruses. Concern that the small, four residue, in-frame insertion may be leaky is unlikely given that it disrupts the highly conserved motif 3 which is crucial to substrate binding (Section 4.2.5). Further confidence is given in that several groups have not found any

HSV-1 specific dUTPase activity in mutant dut⁻1218 (Fisher & Preston, 1986; Williams, 1988).

Even with the reservations in the *in vivo* study by Pyles *et al.* (1992) it is likely that the virally encoded dUTPase has a role in allowing reactivation from latency and viral replication in neuronal cells. HSV-1 mutants in thymidine kinase and ribonucleotide reductase show reduced replication in neuronal tissue as well as a reduction in the capacity to establish latency (Katz *et al.*, 1990; Kosz-Vnenchaak *et al.*, 1990). It is possible that in adult neurons HSV-1 requires dUTPase to redress the nucleotide balance before initiating its own DNA replication.

One of the distinct characteristics of herpesviruses is their ability to establish latent infection. EBV has been shown to encode an active dUTPase during infection and this enzyme may be required for reactivation and replication in resting epithelial tissue or resting B cells (Sommer *et al.*, 1996). Just as mammalian cells may switch off dUTPase to promote nucleotide imbalance and apoptosis (Section 1.4.2), viruses may switch on dUTPase to push dTTP synthesis from dUMP thus allowing viral replication in otherwise resting cells (Pyles *et al.*, 1992). Studies with PRV and BHV-1 indicate the presence of active dUTPases and mutants have been constructed and tested in cell culture (Jons & Mettenleiter, 1996; Liang *et al.*, 1993). As expected, in both cases dUTPase is dispensable in cell culture. *In vivo* studies utilising these mutant viruses is likely to result in reduced replication and/or reactivation from latency.

The herpesvirus dUTPases vary in length particularly between the three subfamilies of herpesviruses. The longest polypeptide chains are found in the α-herpesvirus with VZV the largest at 396 residues. The γ-herpesviruses include shorter chains with EBV at 278 residues and HVS at 287 residues. The β-herpesvirus, HCMV, has been suggested to carry a dUTPase (UL72) on the basis of its location in the genome compared to the other herpesviruses (Chee *et al.*, 1990). There are no convincing homologues of the five motifs found in other dUTPases although it possesses a herpesvirus exclusive motif (McGeoch, unpublished). This protein is the subject of further discussion in the Results section.

All dUTPases which have been characterised to date (excluding the dUTPase of *Leishmania major*), share five short regions of amino acid similarity termed motifs 1-5 (Figure 1.12) (McGeoch, 1990b). Excluding the herpesvirus enzymes, all dUTPases contain these five motifs in the same order over a roughly similar length protein chain around 150 residues. The only major variation from this primary structure in the class I enzymes is found in the retrovirus MMTV where the nucleocapsid domain is fused to the N-terminus of the dUTPase (Bergman *et al.*, 1994). The herpesvirus dUTPases, or class II dUTPases, share the five common motifs found in class I dUTPases but in a different linear order on the polypeptide chain. Motif 3 is displaced to the N-terminal

half of the class II protein with regard to the other four motifs. The class II dUTPases are approximately double the protein chain length of the class I enzymes and exhibit greater length variation. The distance been motifs 4 and 5 is also enlarged compared to the class I enzymes (Figure 1.12).
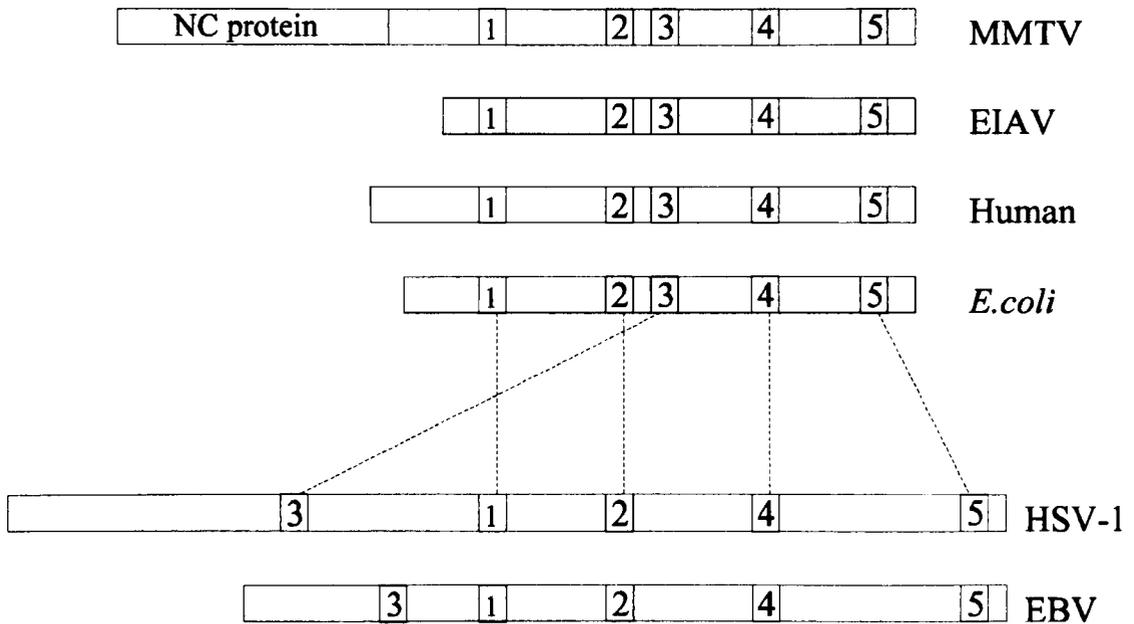
Figure 1.12 Comparison of the order of the five motifs in a subset of class I (MMTV, EIAV, Human and *E.coli*) and class II (HSV-1 and EBV) dUTPases.

HSV-1 dUTPase is the best studied of the class II enzymes and unlike the trimeric class I enzymes has been shown to be active as a monomer (Caradonna & Adamkiewicz, 1984; Persson, 1999, personal communication). EBV dUTPase has also been shown to be a monomer and it is likely that the other class II enzymes which share this primary structure arrangement are also monomers (Persson *et al.*, 1999). It has been suggested that the class II enzymes have arisen by an intragenic duplication of a class I enzyme (McGeoch, 1990b). The proposed evolutionary process linking the two classes of dUTPase is depicted in Figure 1.13. The top figure represents a typical class I dUTPase, such as that of *E.coli*, with five conserved motifs in the class I linear order. Intragenic duplication of this gene creates a double length protein chain with two copies of the conserved motif regions 1-5. Subsequent loss of motifs 1, 2, 4 and 5 from the N-terminal half plus motif 3 from the C-terminal half results in the motif arrangement found in the class II dUTPases.
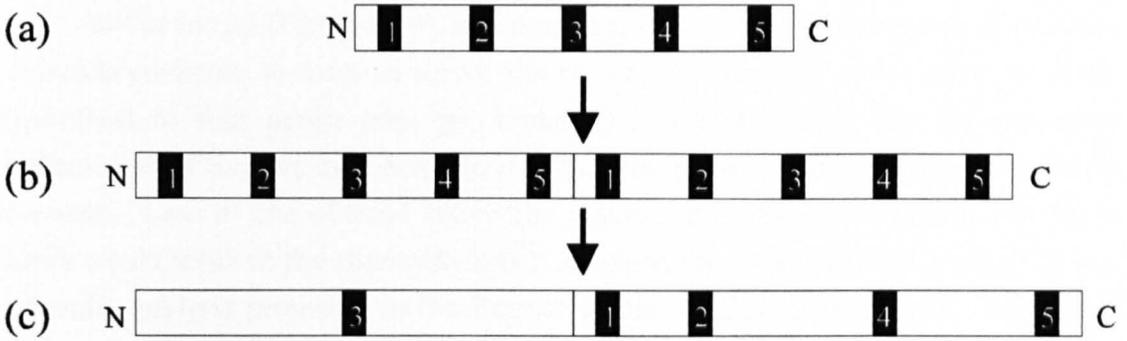
Figure 1.13 Schematic representation of a class I intragenic duplication.
(a) represents the standard class I dUTPase with its conserved motifs 1-5. (b) represents a duplication of the class I gene resulting in a double length protein, possessing two copies of motifs 1-5. (c) represents the genuine arrangement of the class II dUTPases. The transition from (b) to (c) requires the loss of motifs 1,2,4 and 5 in the N-terminal half of the protein and motif 3 in the C-terminal half. There must also be an insertion of novel sequence between motifs 4 and 5 to account for the extension.

A model accounting for the observed motif rearrangement in the class II enzymes was proposed by (McGeoch, 1990b). This model was based on the assumption that the *E.coli* dUTPase class I enzyme was a tetramer (Shlomai & Kornberg, 1978). As described in Section 1.4.7. it is now clear that the *E.coli* dUTPase is, in fact, a trimer. This does not, however, affect the basic idea of the original model since the principles are easily transposed onto a trimeric class I molecule.
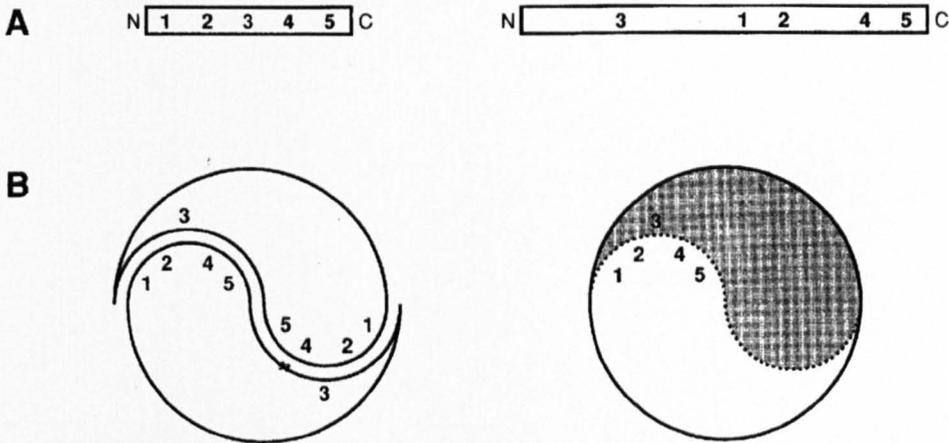
Figure 1.14 Arrangement of motifs and model for dUTPase quaternary structure.
A. Linear arrangement of the motifs of class I (left) and class II dUTPases (right).
B. The left cartoon represents the model of McGeoch (1990b), assuming a tetrameric structure for the class I dUTPase. The proposed motif positions are shown on two of the four subunits. Each of the two active site regions depicted is composed of motifs 1, 2, 4, and 5 from one subunit and motif 3 from the other. The right cartoon represents a class II dUTPase monomer, with the C-terminal region in white and the N-terminal region shaded. Diagram from McGeoch (1990b).

In this model (Figure 1.14), it is proposed that the motifs from each of the class I subunits condense to form an active site region, creating two active sites per dimer (equivalent to four active sites per tetramer). It is proposed that an equivalent condensation of the five motifs would result in the formation of two active sites in the monomer. Loss of one of these active site regions, and subsequent loss of one set of motifs would result in the observed class II arrangement. The structural modelling and molecular analysis presented in the Results section of this thesis provide insight into how the transition from a class I to a class II enzyme could be achieved. This is based on the current knowledge that the *E.coli* and other class I enzymes are trimeric and each of their three active sites is composed of motifs from all three subunits.

Chapter 2 - Materials

2.1 Chemicals and reagents

All chemicals and reagents were purchased from BDH Chemicals UK or Sigma with the exception of those listed below and referred to specifically in the Methods section.

Amersham Life Science	Rainbow coloured protein MW markers
Beecham Research	Ampicillin sodium B.P. (Penbritin)
Bio-Rad Laboratories	Ammonium persulphate, Coomassie brilliant blue R250, TEMED
Boehringer Mannheim	leupeptin, pepstatin, PMSF, agarose MP, Tris (2-amino-2(hydroxymethyl)-1,3-propanediol)
Fluka	Formamide
FMC Bioproducts	NewSieve GTG agarose
Gibco BRL Life Technologies	IPTG
Melford Laboratories	CsCl
National Diagnostics	Ecoscint A scintillation fluid
Pharmacia Biotech	7-deaza-dGTP, dNTPs, ddNTPs
Prolabo	Boric acid, butanol, chloroform, ethanol, glacial acetic acid, glycerol, hydrochloric acid, isopropanol, methanol
United States Biochemical	Glycerol tolerant gel buffer
Whatman International	P11 cellulose phosphate

2.2 Solutions

2.2.1 Standard solutions

Coomassie stain	0.2% (w/v) Coomassie Brilliant Blue R250 in methanol:water:acetic acid in a 50:50:7 ratio
Destain	5% (v/v) methanol, 7% (v/v) acetic acid in water
DF dyes	50% (v/v) TBE, 50% (v/v) glycerol, 0.1% (w/v) bromophenol blue
dNTPs	1.25mM stock (312.5 μM of each of dATP, dCTP, dGTP, dTTP)
Formamide dyes	10mM EDTA, 1mg/ml xylene cyanol FF, 1mg/ml bromophenol blue, 80% (v/v) formamide, pH 8.0
Kinase buffer (5X)	350mM Tris-HCl (pH 7.5), 50mM MgCl$_2$, 25mM DTT
Ligation buffer (5X)	250mM Tris-HCl (pH 7.6), 50mM MgCl$_2$, 5mM DTT, 5mM ATP, 25% (w/v) PEG 8000

PCR buffer (10X)	100mM Tris-HCl, 15mM MgCl$_2$, 500mM KCl, pH 8.3
PBS(A)	170mM NaCl, 3.4mM KCl, 10mM Na$_2$HPO$_4$,
	1.8mM KH$_2$PO$_4$, pH 7.2
PBS-complete	PBS(A) plus CaCl$_2$.2H$_2$O and MgCl$_2$.6H$_2$O at 1g/l
SEQ buffer (5X)	200mM Tris-HCl, 80mM MgCl$_2$, pH 7.5
Sequencing buffer	200mM Tris-HCl, 100mM MgCl$_2$, pH 7.5, 250mM NaCl
(for Sequenase)	
TAE (50X)	0.2M Tris, 0.05 EDTA (pH8.0), pH to 8.0 with glacial
	acetic acid
TBE (10X)	0.8M Tris, 0.3M boric acid, 0.2mM, EDTA
TE	10mM Tris-HCl, 1mM EDTA, pH 8.0
TFB	10mM MES, 100mM RbCl, 45mM MnCl$_2$.4H$_2$O,
	10mM CaCl$_2$.2H$_2$O, 3mM hexammine cobaltic chloride,
	pH6.3

2.2.2 Glycine SDS PAGE buffers:

Loading buffer (3X)	29% (v/v) SGB, 6% (w/v) SDS, 2M β-mecaptoethanol,
	29% (v/v) glycerol, 1mg/ml bromophenol blue
Resolving gel buffer (RGB)	1.5M Tris-HCl, 0.4% (w/v) SDS, pH 8.9
Stacking gel buffer (SGB)	0.5M Tris-HCl, 0.4% (w/v) SDS, pH 6.7
Tank buffer	0.05M Tris, 0.05M glycine, 0.1% (w/v) SDS

2.2.3 Lysis Buffers for plasmid sequencing

Lysozyme solution	0.4% (w/v) Lysozyme, 50mM sucrose, 25mM Tris,
	pH 8.0
NaOH/SDS	0.2M NaOH, 1% (w/v) SDS
Potassium acetate solution	3M KAc, 2M acetic acid
STET	8% sucrose, 0.5% triton X-100, 50mM EDTA (pH 8.0),
	50mM Tris-HCl, pH 8.0

2.2.4 Extraction / purification buffers

IBI extraction buffer A	50mM Tris-HCl (pH 8.0), 5mM EDTA, 0.25mg/ml
	lysozyme, 50μg/ml sodium azide
IBI extraction buffer B	1.5M NaCl, 0.1M CaCl$_2$, 0.1M MgCl$_2$,
	0.02 mg/ml DNAse I, 50μg/ml sodium azide
IBI periplasmic EB	0.5M sucrose, 0.03M Tris, 1mM EDTA at a final
	pH of 8.0 after resuspension

pET extraction buffer	20mM Hepes(pH 7.3), 50mM NaCl, 10%(v/v) glycerol, 1mM EDTA, 1mM EGTA, 1mM DTT, 1mM PMSF, 2µg/ml leupeptin, 2µg/ml pepstatin A, 0.1% polyoxyethylene-10-tridecyl ether
FPLC buffer I (Phosphocellulose)	20mM Hepes(pH 7.3), 50mM NaCl, 10%(v/v) glycerol, 1mM EDTA, 1mM EGTA, 1mM DTT, 1mM PMSF, 2µg/ml leupeptin, 2µg/ml pepstatin A , 0.01% polyoxyethylene-10-tridecyl ether
FPLC buffer II (Mono S)	50mM MES(pH 6.5), 100mM NaCl, 10%(v/v) glycerol, 1mM EDTA, 1mM EGTA, 1mM DTT, 1mM PMSF, 2µg/ml leupeptin, 2µg/ml pepstatin A , 0.01% polyoxyethylene-10-tridecyl ether

2.3   Plasmids

Commercially available plasmid, pFLAG.ATS, was used in the Kodak IBI expression system and plasmids, pET3a and pET23a, in the Novagen pET expression system. The plasmid construct pET3a/UL50, containing the HSV-1 dUTPase ORF, was a kind gift from O. Björnberg, Dept. Biochemistry, University of Lund, Sweden. Details of this construct are given in Section 5.1.3.2.

2.4   Enzymes

Restriction enzymes were obtained from Boehringer Mannheim or New England Biolabs. DNase I and lysozyme were purchased from Sigma. T4 DNA ligase, T4 polynucleotide kinase, calf intestinal phosphatase, DNA polymerase I (Klenow fragment) T7 DNA polymerase and Taq polymerase were purchased from Boehringer Mannheim. Gene 32 protein was purchased from Pharmacia.

2.5   Synthetic oligonucleotides

Oligonucleotides were synthesised for PCR, sequencing and mutagenesis on a Cruachem PS250 automated synthesiser mainly by myself but also by Mr R. Van Deursen and Mr R. Reid. A detailed description of their synthesis is given in section 3.1.1.

2.6   Peptides

Peptides were prepared by Mrs K. McAulay and Miss G. McVey using a Shimadzu PSSM-8 peptide synthesiser. Peptides with purities below 70% were purified by HPLC (Beckman System Gold).

2.7   Bacteria

E.coli strains

DH5α [F⁻,*sup*E44 Δ*lac*U169 (φ80*lac*Z ΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1
   *rel*A1]

Used for maintenance and propagation of plasmid DNA

BL21(DE3)pLysS [*hsd*S *gal* (λcIts857 *ind*1 sam7 *nin*5 *lac*UV5-T7 gene 1)]

Used for expression of recombinant proteins using the pET T7 system
(Novagen, Inc.) T7 RNA polymerase is carried on bacteriophage λDE3 which
is integrated into the chromosome of BL21 (Studier *et al.*, 1990).

CHS26 DAM⁻ [F⁻, *ara*B/(*lac pro*) *thi, str*A]

Used for propagation of plasmids with DAM methylation sites allowing efficient
restiction digestion.

NM522 [*rec*A⁺, *sup*E, *thi*, Δ(*lac⁻pro*AB), *hsd*5(F', *pro*AB, *lac*IQ, *lac*ZΔM15)]

Used as a recipient host for mutagenic plasmids during Kunkel mutagenesis.

BW313 [*dut⁻, ung⁻, rec*A1, *thi*-1, *rel*A *spo*T1/F'*lys*A, Hfr KL16(P045)]

Used for the production of uracil enriched plasmid DNA during Kunkel
mutagenesis and as a dUTPase negative host and control strain.

2.8   Bacteria Culture Media

L-Broth   10g/l NaCl, 10g/l Difco Bactotryptone, 5g/l yeast extract, pH 7.5
2YT Broth   5g/l NaCl, 16g/l Difco Bactotryptone, 10g/l yeast extract, pH 7.0
L-Broth Agar 1.5% (w/v) agar in L-broth

Antibiotics were used at the following concentrations as required;
Ampicillin   100μg/ml for bacteria containing pFLAG.ATS and pET plasmids
Chloramphenicol 25μg/ml for strain BL21(DE3) harbouring the pLysS plasmid
Kanamycin   70μg/ml was used for propagation of the K07 helper phage
Tetracycline   20μg/ml was used for VZV dUTPase donor plasmid KpnC29

2.9   Radiochemicals

Radiochemicals were purchased from Amersham International plc. with the following specific activities:

Sequencing label:
[α-^{35}S] deoxyadenosine thiotriphosphate, 1000Ci/mmol (10μCi/μl)
[α-^{35}S] deoxycytosine thiotriphosphate, 1000Ci/mmol (10μCi/μl)

Terminator cycle sequencing labels:
[α-^{33}P] ddGTP   ⎤
[α-^{33}P] ddATP   ⎟   all at 1500Ci/mmol, 450μCi/ml
[α-^{33}P] ddTTP   ⎟
[α-^{33}P] ddCTP   ⎦

dUTPase assay label:
deoxy[5-^{3}H] uridine 5'triphosphate, 19Ci/mmol (1μCi/μl)

2.10   Other materials

DNA purification
Pharmacia Biotech Sephaglas Band Prep Kit
DNA sequencing
Amersham Life Science, United States Biochemical (USB)
          -Sequenase Version 2.0 DNA Sequencing Kit
          -Sequenase Quick-Denature Plasmid Sequencing Kit
Amersham Life Science
          -Thermo Sequenase Cycle Sequencing Kit
          -Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit
High efficiency ligation
Amersham T4 Ligation Kit
Preparation of plasmid DNA
Qiagen QIAprep Spin Plasmid Mini-Prep Kit (small scale)
Promega Wizard DNA Maxi-Prep Kit (large scale)
Protein concentration
Amicon Centriplus Concentrators
Spectrum Spectra/Por 3 dialysis membrane

Chapter 3 - Methods

3.1 DNA manipulation

3.1.1 Oligonucleotide synthesis and purification

<u>Chemical Synthesis</u>

Synthesis was by cyanoethyl phosphoramidite chemistry on controlled pore glass (CPG) columns. The first nucleoside which forms the 3' end of the synthetic DNA is attached to the CPG surface via an ester linkage and a hydrocarbon spacer. To allow the correct polarity of chemical synthesis (3' → 5') the 5'OH is protected with a dimethoxytrityl group (DMT). Mononucleotides are added sequentially using the following method:

1. Deprotection - the 5' DMT acid labile protecting group is removed producing the 5' OH to react with the next base.
2. Coupling - phosphoramidite is activated by tetrazole and covalent bond formed
3. Capping - chains with failed additions are capped
4. Oxidation - the phosphite internucleotide bond is oxidised.

Steps 1-4 are repeated for each base addition.

The ester linkage to the CPG can then be cleaved with ammonium hydroxide (0.88 specific gravity) leaving the DNA with a free 3' OH. The protecting groups of A, G and C are removed by incubation for 5hr at 55°C and the DNA is lyophilised under vacuum. For standard sequencing reactions, oligonucleotides were resuspended in water and used directly. For more stringent applications such as mutagenesis, oligonucleotides were resuspended in deionised formamide and TBE then gel purified.

<u>Oligonucleotide purification</u>

Purification was performed by denaturing polyacrylamide gel electrophoresis on a 12% gel (acrylamide:bisacrylamide 19:1) with 8M urea, and 1X TBE. Polymerisation was by addition of 0.01vol 10% APS and 0.001vol TEMED. Samples were heated at 100°C for 1min with an equal volume of formamide before loading. Formamide dyes were loaded in adjacent wells. Samples were electrophoresed in gels (1.5mm thick x 25cm long) at 250 V for 3-5hr in 1X TBE running buffer. Bands were visualised under short wave UV light by shadowing with an intensifying screen and excised from the gel. Urea was removed from the slices by washing for 2min in 3ml of water. DNA was passively eluted by washing the slice in 3ml TE buffer overnight at 37°C with shaking.

Fragments of acrylamide were removed from the solution by spinning for 5min in a microfuge and transferring the supernatant to a 30ml siliconised glass Corex tube. DNA was precipitated with 2.5vol of 100% ethanol plus 0.1vol of 3M sodium acetate (pH 4.6), incubated on dry ice for 1hr and spun at 10K (Sorvall, SS34 rotor) for 10min. Pellet was recovered in 100μl of water and diluted accordingly based on the absorbance at 260nm.

3.1.2 Polymerase chain reaction (PCR) amplification of DNA

PCR was carried out using various primer pairs designed to amplify the chosen target sequence, give restriction sites for cloning and introduce stop codons in the target open reading frame. Taq DNA polymerase was employed in the majority of PCR reactions. Pfu DNA polymerase, which lacks terminal transferase activity, was used occasionally depending on the cloning protocol. A standard 50μl reaction was as follows:

Reagent	Final concentration
Buffer (10X)	1X
Template DNA	< 1μg/100ml
Primers	0.1-1.0μM each
dNTPs	20-200μM each
Taq DNA polymerase	0.5-2.5U
H_2O	up to 50μl

The standard 10X PCR buffer supplied by Boehringer Mannheim contains 15mM $MgCl_2$. PCR buffer was prepared from stock chemicals when titrating Mg^{2+} levels down to 1mM. All samples were overlaid with 50μl mineral oil to prevent evaporation. A standard set of cycling temperatures is as follows:

(95°C, 4min / 55°C, 2min / 72°C, 3min) x 1
(95°C, 1min / 55°C, 2min / 72°C, 3min) x 25

These reaction concentrations and cycling conditions were varied for each template / primer set in order to obtain a good product yield without non-specific primer annealing. Titration of template DNA, primers, and Mg^{2+} concentrations proved to be the most useful.

3.1.3 Agarose gel electrophoresis

DNA fragments generated by PCR or restriction digests were analysed on 0.8-1.5% agarose gels containing 0.5μg/ml ethidium bromide in 1X TBE. Samples were prepared in loading buffer (DF dyes) and electrophoresis was performed at 100V for 1-3 hr. DNA bands were visualised using short wave UV light and photographed with either Polaroid film (type 667) or using a digital camera.

3.1.4 Purification of DNA fragments

DNA fragments were separated by electrophoresis using 3% NuSieve GTG Agarose to allow good separation and easy melting of the gel. Fragments were visualised under long wave UV light for very short periods to minimise DNA damage. Appropriate bands were excised and purified using the Sephaglas Band Prep kit (Pharmacia). In this method DNA was bound to a silica matrix in the presence of a high concentration of sodium iodide (Vogelstein & Gillespie, 1979).

Excised agarose blocks were dissolved with 1μl of gel solubiliser (NaI solution) per mg of agarose. The Sephaglas (silica matrix) was added (5μl per μg DNA), vortexed and left to bind at room temperature for 5min. Three rounds of brief centrifugation and washing with wash buffer (8X volume of Sephaglas added) were performed before a final elution in 20μl (or 4X volume of Sephaglas) of elution buffer at RT for 5min. The matrix was then pelleted by a brief centrifugation and the supernatant, containing the purified DNA, was recovered and stored at -20°C.

3.1.5 DNA restriction digests

Restriction enzyme digestion was carried out in 20μl volumes at 37°C (or temperature specified by the supplier). Samples were digested in the appropriate buffer (Boehringer Mannheim system) using approximately 1 unit of enzyme per 0.5μg DNA for 1-3hr. Generally 0.5μg of DNA was digested for agarose gel analysis and up to 10μg for the isolation of specific restriction fragments. Samples prepared using the miniprep method (section 3.1.10) were digested in the presence of RNaseA at 10μg/ml. Digestion with the restriction enzyme KpnI was performed with the addition of 100μg/ml BSA.

3.1.6 DNA cloning

Plasmid DNA was linearised for cloning using the appropriate restriction enzymes. Double digestion was monitored by checking the efficiency of each restriction enzyme separately and also testing for the re-ligation of double digested vectors. End repair (phosphate removal from 5' ends) was performed as necessary using calf intestinal phosphatase (30min, 37°C) followed by phenol/chloroform extraction

and ethanol precipitation. Restriction fragments generated by enzymes that produce 5' extensions were made blunt ended by incubating with 50µM of each dNTP and 0.5 unit/ µg of Klenow polymerase for 30min at room temperature. Again, samples were extracted with phenol/chloroform and ethanol precipitated. Ligation was performed in 20µl volumes at 16°C for 3hr using 2 units of T4 DNA ligase in ligation buffer with various ratios of vector:insert. Alternatively the Amersham T4 Ligation system was employed allowing higher ligation efficiency with an incubation time of approximately 30min (Hayashi *et al.*, 1986).

3.1.7 Transformation for growth and maintenance of plasmid DNA

The method of Hanahan (1983) was used to make competent *E.coli* giving up to 100 times the number of recombinants obtained compared to standard calcium chloride methods. An individual colony from a streaked plate of bacteria was picked into 10ml of 2YT broth and grown up overnight at 37°C with shaking. 100ml of 2YT was inoculated with 1ml of the overnight culture and grown at 37°C, with shaking, to an OD_{600} of 0.4-0.6. 30ml of culture was spun at 2K rpm (Sorvall SS34 rotor) for 10min at 4°C. The pellet was gently resuspended in 2.5ml cold TFB and incubated on ice. 100µl of DMSO was added after 15min, followed by 100µl DTT/KAc solution after a further 10min. A final 100µl aliquot of DMSO was added after a further 5min. Cells were used directly and not stored. Generally 10µl of the ligation mix, about 10ng plasmid (section 3.1.6), was added to 200µl of competent cells and incubated on ice for 45min. Cells were heat shocked for 3min at 42°C before spreading on L-broth agar plates containing appropriate antibiotics. Plates were incubated overnight at 37°C and screening was performed by picking single colonies and extracting DNA by the miniprep method (section 3.1.10). Clones were then analysed by mobility change or restriction profile using agarose gel electrophoresis (section 3.1.3). Positive clones were sequenced to determine the correct positioning of the insert in the vector and to check for the absence of mutations.

3.1.8 Transformation for protein expression

A modification to the protocol in 3.1.7 was used when transforming bacteria for expression of recombinant proteins in the pET system. Glycerol stocks of BL21(DE3)pLysS were streaked to single colonies on LB-agar plates with antibiotics and grown overnight at 37°C. One isolated single colony was picked and streaked on another LB-agar plate and grown at 37°C overnight. This step was repeated again. Several single colonies were picked into 200µl cold TFB and incubated for 30min before addition of the transforming plasmid. No DMSO or DTT/KAc solutions were

used. Transformation was as in section 3.1.7. Repeated streaking of bacteria to single colonies improved expression and reduced variation between different cultures.

3.1.9 Glycerol stocks

Glycerol stocks were prepared by inoculating 10ml of L-broth containing appropriate antibiotics with a single bacterial colony and incubating overnight at 37°C in an orbital shaker. Aliquots of 1.5ml were spun briefly in a microfuge and pellets were resuspended in 0.75ml of 2% Bactopeptone plus 0.75ml of 80% glycerol. Stocks were kept at -20°C and -70°C. All glycerol stocks were streaked to single colonies before use.

3.1.10 Miniprep plasmid DNA preparation

Small scale preparation of plasmid DNA was performed using three methods depending on the final quality of DNA required.

Method 1 - for bulk screening

10ml of bacteria were grown up at 37°C with shaking overnight. A 1.5ml aliquot was pelleted by spinning in a microfuge for 20sec at 13,000 rpm and resuspended in 350µl of STET buffer plus 25µl (10mg/ml) of freshly prepared lysozyme solution. Tubes were vortexed and then placed in a boiling water bath for 40sec. Cell debris was pelleted by centrifugation for 15min. Pellets were removed with a toothpick and discarded. DNA was precipitated by addition of 40µl 2.5M NaOAc plus 420µl isopropanol and samples were incubated on dry ice for 30min. The solution was then spun at 13K in a microfuge for 10min and pellets were recovered, air dried and dissolved in 10-20µl of 1X TE or water. Approximately 25% of sample was electrophoresed and visualised on an agarose gel.

Method 2 - for sequencing

15ml of bacteria were grown up at 37°C with shaking overnight. Bacteria were spun at 10K rpm (Sorvall SS34 rotor) for 15min, resuspended in 1ml lysozyme solution and left at room temperature for 5min. 2ml of NaOH/SDS solution was added and incubated for 5min on ice. The solution was neutralised by the addition of 1.5ml KAc solution, left on ice for 5min then spun at 10K for 15min at 4°C. The supernatant was extracted with phenol/chloroform then precipitated for 30min after the addition of two volumes of ethanol. The sample was spun at 10K for 10min at 4°C and the pellet was washed with 70% ethanol, dried and dissolved in 100µl TE. 2µl of 10mg/ml RNase A was added and incubated at 37°C for 30min. DNA was precipitated by the addition of 60µl 20% PEG/2.5M NaCl and left on ice for 1hr. Sample was spun in a microfuge for

5min then washed with 70% ethanol and dried. Pellet was resuspended in 18μl TE and DNA was denatured by adding 2μl of 2M NaOH and incubating at room temperature for 5min. Solution was neutralised by the addition of 8μl 5M ammonium acetate and precipitated with 100μl of ethanol on dry ice for 5min. The sample was spun in a microfuge for 10min then washed again with 70% ethanol and dried. The pellet was dissolved in 20μl TE and 2μl were used per sequence reaction.

Method 3 - for sequencing

Plasmid DNA was also prepared using Qiagen mini-columns using the manufacturer's standard protocol. Basically overnight cultures were pelleted, resuspended and lysed with NaOH/SDS. The cleared lysate containing the DNA was then adsorbed to a silica-gel membrane which was washed and eluted. In the case of low copy number plasmids the starting culture was increased from the standard 1.5ml up to 10ml to increase the yield of plasmid DNA. DNA produced from these columns was suitably clean to allow sequencing after denaturation.

3.1.11 Large scale plasmid DNA purification

Most large scale preparations of plasmid DNA were done using the Promega Wizard kit which has the advantage of speed and reproducibility and does not involve the large quantity of ethidium bromide used in CsCl purification. The Qiagen large scale plasmid purification system was also tried but was slower and produced a significantly lower yield (10X) of plasmid DNA per litre of bacterial culture.

Purification by Wizard Maxipreps

Method was performed using the standard commercial protocol. Briefly:
A 500ml culture was grown up at 37°C overnight with shaking. Cells were pelleted by centrifugation at 8K rpm (Sorvall GSA rotor) for 10min, 4°C and resuspended in 15ml of resuspension buffer containing 100μg/ml RNase A. Cells were lysed with 15ml lysis solution (0.2M NaOH, 1% SDS) for up to 20min until solution was clear and viscous. 15ml neutralisation solution was added (1.32M KAc, pH 4.8) and the sample was centrifuged at 8K rpm (Sorvall GSA rotor) for 15min, 4°C. The supernatant was decanted and 0.6vol isopropanol were added. Sample was centrifuged at 8K rpm (Sorvall SS34 rotor) for 15min, 4°C and the DNA pellet resuspended in 2ml TE. The DNA solution was then bound to 10ml of the Wizard DNA binding resin and poured into a column. The bound DNA was washed several times with ethanol solution before elution with 1.5ml of preheated (65-70°C) TE. Up to 1mg of plasmid DNA was obtained from 500ml of bacterial culture.

3.1.12 Preparation of ssDNA by phage rescue

The basic method involves infecting a culture with single stranded phage and allowing replication of the plasmid carrying a single stranded origin of replication. Phage is then recovered, precipitated and ssDNA extracted.

A single bacterial colony was inoculated into 5ml 2YT broth with antibiotics and K07 helper phage at 10^8pfu/ml. The culture was grown for 1-2hr at 37°C with vigorous shaking. Kanamycin was added to a final concentration of 70μg/ml and the culture was grown at 37°C for a further 16-24hr. 1.5ml aliquots of bacterial culture were centrifuged for 10min and 1ml supernatant collected in a fresh Eppendorf. Phage particles were precipitated by the addition of 300μl of 20% (w/v) PEG / 2.5M NaCl and incubation on ice for 15min. Samples were microfuged for 5min at 9K rpm and supernatant discarded. This step was repeated to remove all traces of PEG which can produce background smearing during sequencing. The pellet was resuspended in 400μl 0.3M sodium acetate, pH 6.0 / 1mM EDTA and vortexed. DNA was extracted with one vol phenol/chloroform for 1-2min, ethanol precipitated and resuspended in 10μl. Generally 5μl was used per sequence reaction.

3.1.13 DNA sequencing

Sequencing of DNA was performed using several different commercial kits all based on the dideoxy chain termination method of Sanger *et al.* (1977).

Sequenase 2.0 kits (United States Biochemical)

This method is split into four sections: template preparation, annealing, labelling and termination.

Template preparation:

Single stranded template DNA was prepared either as in section (3.1.10) or in one of the following ways:

1. NaOH denaturation

Purified plasmid DNA (0.5-3μg) was denatured at 37°C for 10min in an 11μl reaction volume containing 1μl primer (5pmol) and 2μl 1.0M NaOH. The sample was then cooled on ice and neutralised by the addition of 2μl 1.0M HCl.

2. Glycol heat denaturation

Purified plasmid DNA (0.5-3μg) was combined with primer (5pmol) and a denaturing mix (10mM Tris-HCl, pH 7.5, 1mM EDTA, 50% glycerol, 50% ethylene glycol) up to a total volume of 13μl. Samples were incubated at 90-100°C for 5mins then cooled on ice.

3. Preparation of ssDNA using bacteriophage (See section 3.1.12)

<u>Annealing:</u>

Plasmid reaction buffer (400mM Tris-HCl, pH 7.5, 100mM MgCl$_2$, 250mM NaCl) was added to a total volume of 15µl and the primer/template/buffer mix was annealed at 37°C for 10min.

<u>Labelling:</u>

The commercial labelling mix was diluted to 1.5µM 7-deaza-dGTP, 1.5µM dCTP, 1.5µM dTTP and the following solution was made up:

Ice cold annealed DNA mixture	15µl
DTT, 0.1M	1µl
Diluted labelling mix	2µl
5µCi [α-^{35}S]-dATP	0.5µl
Sequenase T7 DNA Polymerase	2µl (3U)

Solution was mixed and incubated for 2-5min (5-10min for glycol denatured plasmid) at RT.

<u>Termination:</u>

Four aliquots of 4.5µl from the labelling reaction were transferred into 2.5µl of each of the four termination reactions (containing 3 dNTPs and one ddNTP) and incubated at 37°C for 5min. Termination solutions contain: 80µM of three dNTPs, 8µM of the remaining ddNTP and 50mM NaCl. Reactions were stopped by the addition of 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) or freezing at -20°C before subsequent loading.

Thermo Sequenase

Thermo Sequenase is a thermostable DNA polymerase used with the cycle sequencing method. This method uses repeated cycles of thermal denaturation, annealing and extension/termination for increased signal levels relative to standard sequencing. Advantages include the very small amount of template required and the complete denaturing afforded at each cycle allows greater readability of high G+C areas. Two different kits were used.

<u>1. Standard cycle sequencing</u>
Labelling:

An oligonucleotide primer is designed which will allow a brief extension in a reaction mix containing one labelled dNTP and only two non-labelled dNTPs. This results in termination of the labelling step when the missing base is reached in the template. The

reaction is thermally cycled (typically 60°C-90°C, 30-50 times) and results in the production of labelled, extended primer whose length is dependant on the template sequence.

Chain termination:
In the second step the concentration of all the dNTPs is increased and the reaction is split into four tubes each with a different ddNTP. The reactions are thermally cycled (typically 95°C-72°C, 30-60 times) until all the growing DNA chains are terminated by a ddNTP. The final products are denatured with formamide and heating before subsequent loading.

2. Radiolabelled terminator cycle sequencing
The methodology behind this technique is based on standard cycle sequencing with the exception that label is incorporated into the sequencing reaction by the use of four $[\alpha\text{-}^{33}P]$ddNTP terminators in a one-step reaction. Four reactions are set up containing all four dNTPs with a lower concentration of one of the four labelled ddNTPs. Reactions are thermally cycled until all chains have terminated. Efficiency is greater than other methods because only terminated chains are labelled therefore there are no stop artifacts and a there is a reduction of background bands. There is also a greater flexibility as there is no constraints on the primer sequence as in the standard cycle sequencing.

Visualisation of sequenced products
Denaturing electrophoresis for all sequencing methods was performed essentially as in section 3.1.1 with the following modifications. Sequencing gels (0.35mm thick and 35cm long) were prepared using 6% acrylamide and 0.5X TBE. Samples were heated at 100°C in formamide dyes and loaded on a pre-run gel (20min) and then electophoresed at 100W for 1hr. This is a higher wattage than most protocols recommend as it was found that the subsequent increase in temperature in the gel allowed better denaturation and elucidation of compressions. Glycerol tolerant gel buffer (taurine replaces boric acid) was used when the glycol method of plasmid denaturation was performed. Plates were separated and gel was washed on one plate with 10% acetic acid (X3) thus removing the urea. Gels were transferred onto paper and dried under vacuum at 80°C for 1hr. Gels were exposed to film from between 1 and 5 days depending on the intensity of the signal. Autoradiographs were developed for reading using an automatic X-Omat machine.

3.1.14 Kunkel mutagenesis

This method is a version of that of Kunkel (1985; 1987). Uracil enriched DNA was made by transforming the *E.coli* strain BW313 with a recombinant pET-23a plasmid carrying a single stranded origin of replication. BW313 lacks dUTPase and uracil DNA glycosylase allowing incorporation of uracil into newly replicated plasmid. Single stranded uracil enriched DNA is produced by infection and precipitation of the M13 R408 helper phage. A mutagenic olionucleotide was annealed to the recovered plasmid DNA and a new strand was synthesised *in vitro*. This double stranded plasmid was transformed into the *E.coli* strain NM522 which carries an active uracil DNA glycosylase (Mead *et al.*, 1985). This enzyme catalyses the destruction of the uracil containing strand leaving the synthesised mutagenic strand to act as a template for replication. dsDNA plasmid mutants were then recovered from the bacteria using the miniprep method (section 3.1.10).

Preparation of uracil enriched ssDNA:

10ml of 2YT broth plus 100μg/ml ampicillin and 100μg/ml uridine was inoculated with 50μl glycerol stock of *E.coli* BW313 and grown overnight at 37°C with shaking. The overnight culture was added to 200ml 2YT plus 100μg/ml uridine in a 2l baffled flask. The culture was grown at 37°C to an OD_{600} of 0.3 prior to addition of M13 R408 helper phage to 5.04x10^9 pfu/ml. Growth was continued for 8hr at 37°C with shaking. The culture was centrifuged for 20min at 7K rpm (Sorvall GSA rotor), 4°C. Supernatant was collected and the spin was repeated. Supernatant was collected in a 300ml glass bottle and stored at 4°C overnight. Phage was precipitated by addition of 0.25vol 3.75M ammonium acetate / 20% PEG and incubation on ice for 30min. Sample was centrifuged at 9K rpm (Sorvall GSA rotor) for 30min, supernatant discarded and the pellet resuspended in 200ul TE. DNA was recovered by two chloroform extractions, six phenol/chloroform extractions and one final chloroform extraction. The aqueous layer was removed and DNA precipitated with 0.36vol 7.5M ammonium acetate and 2.5vol ethanol and incubated on ice for 20min. Sample was spun for 20min at full speed in a microfuge, and pellet was washed with 70% ethanol and dried before resuspension in 50μl water. DNA was visualised by agarose gel electrophoresis and concentration determined by measuring the OD at 260nm.

<u>Mutagenic strand synthesis:</u>

Mutagenic oligonucleotides were kinased at 37°C for 30min, followed by 15min at 65°C to stop the reaction. A typical reaction with final concentrations was as follows:

Oligonucleotide	500pmol
Ligation buffer	1X
T4 Kinase	10U
H_2O	up to 20µl

80µl of water was added to the stopped reaction to give a kinased oligonucleotide stock of 5pmol/µl. 0.1 pmol/µl of the prepared uracil enriched ssDNA was annealed to 1µl of the kinased oligonucleotide stock with 2µl TM buffer in a 10µl total volume. Samples were annealed at 37°C for 30min. T7 DNA polymerase was found to be the most efficient enzyme for second strand synthesis. A typical reaction is as follows:

Annealing mix (above)	10µl
SEQ buffer (5X)	2µl
dNTPs (5mM)	1µl
ATP (5mM)	1µl
DTT (100mM)	1µl
Acetylated BSA (1µg/ul)	2µl
Gene 32 protein	0.5µl
T4 DNA ligase	3µl (3U)
T4 kinase	0.5µl (5U)
T7 DNA polymerase	1µl (1U)

The reactions were incubated at 37°C for 1hr then stored at -20°C before transformation into the *E.coli* strain NM522. Single colonies were picked and screened by the miniprep method (Section 3.1.10). Initial mutations were designed to allow screening by restriction digest allowing optimisation of the technique. Subsequent mutations were screened by single track sequencing.

3.2   Polypeptide analysis

3.2.1   Expression of recombinant proteins

It was found that the quality of the bacteria carrying the recombinant expression plasmid was vital to maintain consistent and stable expression of recombinant protein. Care was taken to streak host strain several times after recovery from glycerol stocks.

Plasmids were transformed immediately before expression and overnight cultures were avoided as much as possible. Two bacterial expression systems were utilised:

(a) Kodak IBI system

The initial system for the production of recombinant proteins was the Kodak IBI FLAG system. This system was based on a *tac* promotor (a hybrid between the *trp* promoter and the *lac* operator) allowing induction of expression under the standard *lac* system via IPTG. Transcripts were generated as fusion proteins to the FLAG peptide. This octapeptide marker was designed to allow detection and purification of the recombinant fusion protein with specific FLAG monoclonal antibodies. This peptide had an enterokinase cleavage site for removal of the fusion to yield the native protein. The OmpA signal peptide could also be fused to the protein to allow secretion of the protein to the periplasmic space (see Section 5.1.2).

(b) pET system

The majority of proteins were expressed in the pET vector system originally constructed by Studier (1990) and now sold by Novagen Inc. This system utilises the strong bacteriophage T7 promoter. Expression was induced by providing a source of T7 RNA polymerase which is not present in the normal host cell. This allowed propagation of recombinant plasmids without the selective pressure associated with low levels of uninduced expression in other systems. Plasmids were then transferred into the BL21(DE3)pLysS strain which carries a chromosomal copy of the T7 RNA polymerase under *lac*UV5 control. Induction was then performed with the addition of IPTG.

General protocol for expressing recombinant dUTPase

Ampicillin was used at 150μg/ml to maintain the pET construct and chloramphenicol was used at 25μg/ml to maintain the pLysS plasmid. A 10ml culture was prepared by inoculation with a single colony of freshly transformed BL21(DE3)pLysS bacteria containing the appropriate pET construct. This culture was then used to inoculate 300ml of L-broth containing the above antibiotics. Cultures were grown at 30-37°C until they reached an OD of 0.7 at 600nm. Induction of recombinant protein synthesis was performed by the addition of IPTG at concentrations ranging from 0.1-1.0mM. Cultures were grown at temperatures ranging from 20-37°C depending on the construct used. Lower temperatures were found to increase protein solubility in some cases.

3.2.2 Protein extraction

Proteins were extracted in buffers appropriate to the expression system being used. Generally 250ml aliquots of culture were chilled on ice for 10min and then centrifuged for 15min at 10K rpm (Sorvall GSA rotor), 4°C. Supernatant was discarded and the pellet resuspended in 50ml extraction buffer. In the Kodak IBI system, lysis was performed by the incorporation of lysozyme and detergent in the extraction buffer (section 2.2). In the pET system, detergent and freeze thawing (X3) was sufficient to disrupt the cell membrane releasing the indigenous lysozyme of the BL21(DE3)pLys *E.coli* strain resulting in efficient lysis. Variations on the constituents of the extraction buffers and their effects are discussed in the Results section. The viscosity of the lysed solutions was reduced by passing through a thin syringe needle (X3) before centrifugation for 60-90min at 10K rpm (Sorvall SS34 rotor), 4°C. Supernatants containing the soluble protein fraction were stored at -20°C. Samples were centrifuged again immediately prior to purification by FPLC. Protease inhibitors were maintained throughout the extraction process.

3.2.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

For analysis of proteins single concentration slab gels between 9-12.5% acrylamide were used. Stock solutions of 30% acrylamide were prepared using a ratio of 39:1 acrylamide to N,N'-methylene bisacrylamide in water, filtered through Whatman No. 1 filter paper and stored at 4°C. Gels were prepared at the appropriate concentration from this stock solution with freshly made buffer to a final concentration of 375mM Tris-HCl (pH 8.9) and 0.1% sodium dodecylsulphate (SDS). Polymerisation was instigated with the addition of freshly made ammonium persulphate and TEMED to a final concentration of 0.06% and 0.04% respectively.

For large gels (20 x 22cm) the gel solution was poured between two thoroughly washed glass plates separated by 1.5mm thick spacers and sealed with rubber tubing. The gel was left to polymerise at room temperature under a thin layer of butan-2-ol. The butan-2-ol was removed and the stacking gel was poured consisting of 5% acrylamide in 122mM Tris-HCl (pH 6.7), 0.1% SDS with wells formed using a teflon comb.

SDS-PAGE was also performed using the Bio-Rad miniprotean II apparatus allowing faster resolving times and smaller volumes. These gels were not appropriate for the resolution of some recombinantly expressed proteins in a background of closely migrating *E.coli* proteins.

Samples were boiled for 5min in denaturing buffer with sufficient bromophenol blue to visualise the dye front. Electrophoresis was carried out in a buffer containing 52mM Tris, 53mM glycine and 0.1% SDS. Gels were run at 100V until dye front had

reached the bottom, approximately 3hr for large gels and 1hr for mini gels. Proteins were visualised by fixing and staining in Coomassie brilliant blue R250 solution and destained with several washes in 5% methanol and 7% acetic acid.

3.2.4 Purification by FPLC

Samples prepared as in section 3.2.2 were purified by phosphocellulose and Mono S chromatography using the Pharmacia FPLC system. Phosphocellulose (Whatman P11) columns were prepared fresh according to the manufacturers instructions. Phosphocellulose was pretreated with 25vol 1M NaOH for 5min and washed 3 times with water. The phosphocellulose was then washed with 25vol 1M HCl and washed three times in 1X PBS before packing into the columns. Equilibration was performed overnight in the appropriate buffer. Mono S cation exchange columns were from Pharmacia LKB and were used according to the manufacturers instructions. Columns were washed thoroughly and equilibrated before use.

Generally 10-50ml of sample was applied to the phosphocellulose column at a flow rate of 1ml/min. A gradient of 0-1.0M NaCl was used to separate protein species by ion-exchange chromatography and fractions of 1.5ml were collected and analysed for activity. 4ml of the most active fractions were diluted 1/10 in Mono S loading buffer and loaded directly onto the Mono S columns. A gradient of 0-1.0M NaCl was used to separate proteins and fractions of 1-1.5ml were collected. Samples were analysed for activity by enzyme assay (section 3.2.6) and purity by SDS-PAGE (section 3.2.3).

3.2.5 Protein quantification

The total protein concentration was measured using the Bradford assay with BSA as a standard (Bradford, 1976). Reactions were performed in flat bottom microwell plates with a Labtech Anthos HT2 plate reader. This equipment allows instant determination of a standard curve by absorbance at 620nm and calculates the protein concentration of the samples. To quantitate the amount of soluble dUTPase in each purified sample it was necessary to use SDS-PAGE (section 3.2.3). A dilution series of each purified sample was examined by SDS-PAGE and compared with other samples and to standard dilutions of BSA. As each sample contained a different proportion of dUTPase protein compared to the total protein, this allowed quantitation of the dUTPase enzyme specifically. This method was also used as a measure of relative enzyme purity.

3.2.6   Analysis of enzyme activity

Analysis of dUTPase activity was performed using two methods, both utilising tritiated dUTP as a substrate marker. TLC was the preferred method as both substrate (dUTP) and reaction product (dUMP) could both be measured.

dUTPase reaction

A typical reaction solution was as follows:

Potassium phosphate, pH 6.5	0.4M
$MgCl_2$	20mM
DTT	20mM
dUTP	2mM
$[^3H]$dUTP	0.5μCi
H_2O	up to 10μl

Reaction mix was preheated to 37°C and 10μl enzyme sample is added. Note that the final dUTP concentration was 1mM. Reaction was stopped after 10min by the addition of 2μl 80% formic acid.

Filter disk assay

This is a version of the method by Williams and Cheng (1979). Samples were spotted onto Whatman DE81 paper disks and washed thoroughly in 4M formic acid / 1mM ammonium formate three times with a final wash in 95% ethanol. Disks were then dried and the bound, non-decomposed $[^3H]$dUTP was quantified by scintillation counting.

TLC Assay

This is essentially the method of Tye *et al.* (1977). Polyethyleneimine-cellulose sheets (Polygram, Machery-Nagel, Germany) were pre-washed with 1M formic acid and dried. Samples were spotted (4x 0.5μl) onto the sheets with markers (dUTP and dUMP) and TLC was performed in a tank with 1M formic acid and 0.5M LiCl. Plates were dried and spots were visualised under short wave UV light. Spots were excised and quantified by scintillation counting.

Scintillation counting

Quantification of radiolabeled samples was performed by immersing either the DE81 disks or the PEI spots in Ecoscint A. Samples were counted on either a Packard 1600TR or Beckman LS 5000CE liquid scintilation counter.

3.3 Computer-based analysis

3.3.1 Sequence analysis programs

Amino acid primary sequence data was examined using various programs in the GCG program set running under VAX/VMS. Sequences of dUTPases were obtained from the Swiss-Prot database except HSV-2 dUTPase which was sequenced in the MRC Virology Unit. Sequence alignments were performed using Pileup, a simplification of the method of Feng and Doolittle (1987). The program SAGA was also used for alignment comparisons (Notredame *et al.*, 1996). TBALSTX (Altschul *et al.*, 1990), available online at the National Centre for Biotechnology Information (NCBI), was used for primary sequence searches in the Brookhaven Protein Databank (Bernstein *et al.*, 1977). This server can be accessed at www.ncbi.nlm.nih.gov.

3.3.2 Structure prediction

Secondary structure prediction was performed using the PredictProtein server at the Protein Design Group, EMBL (www.embl-heidelberg.de/predictprotein). This server can automatically generate sequence alignments using the MaxHolm method (Sander & Schneider, 1991) although most sequences were pre-aligned using the programs above (section 3.3.1). The program PHDsec (Rost & Sander, 1993; Rost & Sander, 1994; Rost, 1996) was employed to predict secondary structure.

Threading based analysis of specific sequences was performed using the program Threader available at the Biomolecular Structure and Modelling Unit at University College London (Jones *et al.*, 1992). The program Threading Analyst (encompassing the front-end graphical interface, tan.tcl) was used for interpretation of the Threading data (Miller & Thornton, 1995).

Generation of 3D models based on the HSV-1 dUTPase primary amino acid sequence was performed using the ProMod program at the Swiss-Model Protein Modelling Server (Peitsch, 1995; Peitsch, 1996). This server is available online at www.expasy.ch/swissmod/SWISS-MODEL.html.

The calculation of molecular surfaces was performed using the programs NewArea (Lee & Richards, 1971) and Sleuth (Dudek & Ponder, 1995) available at Washington University (ftp://dasher.wustl.edu/pub/sleuth/). Secondary structure assignment based on PDB structural coordinates was performed using the program STRIDE (Frishman & Argos, 1995) with details at the EMBL website, (www.embl-heidelberg.de/argos/stride), available at ftp.ebi.uk/pub/software/unix/stride.

3.3.3 Molecular visualisation tools

The programs that were used in order to generate diagrammatic representations of the molecules analysed in this thesis are listed below. Both the programs Insight II and 'O' were run on a Silicon Graphics Workstation under the Unix operating system. All other viewing programs were run in the PC Windows environment.

- **Insight II** by BIOSYM / MSI
  (www.msi.com)
- **WebLab ViewerPro** by Molecular Simulation Inc.
  (www.msi.com/weblab/)
- **Swiss-PdbViewer** by Glaxo Wellcome Experimental Research
  (www.expasy.ch./spdbv/mainpage.html)
- **Rasmol** by R. Sayle (1994)
  (ftp.dcs.ed.ac.uk:/pub/rasmol)
- **CN3D** by NCBI
  (www.ncbi.nlm.nih.gov/Structure/cn3d.html)

Chapter 4 – Results of Computer Modelling

4.1   General Introduction

The main goal of the work presented in this chapter was the construction of a model for the HSV-1 dUTPase. The main reason for embarking on such a project was the absence of structural data for any of the class II dUTPases. Since the primary objective of this study was to characterise a member of this group, the HSV-1 dUTPase, a working model was considered to be an important parallel to complement mutagenesis studies. Section 1.4.8 introduced the five motif regions found in both class I and class II dUTPases and the intragenic duplication model proposed by D.J. McGeoch (1990b). This study, based on primary amino acid sequence analysis, constitutes the starting point for modelling of the HSV-1 dUTPase. Crystallographic data became available for *E.coli* dUTPase and was used as the original template for the class II model.

In order to construct this model it was necessary to follow a set of logical steps which are set out in this chapter as individual sections. Compilation of the available information revealed a wealth of primary sequence data for both class I and class II dUTPases. The first step was therefore to utilise the class I primary sequences to determine common elements within this group. This in turn was used to identify key structural elements in the *E.coli* structure that were likely to be consistent throughout the class I group. During the course of this study additional class I structures of FIV, human and EIAV dUTPases were solved by X-ray crystallography and allowed subsequent refinement of these data.

The next step was to use extensive alignments with primary sequence data to identify areas of homology and variance within the class II group. Finally the data from both the class I and class II groups were compiled utilising the set of defined hypotheses detailed in Section 4.3.3.

The primary sequence work was expanded to include secondary structure predictions and homology modelling between the two dUTPase classes. Using this basis it was possible to model distinct structural elements of the class II group enzymes from solved crystallographic data from the class I group.

4.2   Class I dUTPases

4.2.1   Introduction

The key to creating a model for the HSV-1 dUTPase was the understanding of the class I enzymes and how they relate to the class II group.   To do this, it was first necessary to define the areas of consistent homology within the class I group.

4.2.2   Comparison of class I primary sequences

Class I dUTPases from a wide variety of both prokaryotic and eukaryotic organisms share remarkably distinct homologous features.   The most obvious of these features is the presence of the five motif regions identified by D.J. McGeoch (1990b).   These regions of localised homology allow diverse class I sequences to be lined up effectively.   Figure 4.1 shows an alignment of a selection of class I sequences including mammalian, yeast, plant, poxvirus, retrovirus and bacterial species.   The Swiss-Prot protein database contains many more homologous class I sequences which have yet to be validated as functional dUTPases.

There are several key observations which are apparent from class I sequence alignments.   The first is the presence of extremely homologous copies of all five motifs in the same order.   In Figure 4.1, comparison of the highlighted motifs (34 residues in total) between the tomato plant, *Lycopersicon esculentum* and the yeast, *Candida albicans* yields a sequence identity of 94%.   Even more distantly related species such as *E.coli* and human share 79% identity over these motif regions.   The high conservation of these loci suggest that not only do they play a vital role in dUTPase functionality but that all class I dUTPases operate by a homologous mechanism.   This is supported by the structural data available so far.

The second observation is the overall conservation of protein chain length.   There is some variation at the sequence termini with the retroviruses possessing the smallest sequences.   The exact N-terminus for some of these dUTPases is not clear and is consistent with the retrovirus dUTPases being translated as a polyprotein and subsequently cleaved.   What is important to note is the variation between the start of motif 1 and the end of motif 5.   In this selection of sequences this core distance varies by a maximum of 10% between species.   This is consistent with all the class I sequences analysed.

Finally it is apparent that although the interlying regions between the motifs are substantially divergent between species at the sequence level, the distance between each motif is highly conserved.   Structural data demonstrate that the active site consists almost exclusively of motif residues.

```
                 1                                                                                        100
Mpmv   .............SLWGSQ..........LCSSQQKQPISKLTRATPGSALDL..CSTSHTVLTPEMGPQALSTGIYGPLPPNTFG.LILGRESITM
Srv1   .............SLWGGQ..........LCSSQQKQPISKLTRATPGSALDL..SSTSHTVLTPEMGPQALSTGIYGPLPPNTFG.LILGRESITI
Jsrv   .............MLWG...........NTAGSKRTIADLCRATRGSALDL..CATSYTVLTPEMGVQTLATGVFGPLPPGTVG.LLLGRESASL
Mmtv   .............GVKGSG..........LNPEAPFFPIHDLPRGTPGSALDL..SSQKDLILSLEDGVSLVPTLVKGTLPEGTTG.LIIGRESNY
Fiv    ....................MIIEGDGILDKRSEDAYDL..LAAKEIHLL.PGEVKVIPTGVKLMLPKGYWG.LIIGKSSIGS
Omvs   ..................SEIFLAREGSGILPKRAEDAYDL..ICPQEVCIP.AGQVRKIPINLRINLKEDQWA.MVGTKSFAS
Visn   ..................SEIFLAKEGRGILQKRAEDAYDL..ICPQEISIP.AGQVKRIAIDLKINLKKDQWA.MIGTKISF.A
Caev   ..................SEIFLAKEGEGILPKREEDAYDL..ICPEEVTIE.PGQVKCIPIELRLNLKKSQWA.MIATKISMAA
Eiav   ..................MLAYQGTQIKEKRDEDAFDL..CVPYDIMIP.VSDTKIIPTDVKIQVPPNSFG.WVTGKKSM.A
Coxb   .......MTHSVQLK......ILDKRLGSEFPLPAYATTGSALDLRACLDEPLKIE.PDETCLISTGLAIYLGHSNVAATILPRESGLGH
Ecol   .......MMKKIDVK......ILDPRVGKEFPLPTYATSGSALDLRACLNDAVELA.PGDTTLVPTGLAIHIADPSLAAMMLPREGLGH
Vaccc  ............MFNMNINSP.........VRFVKETNRAKSPTRQSPGAAYDL..YSAYDYTIP.PGERQLIKTDISMSMPKICYG.RIAPREGLSL
Varv   ............MFNMNINSP.........VRFVKETNRAKSPTRQSPYAAYDL..YSAYDYTIP.PGERQLIKTDISMSMPKFCYG.RIAPREGLSL
Human  ...................MQ.........LRFARLSEHATAPTRGSARAAYDL..YSAYDYTIP.PMEKAVVKTDIQIALPSGCYG.RVAPREGLAA
Orfv   .......MEFCHTET......LQVVRLSQNATIPARGSPGAALDL..CSAYDCVIP.SHCSRVVFTDLLIKPPSGCYG.RIAPRRG.AV
Fadv   .....MDPFGSSSVPPCSTSDLPEPKLYFVRLSPHAVPPVRATHGAAYDL..FSAYDIKVP.ARGRALVPTDLVFQFPPGCYG.RIAPREGLAA
Calb   .........MTSEDQSLKKQKLESTQSLKVYLRSPKGKVPTKGSALAAYDL..YSAEAATIP.AHGQGLVSTDISIIVPIGTYG.RVAPREGLAV
Scer   ...........MTATSDKV........LNIQLRSASATVPTKGSATAAYDI..YASQDITIP.AMGQGMVSTDISFTVPVGTYG.RIAPRSGLAV
Lyces  MAENQINSPEITEPSPKVQKLDHPENGNVPFFRVKKLSENAVLPSRASSLAAYDL..SSAAETKVP.ARGKALVPTDLSIAVPQGTYA.RIAPREGLAW
Ccv    MDFNTDRGSAIPTGDDRGYQCGVIGDSVRFSVFTGYDAADVSIPKISSPGSAFDL..SVLEDREFI.RGCHYRLPTGLAIAVPRGYVG.IITPRES.QA
                                                          -----                                    ----
                                                          Motif 1                                  Motif 2

              101                                                                         197
Mpmv   K.GLQVYP..GVIFNDYTGEIKIMA.KAVNNIVTVSQGNRIALILLLPLIETDNKV.......QQPYRGQGSFGSSDI.................
Srv1   K.GLQVYP..GVIINDYTGEIKIMA.KAVNNIVTVPQGNRIALILLLPLIETDNKV.......QQPYRGQGSFGSSDIYW..............
Jsrv   K.GILIHP..GVIISDYTGEIKILA.SAPNKIIVINAGQRIALLVPLVIQGKTI.......NRDRQDKGFGSSDAYW..............
Mmtv   KKGLEVLP..GVIISDQGEIKVMV.KAAKNAVIIHKGERIALLLLLPYLKLPNPVI.......KEERGSEGFGSPSHVHW..............
Fiv    K.GLDVLG..GVIIEGARGEIGVIMINVSRKSITLMERQKIALIILPCKH.EVLEQG..KVVMDSERGDNGYGSTGVFSSW...........
Omvs   K.GVFVQG..GIIISGYQGIQVVVYNSNDKEVIIPQGRKFALILMPLIH.EDLEAW..GETRRTERGNQGFGSTGAY.............
Visn   NKGVFVQG..GIIISGYQGTIQVVIYNSNNKEVVIPQGRKFALILMPLIH.EELEPW..GETRKTERGSQGFGSTGMYW............
Caev   K.GVFTQG..GIIISGYQGQIQVIMYNSNKIAVVIPQGRKFALILMDKKH.GKLEPW..GESRKTERGKQGFGSTGMY.............
Eiav   KQGLLING..GIIIEGYTGEIQVICTNIGKSNIKLIEGQKFALIILQHHS.NSRQPW..DENKISQRGDKGFGSTGVFW...........
Coxb   KHGIVLGNLVGLIISDQQGPLMVSCWNRGKEPYTINPGDRIALVVPILK.AQFAVV..EEFELTERGAGGFGSSGQN...........
Ecol   KHGIVLGNLVGLIISDQQGLMISVWNRGQDSFTIQPGERIALMIFVPVQ.AEFNLV..EDFDATDRGEGGFGHSGRQ...........
Vaccc  K.GIDIGG..GVIIEDRRGNIGVILNNNGKCTFNVNTGDRIALIYQRIYY.PELEEV..QSLDSTNRGDQGFGSTGLR...........
Varv   K.GIDIGG..GVIIEDRRGNIGVILNNGKYTFNVNTGDRIALIYQRIYY.PELKEV..QSLDSTDRGDQGFGSTGLR...........
Human  KHFIDVGA..GVIIEDRRGNVGVVLFNFGKEKFEVKKGDRIALICERIFY.PEIEEA..QALDDTERGSSGFGSTGKN...........
Orfv   KHFIDVGA..GVIIEDRRGNVGVVLFNFGNSDFEVKKGDRIALICERISC.PAVQEV..NCLDNTDRGDSGFGSTGSGACGGRDTAWYIS......
Fadv   KFFIDVGA..GVIIPDRRGNVSVVLFNFSESSFNIRRGDRVALILERIMV.PELSEL..TQLGETDRGAGGFGFTGMGAVDRNQRSVLEWLTPGSR
Calb   KHGISTGA..GVIBADRRGEVKVVLFNHSEKDFEIKEGDRIALVLEQIVN.ADIKEISLEELDNTERGEGGFGSTGKN...........
Scer   KNGIQTGA..GVIIRDYTGEVKVVLFNHSQRDFAIKKGDRVALILEKIVDDAQIVVV..DSLEESARGAGGFGSTGN.............
Lyces  KYSIDVGA..GVIIADRRGPVGVVLFNHSEVDFEVKVPQKIVT.PEVEQV..DDLDSTVRGSGGFGSTGV...............
Ccv    KNFVST....GVIISDERGHIHIMV..SAIADFSVKKNQRIALLVVTPCLTQSEVVPY..ETLERTRRGTGGFGSSGQ............
          ---------         ---------             ---------
          Motif 3           Motif 4               Motif 5
```

Figure 4.1 Alignment of class I sequences

Over 100 dUTPase homologues are produced from a blastp search of the Swiss-Prot database probed with the *E.coli* sequence. A standard alignment using a selection of 19 class I sequences is shown. Motif regions are highlighted in yellow. Green highlights refer to those residues which make contact directly or indirectly with the substrate in the *E.coli*-dUDP structure (see Section 4.2.5). The following sequences were used: Mpmv (Mason-Pfizer monkey virus), Srv1 (Simian type D retrovirus), Jsrv (Sheep pulmonary adenomatosis virus, Lentivirus), Mmtv (Mouse mammary tumor virus), Fiv (Feline immunodeficiency virus), Omvs (Ovine lentivirus), Visn (Visna lentivirus), Caev (Caprine arthritis-encephalitis virus), Eiav (Equine infectious anaemia virus), Coxb (*Coxiella burnetii,* Eubacteria), Ecol (*E.coli*), Vaccc (Vaccinia virus), Varv (Variola virus), Human (Human), Orfv (Orf virus), Fadv (Avian/Fowl adenovirus CELO), Calb (*Candida albicans*), Scer (*Saccharomyces cerevisiae*), Lyces (*Lycopersicon esculentum,* Tomato), Ccv (Channel catfish virus).

4.2.3   Comparison of trimeric class I structures

There are presently structures available for four class I dUTPases: *E.coli* (1DUP), FIV (1DUT), EIAV (1DUC) and human (not yet in Brookhaven PDB database). The main features of the *E.coli* dUTPase structure are described in the introduction (Section 1.4.7). This section deals with the comparison of overall structures between these four class I dUTPases.

(a) *E.coli*

(b) FIV

(c) human

(d) EIAV
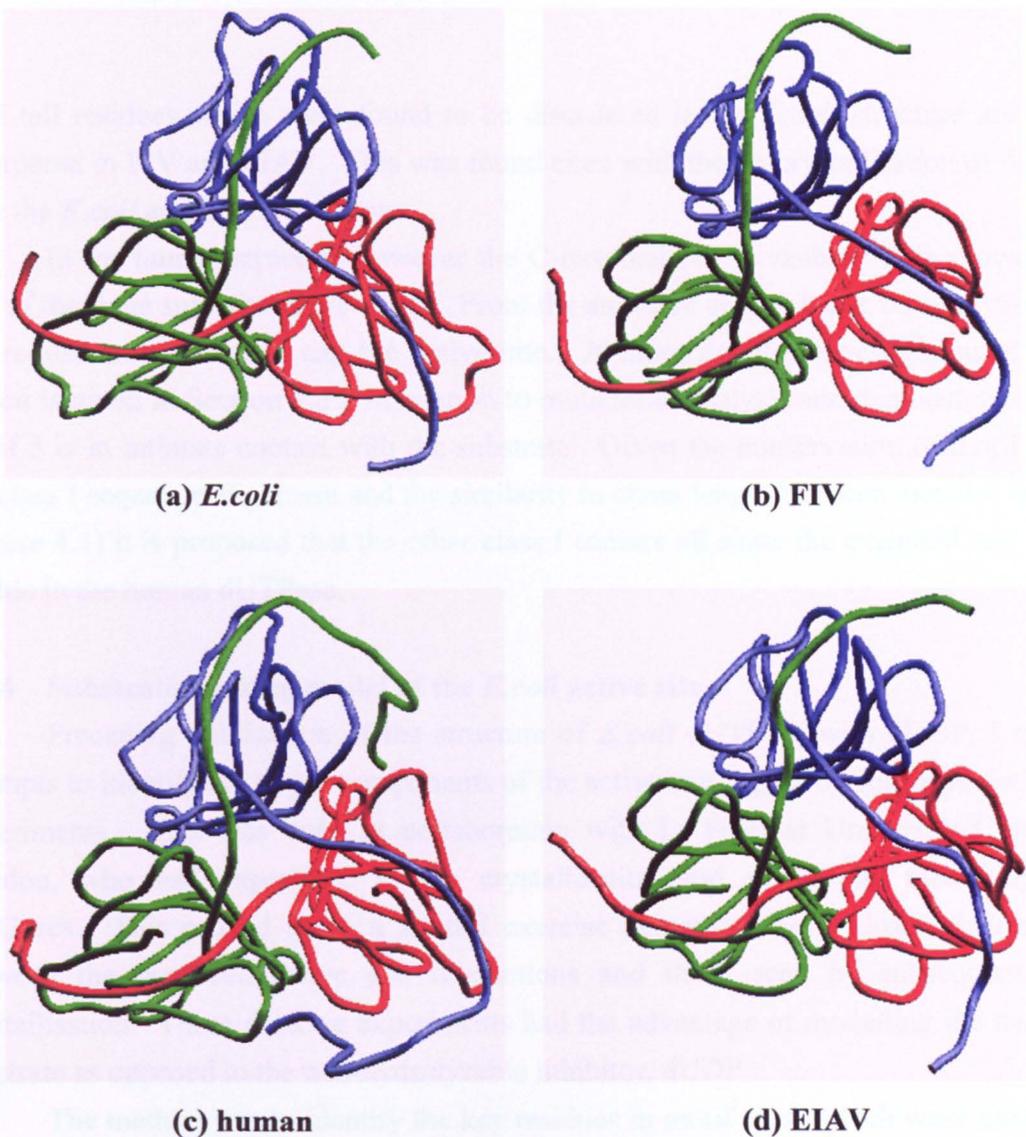
Figure 4.2 Comparison of class I trimeric structures
The four class I trimers are rendered as ribbons using the Swiss-PDB Viewer. Each trimer is orientated approximately down the three fold axis of symmetry. In the human structure the extended tail is shown on two subunits (blue and green) and removed from the other (red) for comparison to the other trimers.

From Figure 4.2 it is clear that the structure detailed for *E.coli* dUTPase in Section 1.4.7 is repeated in other distantly related class I species. Each subunit structure shows an almost identical α-carbon backbone trace allowing the generation of almost identical trimers. Not only is the globular region conserved but also the visible region of the extended C-terminal arm.

Figure 4.2 shows these four structures positioned approximately down the 3-fold axis of symmetry. Each subunit is represented by a different colour. It is clear that in terms of overall quaternary structure these dUTPases are highly homologous. This is an important observation for the modelling of a class II structure. Given that the only areas of primary sequence conservation lie within the motif regions it suggests that inter-motif chain length is an important factor in the overall structure. Interestingly the

final tail residues which were found to be disordered in the *E.coli* structure are also disordered in FIV and EIAV. This was found even with the co-crystallisation of dUDP with the *E.coli* and EIAV structures.

In the human structure however the C-terminal tail is visible. It is shown for two of the three subunits in the figure. From the structure shown it can be seen that the tail region curls round to cap the active site. A more detailed investigation of this region is given in Section 5.8.2 in relation to mutational analysis and demonstrates that motif 5 is in intimate contact with the substrate. Given the conservation of motif 5 in the class I sequence alignment and the similarity in chain length between motifs 4 and 5 (Figure 4.1) it is proposed that the other class I trimers all share the extended arm fold visible in the human dUTPase.

4.2.4   Substrate docking model of the *E.coli* active site

Preceding publication of the structure of *E.coli* dUTPase with dUDP, I made attempts to identify the major components of the active site region by substrate docking experiments. This was done in collaboration with L. Pearl at University College, London, who had experience in the crystallisation and active site modelling of UDGases. This proved to be a fruitful exercise demonstrating a close relationship between the proposed active site interactions and those seen by subsequent co-crystallisation. These docking experiments had the advantage of modelling the natural substrate as opposed to the non-hydrolyzable inhibitor, dUDP.

The method was to identify the key residues in motif areas which were likely to interact with dUTP. Residues were graded on importance by assessing their conservation between the class I species, their relative position in the *E.coli* structure and their properties such as hydrophobicity and charge. Figure 4.3 shows the proposed position of dUTP in a solvent surface and wireframe representation of the *E.coli* active site. The key residues are highlighted with their respective sequence numbers and the wireframe is colour coded in relation to charge. Blue represents positive charge, red is negative and yellow is neutral.

The observations on this model are discussed with reference to the following section (4.2.5) which details the crystal structure of *E.coli* dUTPase with dUDP. One of the major difficulties with the modelling exercise was the inherent flexibility of the natural dUTP substrate molecule. It was realised that the various torsion angles of the dUTP molecule could be significantly altered if the active site produced an energetically favourable environment.
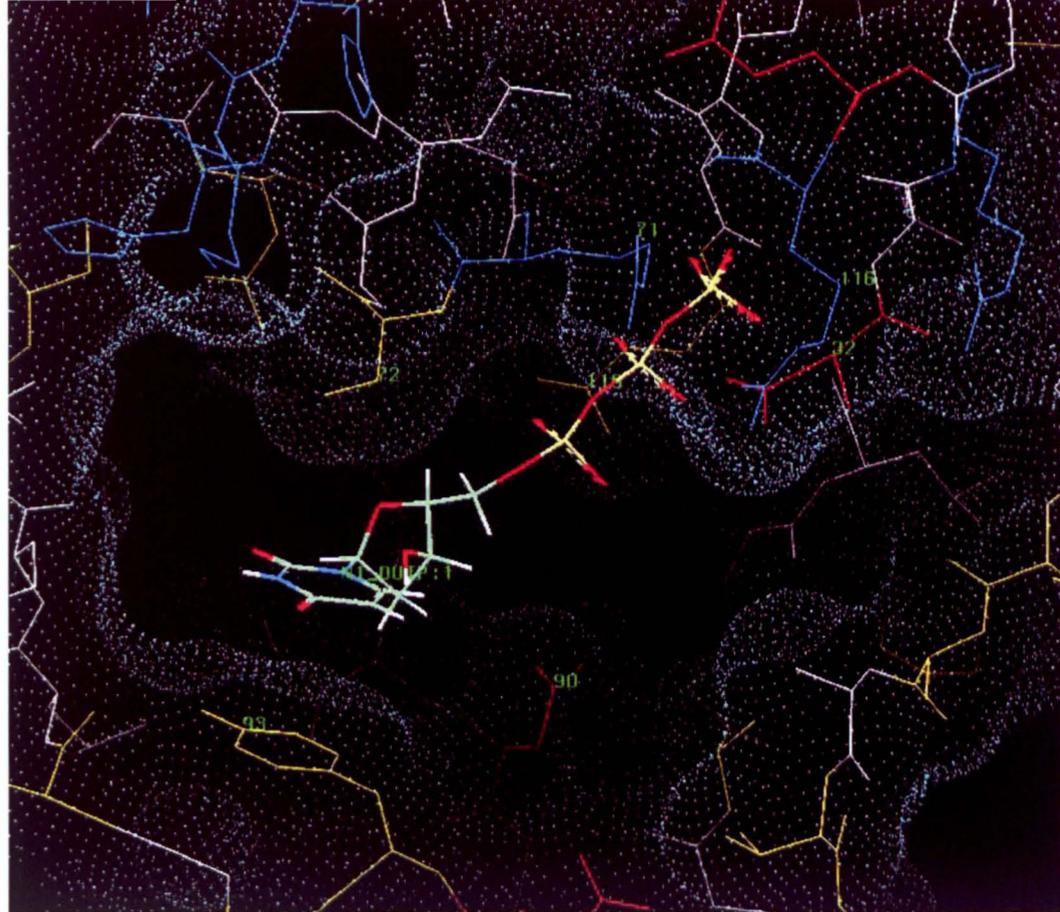
(a)
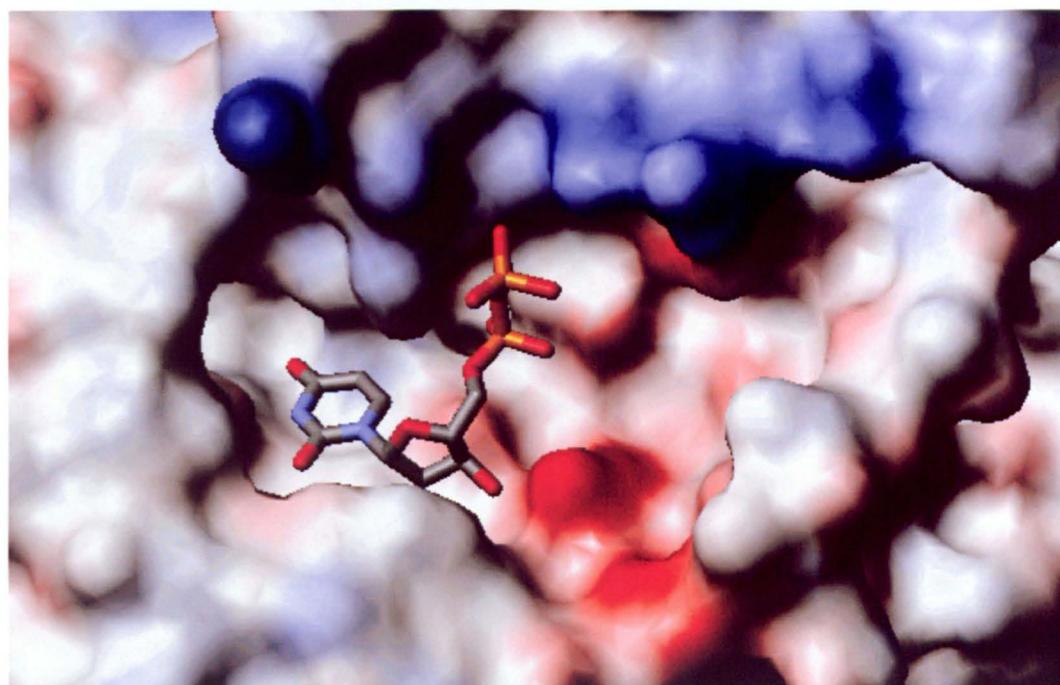
(b)

Figure 4.3 Active site docking model of dUTP and *E.coli* dUTPase

Figure (a) shows the wireframe and dot surface representation of *E.coli* dUTPase active site which was created using the molecular graphics program 'O'. The dUTP substrate was positioned in the active site cavity with respect to residues of stabilising charge (numbered). The uracil ring is shown stacking on the highly conserved residue tyrosine 93. Figure (b) shows the actual position of dUDP in the *E.coli* co-crystal. A solvent surface was created using the WebLab molecular viewer. Residues are coloured by charge: positive in blue and negative in red.

Figure 4.3 shows the docking model prediction of dUTP in the *E.coli* active site (a) compared to the actual structural data from bound dUDP in the *E.coli* dUTPase co-crystal (b). Both diagrams show the active site cavity in the same orientation to allow direct comparison. Charged residues shown as coloured wireframe representations in (a) can be seen as correspondingly coloured molecular surfaces in (b). The cavity area which can be seen in both diagrams was proposed to be the substrate binding area based solely on the condensation of the conserved motif regions in this area. The subsequent co-crystal of *E.coli* dUTPase with bound dUDP demonstrated that this was predicted correctly. The overall shape of this cavity constrained the number of possible substrate orientations in the docking model. The deepest area of the cavity (on the left of the diagrams) was predicted to accommodate the nucleoside portion of the substrate which generated an overall predicted orientation close to the real structure.

The substrate was aligned within the active site cavity with the phosphate groups close to areas of neutralising charge. The ring of Tyr 93 was proposed to generate a stable stacking conformation with the pyrimidine ring of dUTP and was positioned accordingly. It was later found in the crystal structure that Tyr 93 does indeed form a parallel stacking structure but with the sugar moiety of dUDP and not with the pyrimidine as proposed by the docking model. This can be more clearly visualised in Figure 4.5 in the following section. Apart from this error the overall position of the substrate in the proposed active site was reasonable. The sugar ring and α-phosphate was correctly located close to negatively charged Asp 90. This residue interacts directly with the hydroxyl group on the sugar and through a water molecule, indirectly with the α-phosphate. The α-phosphate was also correctly positioned close to Ser 72 (on the upper side of dUTP in the model). This residue was shown to interact with the α-phosphate through a second water molecule in the crystal structure.

In the crystal structure the β-phosphate is highly mobile and the stabilising interactions are not clear. In the proposed model the β- and γ-phosphates are positioned towards the conserved residues Arg71 and Arg116 which, being positively charged, would have the ability to neutralise the phosphate groups directly without the interaction on further water molecules. This area of positive charge can be seen in both diagrams in Figure 4.3 by comparing the position of the blue wireframe residues Arg71 and Arg116 in (a) and the corresponding blue surfaces in (b). It is likely that this area of the model will not be resolved until a non-hydrolyzable substrate containing the α-, β-, and γ-phosphates is co-crystallised with the enzyme. This type of docking analysis is useful at providing a constrained set of possible residue-substrate interactions which can then be tested by mutagenesis studies.

4.2.5   Interactions of the *E.coli* active site with bound dUDP

Two properties of the active site area were highlighted by the docking model: the accommodation of the pyrimidine and sugar rings in a visible cavity and the accommodation of the phosphate tail by neutralising charge. There are however additional attributes necessary for functionality. For modelling purposes it is useful to identify the specific residues involved in substrate specificity and those involved in catalysis. This information became available with the co-crystallisation of *E.coli* dUTPase with dUDP. The main interactions are shown in Figure 4.4 and 4.5.

Seven of the residues in Figure 4.4 appear to be of key functional importance to the enzyme. Four residues (Ile 89, Asp 90, Tyr 93, Gln 119) make direct contact with dUDP in the active site and three residues (Gly 30, Ser 72, Met 98) make indirect contact mediated through water molecules. These residues are almost completely conserved throughout the class I group (highlighted in green in alignment Figure 4.1) with the exception of Met 98. Since this residue hydrogen bonds to the uracil ring of dUDP through main chain atoms it is likely that the conservation of its position is of more consequence than conservation of its side chain. This main chain interaction can be seen in the motif 3 loop shown in Figure 4.5. The water molecule responsible for nucleophilic attack of the α-phosphate has not been determined but it is likely that the conserved residues (Ser 72, Asp 90, Gln 119) play a role by contributing to the hydrogen bonding pattern around the phosphorus atom.
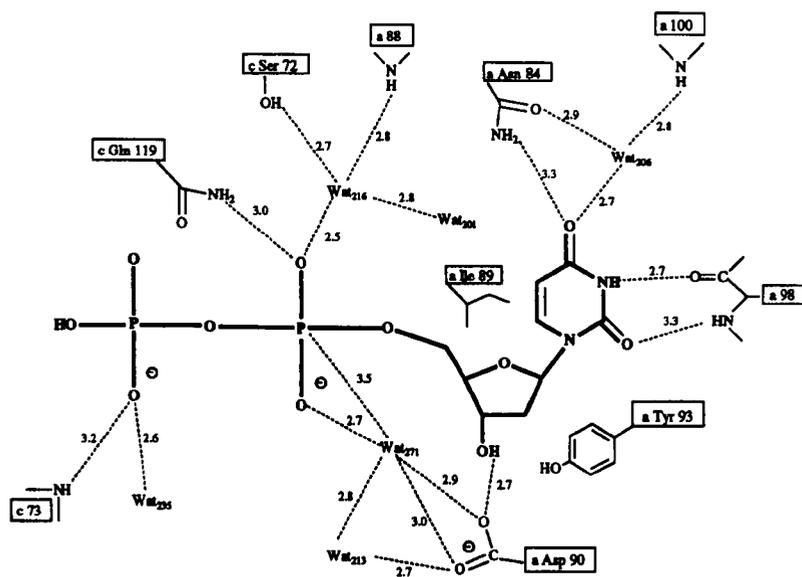
Figure 4.4 Interactions of the *E.coli* active site with bound dUDP.
A diagram of the main interactions of dUDP with the active site in the *E.coli* co-crystal structure is shown (from Larsson *et al.*, 1996). Hydrogen bonds are shown as dashed lines and distances are given (in Å). Residues from different subunits are marked as 'a' or 'c' respectively.

E.coli dUTPase is highly discriminatory with respect to base, sugar and phosphate and does not significantly hydrolyse dCTP, dTTP, UTP or any of the diphosphate relatives. This specificity is accomplished almost entirely by the β-hairpin formed by the highly conserved motif 3 (Figure 4.5). In terms of base specificity it is likely that the main chain of Met 98 provides discrimination between cytosine and uracil. Discrimination against thymine is proposed to be steric given the lack of space available to accommodate a C5 methyl group. Specificity for the sugar is achieved by the positioning of the pentose ring between the conserved residues Ile 89 and Tyr 93 thus removing space to accommodate a 2' OH. The mechanism for triphosphate specificity is not clear at present. The key residues identified as interacting with dUDP are highly conserved within the class I group and are highlighted in the sequence alignment in Figure 4.1.
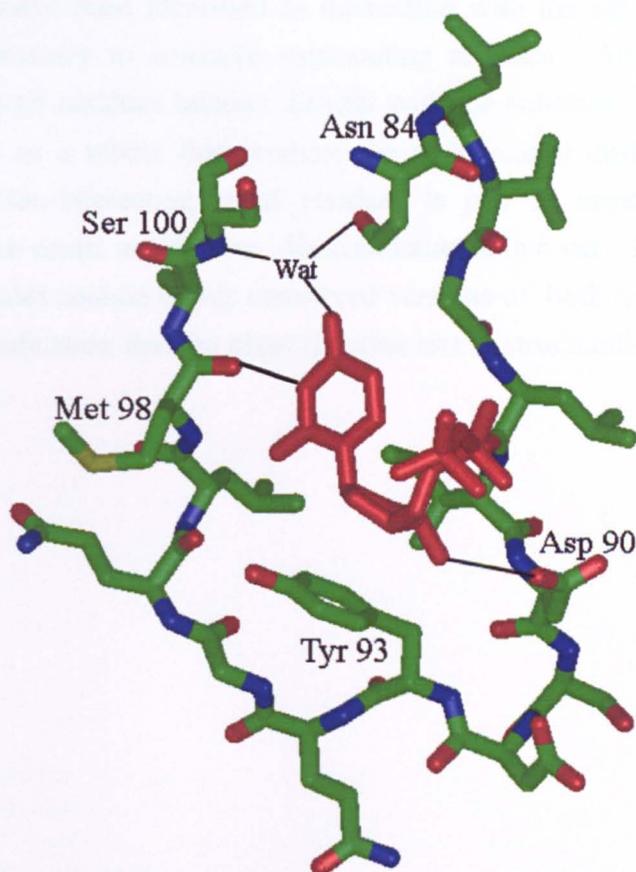
Figure 4.5 Interactions of *E.coli* β-hairpin loop with bound dUDP

The structure of the *E.coli* β-hairpin loop with bound dUDP is shown. The β-hairpin loop structure is formed by the conserved motif 3 region. The main residues involved in the specific recognition of the substrate are indicated in addition to the structurally bound water molecule (Wat). Hydrogen bonds are shown as black lines. Uracil recognition is accomplished by hydrogen bonding to the main chain (Met-98) and the bound water molecule through the side chain of Asn-84 and backbone of Ser-100. Discrimination against ribose is provided by Tyr-93 which faces the deoxyribose moiety. The 3'-hydroxyl group of the sugar donates a proton to the side chain of Asp-90. Graphic representation was generated using the Insight II molecular viewer.

4.2.6 Comparison of class I active sites

Since data became available for the EIAV, FIV and human dUTPases it was of interest to compare the active site regions to identify areas of homology. The condensation of the 4 motif regions visible in Figure 4.6 is remarkably consistent between species. Motif 3 forms the main interaction platform and, as seen in the previous section, forms the β-hairpin loop allowing high substrate specificity. All four active sites shown expose this motif 3 region (red) and utilise motifs 1, 2 and 4 to create the other 3 sides of the active site cavity. The interacting residues of the *E.coli* dUTPase with dUDP depicted in diagram 4.5 represent only one third of the total residues contained in motifs 1-4. Although there are few interactions from motifs 1, 2 and 4, it is clear from these diagrams that they have a role in creating a specific active site environment.

This is a key observation for interpretation of the class II model. Although specific residues have been identified as interacting with the substrate, for this to be achieved it is necessary to conserve surrounding residues. Although only a small minority of the motif residues interact directly with the substrate (see Figure 4.1) it is the motif regions as a whole that produce the architectural design. It appears that conservation of non-interacting motif residues is just as important as interacting residues in order to create an efficient, discriminating active site. It will be shown that the class II molecules contain highly conserved versions of both types of motif residue. Again this is an indication that the class II active site is structurally homologous to the class I active site.

(a) *E.coli*

(b) FIV

(c) Human

(d) EIAV

Species	Motifs				
	1	2	3	4	5
E.coli	29-33	71-77	87-100	114-121	141-150
FIV	16-20	55-61	68-81	95-102	122-131
Human	23-27	62-68	76-89	103-110	130-139
EIAV	17-21	56-61	69-82	96-103	123-132

Figure 4.6 Comparison of class I proposed active site regions

For each molecular species an enlarged view of a single active site is shown. The following colour coding has been used for the motif regions composing the active site area: 1-blue, 2-purple, 3-red and 4-green. The spacefilling representations were created using RasMol. Note that the C-terminal tail has been removed from the human structure to allow an unimpeded view of the active site. The table indicates the residue numbers which constitute the motifs in each species.

4.2.7 Discussion

In summary it is clear that in terms of overall structure and active site conformation, the class I dUTPases are highly similar. The motif regions condense to form an active site cavity with conserved residues providing key functional roles. The conservation of inter-motif distances, observed in the class I alignments, is likely to help define the position of the 5 motifs. Structural comparison of the four available structures highlighted that the inter-motif regions can generate similar conformations even with low sequence conservation. It will be shown in the forthcoming sections that these inter-motif regions possess more homology than is first evident with primary sequence analysis. Although there is only a relatively small percentage of the molecule that actually achieves substrate binding and catalysis, the active site conformation is dependant on the secondary and tertiary structure of the entire molecule. This is an important observation for the class II modelling since there is almost no primary sequence homology out with the motif regions.

Modelling of the *E.coli* active site helped to identify the key residues now apparent in the dUDP co-crystal structure. Comparison of active sites from diverse class I species shows a remarkable conservation of structural elements. Homologous cavities are visible in all the structures with key residues being similarly positioned. The attribution of function and high conservation of these residues provide an ideal probe for investigating the class II enzymes.

4.3　Class I dUTPases as a basis for modelling class II dUTPases

4.3.1　Introduction

In terms of motif arrangement, gene size and general structural arrangement, the class I dUTPases are highly homologous. The class II dUTPases however are a much more heterogeneous group. Members share a common motif arrangement but vary considerably in motif spacing and gene size. It is predicted that this group share a similar structural arrangement although these genetic variations will provide some variations and will be considered with respect to the class II model. This section tackles the class II dUTPases and how to best utilise the data from the class I enzymes for comparison.

4.3.2　Comparison of class II primary sequences

Alignment of the class II sequences was more difficult than the class I relatives due to the heterogeneity outside the motif regions. The motifs themselves are highly conserved throughout the class II group and provide a starting point for sequence alignments. In order to achieve a non biased set of alignments different programs were utilised.

Initially class II sequence alignments were performed using GCG Pileup, a simplification of the method of Feng and Doolittle (1987). This is a very quick and reasonably efficient alignment program and is commonly used. It performs well on sequences with reasonable conservation but is less efficient at aligning multiple sequences with significant variation. Variability in sequence length for example, can be potentially difficult for the program to judge, especially if it occurs in areas of low sequence conservation. The program itself aligns the two most conserved sequences first and subsequently adds additional sequences to the lineup one by one. This means that if there are any significant mistakes in the initial or intermediate alignments they cannot be corrected later as further sequences are added.

In the context of the class II alignments this was thought to pose a potential problem. As will be shown, the C-terminal halves of the class II enzymes are reasonably conserved relative to motif sequence and inter-motif distance. Conversely the N-terminal halves are significantly divergent within class II group. Since this half contains only one motif region (motif 3) this variability is judged by measuring two direct variables: the distance from the N-terminus to the beginning of motif 3 and secondly, the distance between motif 3 and motif 1 (the first motif in the C-terminal half). Figure 4.7 shows such a measurement.
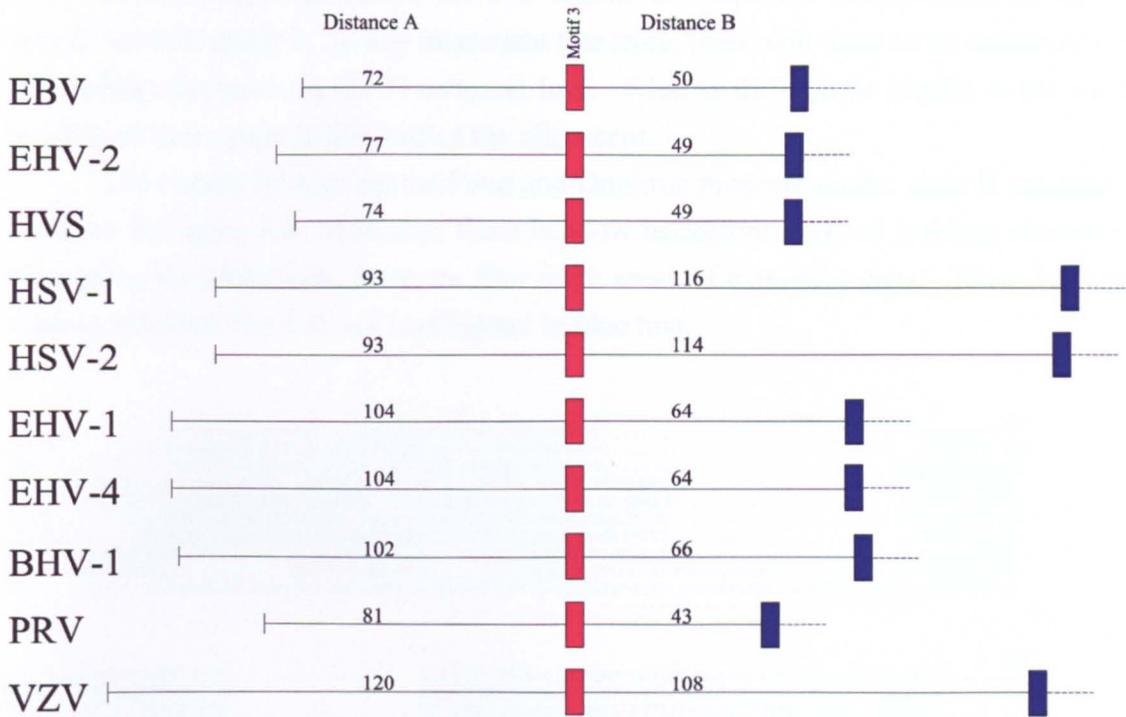
Figure 4.7 Determination of class II variability by inter-motif distances
Two sequence lengths have been calculated for each species: "A", the distance between motif 3 and the N-terminus and "B", the distance between motif 3 and motif 1. The N-terminus is represented as a vertical line, motif 3 by a red box and motif 1 as a blue box. The lines are drawn to scale to allow a graphical as well as numerical visualisation. Sequences are arranged in the same order as Figure 4.8 for comparison.

The first measurement "A", from the N-termini to motif 3, shows a general variability between dissimilar species. Closely related pairs such as HSV-1/HSV-2 and EHV-1/EHV-4 have identical distances. The largest "A" distance is that of VZV while EBV/EHV-2/HVS have the smallest. This overall pattern is mimicked in the second distance "B" between motif 3 and motif 1. Similar species have closely related "B" distances while more distant species show wide variability. HSV-1 and HSV-2 have the longest "B" distance closely followed by VZV.

The key point to note is that distance "A" is not predictive of distance "B". For example, EHV-1 and EHV-2 have a long "A" and a short "B" while EBV and HVS have a short "A" and a short "B". Furthermore, these distances are not conserved between species. For example, HSV-1 and EHV-1 differ by only 11 residues in "A" but differ by 52 residues in "B". A very simple conclusion can be drawn from this data, namely that there is variability with respect to insertions in both regions "A" and "B". However, in terms of sequence alignment this is not a simple problem. As can be seen

from the following alignments, there is almost no sequence conservation in the N-termini outwith motif 3. In any alignment therefore, there will have to be extensive use of padding characters in the N-terminal half. What is difficult to predict is the exact position of these gaps in this half of the alignment.

The results from using the Feng and Doolittle program on the class II sequences is shown in Figure 4.8. Although there is a low background use of padding characters throughout the alignment, there are four main areas of extensive gaps. These variable regions, labelled Var 1-4, are highlighted in blue text.

```
        1                                                                                          100
EBV     ...........MEACPHIRYAFQNDKLLL.QQASVGRLTLVNKTTILLRPMKTTT.............................VDLGLYARPPE.
EHV-2   .........MTLRKQRQEIYYEFTSSSFEITSPPESSKITLTNLYPILVRPYVPAV...........................IPLGIRIIYGNK
HVS     ...........MPYKVPEIYYRFEPQTFYITSPARASNLQLINHNNILVKAGQVTI...........................VSTGI.IFFKET
HSV-1   .MSQWGSGAILVQPDSLGRGYDGDWHTAVATRGGGVVQLNLVNRRAVAFMPKVSGDSGWAVG....................RVSLDLRMAMPA.
HSV-2   ...MSQWGPRAILVQTDSTNRNADGDWQAAVAIRGGGVVQLNMVNKRAVDFTPAECGDSEWAVG..................RVSLGLRMAMPR.
EHV-1   .......MASVTNLVDSIVVVECGERWRARAEAAGRLLVLINNHTVELSGEHGSAGEFYSVL....................TDVGVRVACSS.
EHV-4   .......MASATNLADNIVVVECSNGWEARAEADGRLLVLINNHTVELSAGLGSAGEFYSVL....................TDVGVRVACSS.
BHV-1   MANSAAATTATMSGDRGILVVELNAEAAPWRLESCCEPDSLALWGPIAPAAKRDETAPSGSLLYSR................LINLNMKAAAPG.
PRV     MEESAGATSA.............QSAATSVSESPAE.......ETILVCASEPVTVDGGRLLVCRSPGPEGF..........YKVPLGLKVALPT.
VZV     ...MNEAVIDPILETAVNTGDMFCSQTIPNRCLKDTILIEVQPECADTLQCVLDDKVSRHQPLLLRNHKKLELPSEKSVTRGGFYMQQLELLVKSAPPN.
                                                                        -----------------------------
                                                                                  VAR 1

        101                                                                                        200
EBV     GHGLMLWGSTSRPVTSH...................VGIIDPGYTGELRLILQNQRRYNSTLRPSELKIHLAAFRYATPQMEEDKGPINH.........
EHV-2   GQGFILAGSSQKKVFCH...................TGLIDPGYRGEIKLIVLNTTKYNVTLFAGELRVSLFSFFFSTPII.YDYDLLNR.........
HVS     SFAFILYGKSAKSIFCH...................TGLIDPGFQGELKLIVLNKTEDDITLFENDLRVSVTAFVYGVPKL.HDYSDLCP.........
HSV-1   DFCAIIHAPALASPGHH...................VILGLIDSGYRGTVMAVVVAPKR.TREFAPGTLRVDVTFLDILATPPALTEPISLRQF...PQLA
HSV-2   DFCAIIHAPAVSGPGPH...................VMLGLIDSGYRGTVLAVVVAPNG.TRGFAPGALRVDVTFLDIRATPPTLTEPSSLHRF...PQLA
EHV-1   GYAIVLTQISGLLPVEPEPGNFSNVTFPENSAKYYTAYGIVDSGYRGVVKAVQFAPGI.NTSVPPGQMSLGLVLVKLARKSIHVTSIGSTRDG...RTSE
EHV-4   GYAIVLAQISGLPHVGREPGNFSNITFTGNLANYYTAYGIVDSGYRGVVKAVQFANGV.NTVVPPGCMSLGLVLVKLSTETINVTNINLTENG...RSPR
BHV-1   GYAIIMSQMRSGDTHMPRPPAVAV.............GIVDSGYSGILRAIVWAPES.AAAAPPAGL...ALRLTLARLTTTLPRLIAVDDD...ANAG
PRV     GYAMLVAQRGGGRT.........TN............GIVDAGFRGEVQAIV......APGRPRAQFYCTPLRL.................A
VZV     EYALLLIQCKDTALADED...............NFFVANGVIDAGYRGVISALLYYRPG.VTVILPGHLTIYLFPVKLRQSRLLPKNVLKHLDPIFKSIQV
                             --------------------  ---------                                        ----------
                                    VAR 2         MOTIF 3                                            VAR 3...

        201                                                                                        300
EBV     ...........................................................PQYPGDVGLDVSLPKDLALFPHQTVSVTLTV..P.PPSI
EHV-2   ...........................................................PQYSDDAGYDLYLQEDLMLFPQASTTVTIDS..RVPTTT
HVS     ...........................................................PRYSKDAGFDLYLPTDVTVKPRVPNRYSVNI..CCPAQL
HSV-1   PPPPTGAGIREDPWLEGALGAPSVTTALPARRRGRSLVYAGELTPVQTEHGDGVREAIAFLPKREEDAGFDIVVRRPVTVPANGTTVVQPS.LRMLHADA
HSV-2   PSPL..AGLREDPWLDGALATAGGAVALPARRRGGSLVYAGELTQVVTTEHGDCVHEAPAFLPKREEDAGFDILIHRAVTVPANGATVIQPS.LRVLRAAD
EHV-1   ANL........FYD.......................................YFAPKRVEDAGYDISAPEDATIDPDESHFVDLP.I.VFANSN
EHV-4   VNI........FYD.......................................YFAPKRDEDAGYDISAQTNATIEPDESYFVELP.I.VFSSSN
BHV-1   TEAGVEVPFFA.........................................TFAPKRDEDAGYDIAMPYTAVLAPGENLHVRLP.V.AYAADA
PRV     PGIATDVPFFE.........................................VFAPKRDEDAGYDIPCPRELVLPPGGAETVTLP.V.HRTDGR
VVZ     QPLSNSPSNYEKPVIPEFADISTVQQGQPLHRDSAEYHIDVPLT...........YKHIINPKRQEDAGYDICVPYNLYLKRNEFIKIVLPIIRDWDLQH
        --------------------------------------------                     -----
                       VAR 3 continued                                  MOTIF 1

        301                                                                                        400
EBV     PHHRPTIFGRSGLAMQGILVKPCRWRRGGV.DVSLTNFSDQTVFLNKYRRFCQLVYLHKHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEVPSLAM....
EHV-2   KFFKPVVFGRSGLATRGVVVDVVKWTHSPL.TLKIYNFTDNTLRYSAGTRICQVVFVHRRHFPSKLKHFFTYINLNSKTSFYWANVSFVDCQNDAYRSLV
HVS     KSYKPVLFGRSGLAAKGLTIKVSRW.QNQL.QIIFYNYTKSQITYTARTRIAQVVFMHKKHLPTTLTRLKPTMHLSENIKYSWARVSFQDIKTFPVQDEK
HSV-1   GPAACYVLGRSSLNARGLLVVPTRWLPGHVCAFVVYNLTGVPVTLEAGKVAQLLVAGADALPWI....PPDNFHGTKALRNYPRGVPDSTAEPRNPPLL
HSV-2   GPEACYVLGRSSLNARGLLVMPTRWPSGHACAFVVCNLTGVPVTLQAGSKVAQLLVAGTHALPWI....PPDNIHEDGAFRAYPRGVPDATATPRDPPIL
EHV-1   PAVTPCIFGRSSMNRRGLILVLPTRWVAGRTCCFFILNVNKYPVSITKGQRVAQLLLTEDIDDALI....PP.TVNYDNPFPTYSPS..ESTKAPQSPVLW
EHV-4   PAVTPCIFGRSSMNRRGLIVLPTRWVTGRTCCFFILNINKYPVYITKGQRVAQLVLTEDIDEALI....PT.NVNYNTPFPTYSPT..GAVKHNPTPILW
BHV-1   HAAAPYVFGRSSCNLRGLIVLPTAWPPGEPCRFVLRNVTQEPLVAAAGQRVAQLLLLARRLE.WI....PS.GLNDREPFPTSPRA..APPAPGAPRLRW
PRV     HWA..YVFGRSSLNLRGIVVFPTPWESG.PCRFRIQNRGAHPVTLESGQRVAQLVLTREPLG.WI....T....GRSPFPATPRA..PMQHRPAWLFAR
VZV     PSINAYIFGRSSKSRSGIIVCPTAWPAGEHCKFYVYNLTGDDIRIKTGDRLAQVLLIDHNTQIHL.....KHNVLSNIAFPYAIRGKCGIPG.....VQW
                --------                                  ---------       -------
                MOTIF 2                                   MOTIF 4         VAR 4

        401                      436
EBV     ...GDSGLSEALEGRQGRGFGSSGQ..........
EHV-2   TLPCQEDTDRG..YRGDSGFGSSGMR.........
HVS     LYSSSKDTSDSQMSRGDAGLGSSGLM.........
HSV-1   VFTNEFDAEAPPSERGTGGFGSTGI...........
HSV-2   VFTNEFDADAPPSKRGAGGFGSTGI...........
EHV-1   KFTTDFDREAPSSLRADGGFGSTGL...........
EHV-4   KFTEAFDHDAPSSARSEGGFGSTGL...........
BHV-1   RRVADLAAAVPPSARGPRGFGSTGL...........
PRV     DFVA......PSSARGARGFGSTGL...........
VZV     YFTKTLDLIATPSERGTRGFGSTDKETNDVDFLLKH
                          ---------
                          MOTIF 5
```

Figure 4.8 Class II sequence alignment by Feng and Doolittle method
This alignment was generated by GCG Pileup using the method of Feng and Doolittle with standard default values (gap weight: 3.0, gap length weight: 0.1). Motif regions are highlighted in yellow. Variable regions are shown as blue text.

To allow a more objective approach to the class II alignments a different program was used to process identical input data. A multitude of alignment programs are available but many of them utilise similar progressive methods to that of Feng and Doolittle (FD). The program SAGA (sequence alignment by genetic algorithm) was chosen since it employed a novel approach to sequence alignment.

The program works on an evolutionary basis, generating a large population of alignments and gradually improving them by measuring their quality. The calibre of each alignment is measured by an objective function which is generated by assigning a cost for each column of aligned residues and a cost for each gap. Using this objective function it is then possible to apply a selection process to the population of alignment solutions. Parental alignments which score highly generate more children than alignments of poor quality. The overall size of the population is kept constant creating a selection pressure for better alignments. Over many generations the regions of high quality are conserved until the alignment can no longer be improved. The results of such an alignment is shown in Figure 4.9.

The alignment generated by SAGA has been highlighted in the same way as the original alignment of FD to allow direct comparison. In overall terms the alignments were generally similar despite using two different methods. This was encouraging and gave some assurance that the alignments were close to optimal.

The C-terminal half of the alignments, from motif 1 to motif 5, was aligned efficiently by both methods. Although the inter-motif regions share only scarce sequence conservation, their lengths are highly conserved. Once motifs 1, 2, 4 and 5 are aligned, there are few gapped regions necessary to generate a reasonable alignment in the C-terminal half.

As was shown graphically in Figure 4.7, there is considerable variation in sequence length in the N-terminal half of the alignments. Both programs readily recognised motif 3 but had difficulty with the regions to either side. There is very little sequence homology in these regions which is complicated by the fact that some individual species have substantial deletions and insertions. The two methods used generated similar alignments for the N-terminal half in terms of overall structure. There were however distinct differences.

The main difference was the splitting of variable region 1 into two parts in the SAGA alignment. This is interesting in terms of how these two methods interpret overall conservation. The key feature of this region is that both programs are obviously constrained by the excessive length of the VZV sequence compared to the others. In both cases the VZV sequence is run almost straight through this region with very few gaps. In the FD alignment it can be seen that additional gaps were added to each additional species from top to bottom in order to accommodate the VZV sequence. The

SAGA program has allowed refinement of this region providing a greater number of paired alignments. This can be measured in terms of the quantity of gaps inserted within the bounds of variable region 1. The FD program created 221 gaps compared to 162 from SAGA. Even with these differences, the overall consensus for this region is reasonably high between the two alignments.

```
         1                                                                                          100
EHV-2  ................MTLRKQRQEIYYEFTSSSFEITSPPESS...KITLTN....LYPILVRPYVPA................VIPLGIRIIYGNKGQ
HVS    ................MPYKVPEIYYRFEPQTFYITSPARAS...NLQLIN...HNNILVKAGQVT............IVSTGIIFP.KETSF
EBV    ............MEACPHIRYAFQNDKLLLQQASVGR...LTLVN...KTTILLRPMKTT............TVDLGLYAR.PPEGH
EHV-1  ........MASVTNLVDSIVVVECGERWRARAEAAGR...LLVLIN...NHTVELSGEHGSAGEFYSV.........LTDVGVGRVA.CSSGY
EHV-4  ........MASATNLADNIVVVECSNGWEARAEAADGR...LLVLIN...NHTVELSAGLGSAGEFYSV.........LTDVGVRVA.CSSGY
BHV-1  ........MANSAAATTATMSGDRGILVVELNAEAA...PWRLES.CCEPDSLALWGPIAPAAKRDETAPSGSLLYSRLINLNMKAA.APGGY
HSV-1  ......MSQWGSGAILVQPDSLGRGYDGDWHTAVATRGGGVVQLNLVN...RRAVAFMPKVSGDSGWAVG.....RVSLDLRMA.MPADF
HSV-2  ......MSQWGPRAILVQTDSTNRNADGDWQAAVAIRGGGVVQLNMVN...KRAVDFTPAECGDSEWAVG.........RVSLGLRMA.MPRDF
PRV    ......MEESAGATSAQSAATSVSESPAEETILVCASEP....VTVDG....GRLLVCRSPGPEGFY...........KVPLGLKVA.LPTGY
VZV    MNEAVIDPILETAVNTGDMFCSQTIPNRCLKDTILIEVQPECADTLQCVLDDKVSRHQPLLLRNHKKLELPSEKSVTRGGFYMQ.QLELLVKSA.PPNEY
                                                            --------------        -----------------
                                                               VAR 1A                   VAR 1B

       101                                                                                          200
EHV-2  GFILAGSSQKKV.................FCHTGLIDPGYRGEIKLIVLNTTKYNVTLFAGELRVSLFSFFFSTPIIY................
HVS    AFILYGKSAKSI.................FCHTGLIDPGFQGELKLIVLNKTEDDITLFENDLRVSVTAFVYGVPKLH................
EBV    GLMLWGSTSRPV.................TSHVGIIDPGYTGELRLILQNQRRYNSTLRPSELKIHLAAFRYATPQME................
EHV-1  AIVLTQISGLLPVEPEPGNFSNVTFPENSAKYYTAYGIVDSGYRGVVKAVQFANGV.NTSVPPGQMSLGLVLVKLARKSIH............
EHV-4  AIVLAQISGLPHVGREPGNFSNITFTGNLANYYTAYGIVDSGYRGVVKAVQFANGV.NTVVPPGCMSLGLVLVKLSTETIN............
BHV-1  AIIMSQMRSGDTHMPRPP.............AVAVGIVDSGYSGILRAIVWAPES.AAAAPPAGLALRLTLARLTTTLPR............
HSV-1  CAIIHAPALASPGH.................HVILGLIDSGYRGTVMAVVVAPKR.TREFAPGTLRVDVTFLDILATPPALTEPISLRQFPQLAPPPPT
HSV-2  CAIIHAPAVSGPGP.................HVMLGLVDSGYRGTVLAVVVAPNG.TRGFAPGALRVDVTFLDIRATPPTLTEPSSLHRFPQLAP..SP
PRV    AMLVAQRGGG...................RTTNGIVDAGFRGEVQAIVAPGRP.RAQFYCTPLRLA......PGIATD................
VZV    ALLLIQCKDTALADEDNF.............FVANGVIDAGYRGVISALLYYRPG.VTVILPGHLTIYLFPVKLRQSRLL.........PKNVLKHLDPI
       --------------------              -----------                                          --------------------
             VAR 2                        MOTIF 3                                                VAR 3...

       201                                                                                          300
EHV-2  ...............................DYDLLNRPQYSDDAGYDLYLQEDLMLFPQASTTVTIDSRV..PTTTKFFKP
HVS    ...............................DYSDLCPPRYSKDAGFDLYLPTDVTVKPRVPNRYSVNICC..PAQLKSYKP
EBV    ...............................EDKGPINHPQYPGDVGLDVSLPKDLALFPHQTVSVTLTVP..PPSIPHHRP
EHV-1  ..............VTSIGSTRDGRTSEANLFYDYFAPKRVEDAGYDISAPEDATIDPDESHFVDLPIVF..ANSNPAVTP
EHV-4  ..............VTNINLTENGRSPRVNIFYDYFAPKRDEDAGYDISAQTNATIEPDESYFVELPIVF..SSSNPAVTP
BHV-1  ..............LIAVDDDANAGTEAGVEVPFFATFAPKRDEDAGYDIAMPYTAVLAPGENLHVRLPVAY..AADAHAAAP
HSV-1  GAGIREDPWLEGALGAPSVTTALPARRRGRSLVYAGELTPVQTEHGDGVREAIAFLPKREEDAGFDIVVRRPVTVPANGTTVVQPSLRML.HADAGPAAC
HSV-2  LAGLREDPWLDGALATAGGAVALPARRRGGSLVYAGELTQVTTEHGDCVHEAPAFLPKREEDAGFDILIHRAVTVPANGATVIQPSLRVL.RAADGPEAC
PRV    .....................VPFFEVFAPKRDEDAGYDIPCPRELVLPPGGAETVTLPV....HRTDGRHWA
VZV    FKSIQVQPLSNSPSNYEKPVPIPEFADISTVQQGQPLHRDSAEYHIDVPLTYKHIINPKRQEDAGYDICVPYNLYLKRNEFIKIVLPIIRDWDLQHPSINA
       -----------------------------------------------        -----
              VAR 3 continued                                Motif 1

       301                                                                                          400
EHV-2  VVFGRSGLATRGVVVDVVKWTHS.PLTLKIYNFTDNTLRYSAGTRICQVVFVHRRHFPSKLKH.FFTYINLNSKTSFYWANVSFV..DCQNDAYRSLVTL
HVS    TIFGRSGLAAKGLTIKVSRWQN..QLQIIFYNYTKSQITYTARTRIAQVVFMHKKHLPTTLTR.LKPTMHLSENIKYSWARVSFQDIKTFPVQDEKLYSS
EBV    TIFGRSGLAMQGILVKPCRWRRG.GVDVSLTNFSDQTVFLNKYRRFCQLVYLHKHHLTSFYSP.HSDAGVLGPRSLFRWA......SCTFEEVPSLAMG
EHV-1  CIFGRSSMNRRGLIVLPTRWVAGRTCCFFILNVNKYPVSITKGQRVAQLLLTEDIDDALIPPT.VNYDNPFPTYSPSEST......KAPQSPVLWKFTT
EHV-4  CIFGRSSMNRRGLIVLPTRWVTGRTCCFFILNINKYPVNITKGQRVAQLVLTEDIDEALIPTN.VNYNTPFPTYISPTGAV.....KHNPTPILWKFTE
BHV-1  YVFGRSSCNLRGLIVLPTAWPPGEPCRFVLRNVTQEPLVAAAGORVAQLLLL.ARRLEWLPSG.LNDREPFPTSPRAAPP.....APGAPRLRWRRVA
HSV-1  YVLGRSSLNARGLLVVPTRWLPGHVCAFVVYNLTGVPVTLEAGAKVAQLLVAGADALPWIPPDNFHGTKALRNYPRGVPDST.....AEPRNPPLLVFTN
HSV-2  YVLGRSSLNARGLLVMPTRWPSGHACAFVVCNLTGVPVTLQAGSKVAQLLVAGTHALPWIPPDNIHEDGAFRAYPRGVPDAT.....ATPRDDPILVFTN
PRV    YVFGRSSLNLRGIVVFPTPWESG.PCRFRIQNRGAHPVTLESGQRVAQLVLT.REPLGWITGR....SPFPATPR.......APMQHRPAWLFAR
VZV    YIFGRSSKSRSGIIVCPTAWPAGEHCKFYVYNLTGDDIRIKTGDRLAQVLLIDHNTQIHLKHN.VLSNIAFPYAIRG.........KCGIPGVQWYFTK
       ----                                        ------            -------
       MOTIF 2                                     MOTIF 4           VAR 4

       401                      432
EHV-2  PCQEDTDRGYRGDSGFGSSGMR..........
HVS    SKDTSDSQMSRGDAGLGSSGLM..........
EBV    DSGLSEALEGRQGRGFGSSGQ...........
EHV-1  DFDREAPSSLRADGGFGSTGL...........
EHV-4  AFDHDAPSSARSEGGFGSTGL...........
BHV-1  DLAAAVPPSARGPRGFGSTGL...........
HSV-1  EFDAEAPPSERGTGGFGSTGI...........
HSV-2  EFDADAPPSKRGAGGFGSTGI...........
PRV    DF..VAPSSARGARGFGSTGL...........
VZV    TLDLIATPSERGTRGFGSTDKETNDVDFLLKH
                 --------
                 MOTIF 5
```

Figure 4.9 Class II sequence alignment using SAGA

This alignment was generated by the program SAGA with the standard set of 10 class II sequences. Motif regions are highlighted in yellow and variable regions in blue text. Note that the order of sequences is different than in Figure 4.8.

Variable region 2 is almost identical in both alignments. Alignment in this region is constrained by insertions in the EHV-1 and EHV-4 sequences. Region 3 is the main area of variability between class II species and of significant importance to the class II model. Both alignments show very similar resolutions, notably constrained by insertions in the HSV-1 and HSV-2 sequences and to a lesser degree the VZV sequence. Between the two alignments the position of the gap regions varies by only a few residues in all the sequences. It is the position and length of this gap that is of importance and both alignments are in general consensus by these criteria. This region of class II variability is considered in depth in the following sections.

Variable region 4 is found between motifs 4 and 5. In comparison to the FD result, the SAGA method favours the condensation of single gaps into one region. These gaps are fairly small compared with the other variable regions but will be taken into account in the model. Finally it is of consequence to note that there is variability in terms of sequence length at the N-terminus while at the C-terminus, only VZV has additional residues.

4.3.3 Proposed relationship between class I and class II dUTPases

The intragenic duplication model proposed by D.J. McGeoch (1990b) accounts for the genetic rearrangement of the five motifs in class II dUTPases compared to the class I group. The genetic evidence for this rearrangement in addition to the structural data from the class I enzymes supports the following hypotheses:

- the five motifs which make up the active site in the class I dUTPases constitute a highly homologous active site in the class II dUTPases
- the active site which is assembled from motifs donated by three subunits in the class I trimer can be successfully duplicated in the class II monomer by a single protein chain

To further these hypotheses in a structural context it was necessary to provide a model to link the class I and class II dUTPases. It has been shown that the class I genes encode a subunit structure which subsequently assembles into a trimeric arrangement. Since the class II genes are approximately double the length of the class I genes and encode a monomer, a more defined hypothesis was proposed:

- duplication of the ancestral class I gene led to a duplication of the trimer subunit structure to create a double length class II monomer

This hypothesis has implications for the class II model. In terms of overall structural arrangements, the class II dUTPase is proposed to be constructed by two

trimer subunits joined as a single protein chain. For this new structure to be viable it would have to be able to condense all five of its single copy motifs into one functional active site.

To test this model it was necessary to provide sequence lineups between the class I and class II dUTPases. Since it was proposed that the class II arrangement originated by a duplication of the class I gene, the two classes were compared directly. This was achieved by duplicating the *E.coli* primary sequence to create a single chain doublet. This double length *E.coli* sequence was then used to align with the class II sequences. A representation of this alignment is shown in Figure 4.10.

Figure 4.10 Comparison of class I and class II sequences utilising an *E.coli* doublet
Figure shows a representation of the *E.coli* class I and the HSV-1 class II sequences indicating the position of the motif regions. The class I sequence is effectively duplicated and joined to create a double length chain. To permit correct alignment of the motif regions, gaps must be inserted for the extended joining region and the extension between motifs 4 and 5 in the HSV-1 sequence. This allows direct sequence comparison between the two classes. The motif region labelled 'X' is discussed in Section 4.7.

In this diagram the single copy of the five class II motifs is aligned with the respective *E.coli* motifs. What cannot be seen is how closely related these motifs are between the two classes. In fact when they are compared directly it can be seen that all of the key residues identified in the class I structures are present in the class II motifs.

Figure 4.11 Direct comparison of *E.coli* and HSV-1 motifs
E.coli motifs are shown on top with the equivalent HSV-1 motifs below. The residues which interact directly with dUDP in the *E.coli* co-crystal are shown in yellow. Note that these residues are almost totally conserved throughout the class II group as well as the class I group.

4.3.4 Discussion

The class II dUTPases have been shown to be more heterogeneous between species than their class I counterparts. The use of alignment programs has allowed them to be compared at the primary sequence level. By using different methods to analyse the same data set it was possible to identify distinct areas of homology and variability within the class II group with a degree of confidence. The use of the program SAGA for this purpose proved to be very useful. This type of alignment program has only become possible with the advances in computing power and even with a powerful VMS system, processor time for a single alignment was in the region of 10 hours.

Finally a method has been developed to allow direct comparison of the two dUTPase classes at the primary sequence level. Although this method is based on the specific hypotheses in Section 4.3.3, an element of caution must be used. It should be noted that the intragenic duplication is likely to have occurred from a common ancestor of the class I and class II dUTPases. It is therefore probable that the class I enzymes have diverged to some degree since that time. This should be taken in the context of the available information, and given that the structural conservation between diverse class I species is high, it is likely that significant changes are few. In terms of this study, the doublet molecule generated from the *E.coli* gene must be considered to be at least partially divergent from the original ancestral doublet. This is considered more thoroughly in Section 4.6 but at this stage only major structural elements which are more likely to have been conserved through time will be examined. The extended regions present in the HSV-1 sequence compared to the *E.coli* doublet are discussed in the next section.

4.4   Secondary structure prediction

4.4.1   Introduction

As will be shown, the class I subunit structures all possess a signature secondary structure arrangement. If the hypothesis in the previous section holds true then it is likely that each half of the class II enzyme will contain some of these secondary structural elements. This was tested by using computer analysis to predict secondary structure from primary sequence lineups of the class II dUTPases. The predicted class II β-sheet arrangement was then compared with the class I doublet of known structure.

4.4.2   Class I β-sheet structure

E.coli dUTPase was the first class I structure to be solved by crystallography and the secondary structure is depicted in the introduction Figure 1.11. Given the low inter-motif sequence homology between the class I enzymes it was of interest to compare these structural elements with the other subsequently solved structures. It was immediately obvious, as in the comparison of quaternary structures in Section 4.2.3, that the class I group share highly homologous secondary structure arrangements. This section details these conserved secondary structural elements in order to obtain a consensus to probe the class II enzymes.

It was first noted in the *E.coli* structure that the arrangement of β-strands conforms approximately to a 'jelly roll' fold (Cedergren-Zeppezauer *et al.*, 1992) which is depicted in Figure 4.12. Each subunit possesses 10 β-strands and a short α-helix. Eight β-strands form two β-sheets and create an overall β-barrel arrangement. The two remaining short anti-parallel β-strands are not part of the β-barrel. The eighth β-strand in the 'jelly roll' fold configuration is donated by an adjacent subunit (more clearly visualised in the introduction Figure 1.11).

(a) Idealised jelly roll fold

(b) *E.coli* secondary structure

(c) Comparison of four class I secondary structures relative to primary sequence
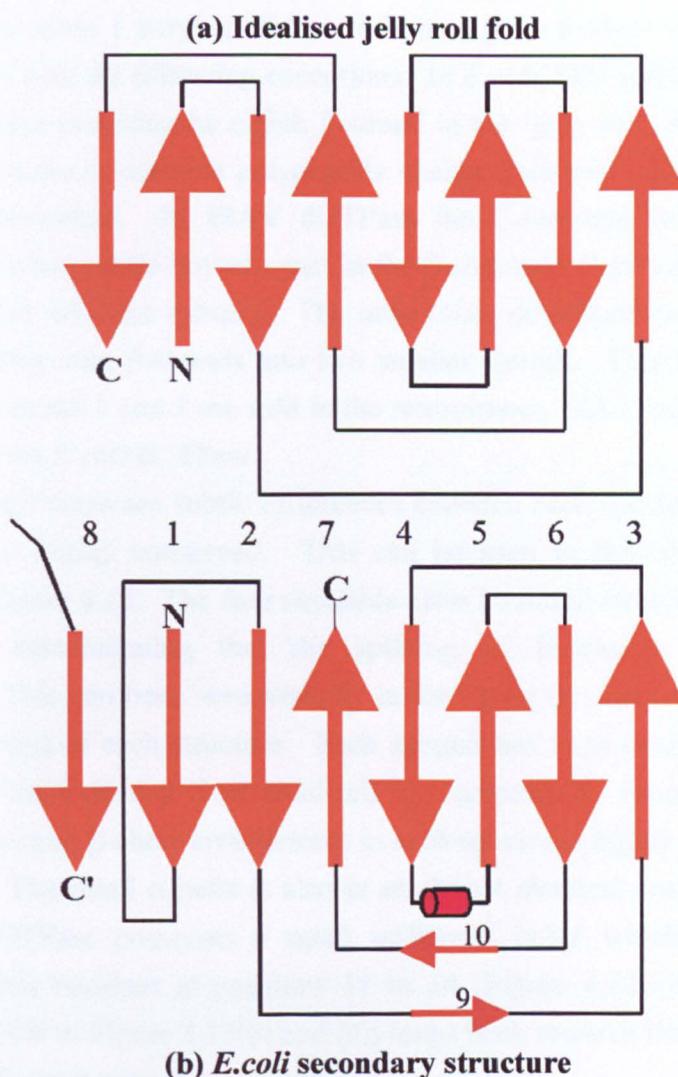
```
            -------1-------            --2-- ---9---    ------3-------   ---4---   HELIX
EIAV     MLAYQGT   QIKEK    RDEDAGFDLCVPYD IMIPVSDTKIIPTDVKIQVP PNSFGWVTGKSSMA KQG
FIV       MIIEGD   GILDK    RSEDAGYDLLAAKE IHLLPGEVKVIPTGVKLMLP KGYWGLIIGKSSIGSK G
Human    MQLRFARL  SEHATAPTRGSARAAGYDLYSAYD YTIPPMEKAVVKTDIQIALP SGCYGRVAPRSGLAAKHF
E.coli MMKKIDVKILDPRVGKEFPLPTYATSGSAGLDLRACLNDAVELAPGDTTLVPTGLAIHIADPSLAAMMLPRSGLGHKHG
                                        -----                              ----
                                       Motif 1                           Motif 2

            -a---5--b--     ---6------   -10-   -----7----- ---8----
EIAV     LLIN  GGIIDEGYTGEIQVICTNIGKSNIKLIEGQKFAQLIILQHHSNSRQPWDENKISQRGDKGFGSTGVF
FIV      LDVL  GGVIDEGYRGEIGVIMINVSRKSITLMERQKIAQLIILPCKHEVLEQGKVVMDSERGDNGYGSTGVF
Human    IDVG  AGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIAQLICERIFYPEIEEAQALDDTERGSGGFGSTGKN
E.coli IVLGNLVGLIDSDYQGQLMISVWNRGQDSFTIQPGERIAQMIFVPVVQAEFNLVEDFDATDRGEGGFGHSGRQ
         ---------              --------          --------
          Motif 3                 Motif 4           Motif 5
```

Figure 4.11 Comparison of the class I secondary structure to a "jelly roll" fold
Two topological diagrams are shown: (a) an idealised 'jelly roll' fold and (b) the structure of *E.coli* dUTPase. β-strands are depicted as arrows representing their direction and the α-helix as a cylinder. A direct comparison of β-strand positions in relation to primary sequence is given in (c). The eight β-strands composing the jelly roll fold are highlighted in yellow with the additional strands in grey. α-helices are highlighted in red. A four residue insertion in the *E.coli* sequence is highlighted in cyan (see text below for details). β-strands are numbered 1-10 for comparison.

All four class I enzymes follow a homologous β-sheet profile as depicted in Figure 4.12 (b) with the following exceptions. In *E.coli*, FIV and human dUTPases, the C-terminal region provides the eighth β-strand in the 'jelly roll' fold. This means that each globular domain contains polypeptide chains from two different subunits which are tightly interlocked. In EIAV dUTPase the C-terminal strand donated by the neighbouring subunit does not take part in the β-sheet structure but still makes intimate contact with the adjacent subunit. The only other deviations from the pattern are a result of splitting long β-strands into two smaller strands. This is apparent in Figure 4.12(c) where strand 1 and 3 are split in the retroviruses, EIAV and FIV and strand 5 is split in all but the *E.coli* dUTPase.

Although there are subtle differences between each species the general pattern of β-strands is highly conserved. This can be seen in the comparison of the 3D structures in Figure 4.13. The four available class I subunit structures closely resemble one another demonstrating that the splitting of β-strands has little structural consequence. This can been seen visually in the figure by comparing the bottom right area (N-terminus) of each structure. Each subunit has been positioned relative to the orientation of the extended C-terminal tail for comparison. From this view it can be seen that the overall β-sheet arrangement in each species is highly conserved relative to this position. The small α-helix is also in an almost identical position in each species. The *E.coli* dUTPase possesses a small additional bulge which corresponds to the insertion of four residues at positions 11 to 14 (Figure 4.12c in blue). This chain extension (visible in Figure 4.13(a) and (e)) loops back towards the main body and does not affect the β-sheet pattern.

Figure 4.13(e) shows all four class I structures superimposed without any structural modifications. Apart from the small extended loop visible in the *E.coli* dUTPase and to a lesser degree in the human dUTPase the structural backbones match almost exactly. This is a fascinating observation given the diversity of these four species. It highlights the relative unimportance of inter-motif primary sequence conservation in comparison to conservation of polypeptide chain length to the maintenance of the subunit structure.

(a) *E.coli*

(b) FIV

(c) Human

(d) EIAV

(e) Superimposed ribbon traces
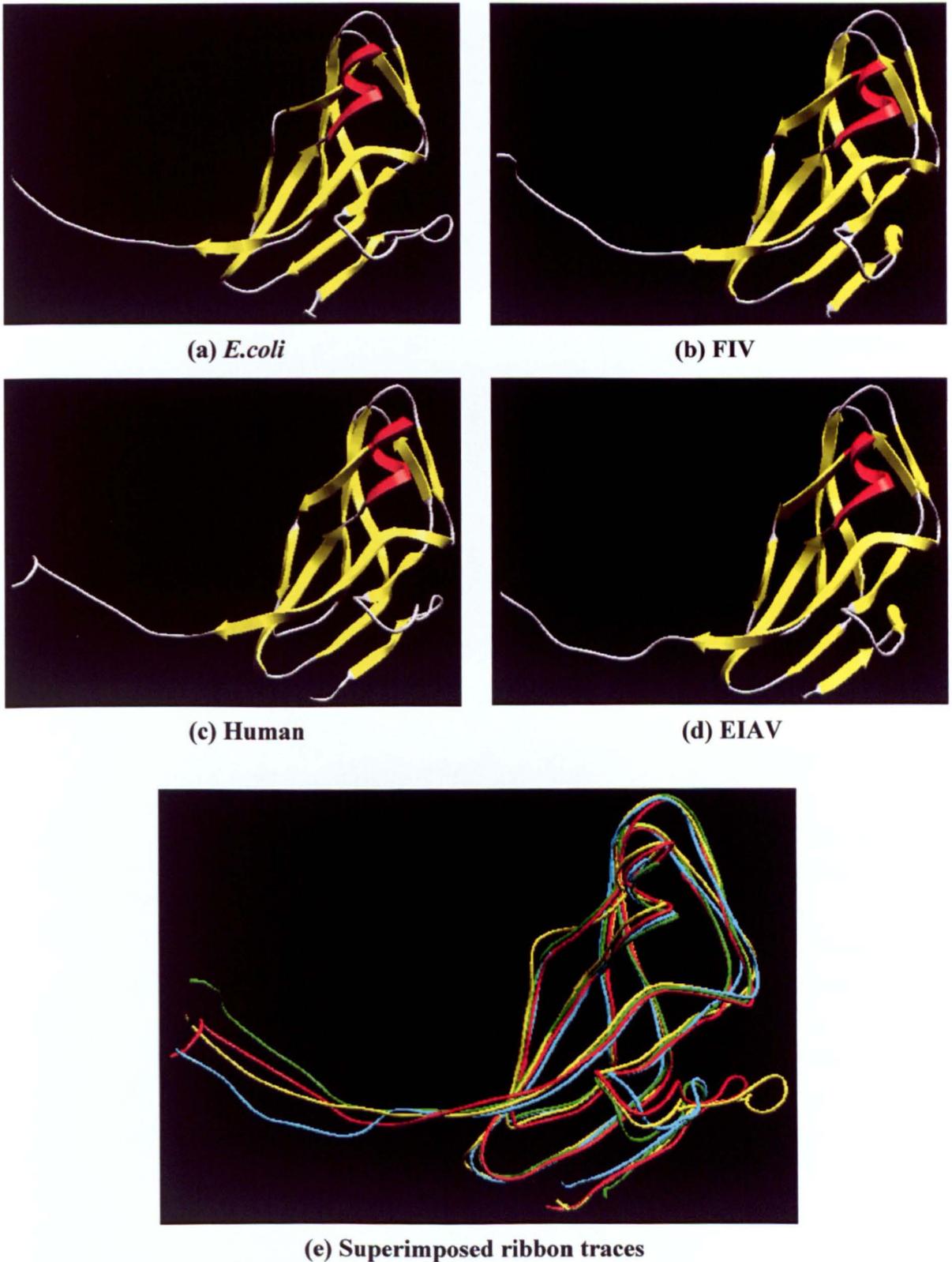
Figure 4.13 β-sheet homology between class I dUTPases
Comparison of secondary structure of the class I dUTPases. Each subunit is depicted as an α-carbon ribbon. β-strands are shown as yellow arrows depicting their direction while α-helical regions are in red. In part (e) all four species are superimposed in the following colours: *E.coli*-yellow, FIV-green, human-red, EIAV-cyan. The Swiss-PDB Viewer was used to generate these ribbons and to superimpose the four class I species by aligning α-carbon backbones.

Given the strong conservation of these secondary structural elements in the class I dUTPases it was of interest to compare this pattern to the class II enzymes. Since there was no structural data available for any member of the class II group it was necessary to employ secondary structure prediction programs. The results of this analysis are given in the following sections.

4.4.3   Secondary prediction of class II dUTPases

The program used for secondary structure prediction of the class II group was PredictProtein at the automated server at EMBL, Heidelberg. This program is stated to give expected 3 state accuracies (helix, strand, other) of over 72% depending on the input data (Rost and Sander, 1993 and 1994). This program is based on a system of trained neural networks which allows greater reliability than single sequence or statistical alignment inputs.

The basis behind the prediction method is the utilisation of multiple sequence alignments. This program was therefore appropriate for comparing the class I and II dUTPase using the previously refined alignments. The program can use a single query sequence or a multiple alignment can be supplied. For single sequence input the program scans the Swiss-Prot database for homologues and generates an alignment using the MaxHolm method (Sander and Schneider, 1991).

The main objective of this work was the secondary structure prediction of HSV-1 dUTPase. Initial experiments were therefore carried out using whole or part of the HSV-1 primary sequence as single input data. This allowed the PredictProtein program to calculate its own alignments with the MaxHolm program. Given the potential for alignment problems using single sequence input this strategy was improved.

Experiments using the HSV-1 sequence as single input data were unreliable. This was due to the lack of conserved sequences detected and aligned with the HSV-1 sequence by the automated MaxHolm program. Improved alignments could be generated by splitting the HSV-1 sequence into smaller fragments. This allowed the MaxHolm program to align the C-terminal half of the HSV-1 protein to class I species due to the local similarities in motifs 1, 2, 4 and 5. However, the N-terminal half could not be satisfactorily aligned by the MaxHolm program. The lack of conserved sequences found in this area was below the criteria recommended for secondary structure prediction. This is not surprising given that there is no sequence homology with the class I structures outside motif 3 in this half of the HSV-1 sequence.

The PredictProtein program has the ability to utilise ready made alignments for its prediction. Although it was specifically the HSV-1 sequence which was of interest, a better prediction would be realised with refined class II lineups. Given the primary sequence differences between the class II dUTPases it was decided to use a bank of

alignments including all the available class II sequences and subsets based on phylogenetic relationships. This was deemed appropriate since the Feng and Doolittle alignment program had successfully partitioned each of the class II dUTPase into the α and γ subfamilies directly from dUTPase primary sequences.

An example prediction using pre-aligned sequences as input data is given in Figure 4.14. The most obvious observation from the ProteinPredict output data is the abundance of predicted β-strands throughout the lineup. It is important to note that this is not a prediction of the EHV-1 sequence alone and this sequence is given only as an alignment marker. The prediction is in fact from an alignment of 8 class II sequences. An overall reliability index is shown (Rel sec) and is relatively high in areas of predicted β-strands ('E'). This reliability index is broken down into helix, strand and loop and again the majority of β-strand are predicted with a high probability. Further confidence is given by the final line of output data which indicate areas where the prediction is likely to be over 82% accurate. Predicted β-strands have been highlighted in yellow for clarity.

It must be noted that the reliability index is based on an efficient alignment with sequence conservation varying from 80% identity down to 20% and any deviation from these criteria will result in a lower reliability. Since the sequences used here do not align with great efficiency it is likely that the secondary structure predictions will be several percentage points lower than the 72-82% accuracy range. Even so, although the alignment is constructed from the class II group exclusively, species are included from the two different subfamilies α and γ. Members from these groups differ substantially at the primary sequence level which is actually advantageous to the prediction program as it more closely satisfies the ProteinPredict variability criteria. Additionally greater weight can be inferred from the predictions by using several independent alignments.

In addition to the β-strand prediction shown, there are several predicted extended loop regions. Interestingly these areas correspond approximately to variable regions 2 and 3 (see previous section) which are highlighted in blue. This observation supports the theory that variation in sequence length between the class II species is caused by the insertion of additional loop regions and not by a change in the overall β-sheet arrangement.

```
                1                                                                                  100
AA      |.................MTLRKQRQEIYYEFTSSSFEITSPPESS...KITLTN....LYPILVRPYVPA................VIPLGIRIIYGNKGQ|
PHD sec |    E    E EE        EEEE     EEEEEE      EEEEEEE        EEEEE              HHHHHHEEEEEEEEE          E|
Rel sec |9661147311225144553123542132136651789952699869999846767652257524778887754346777523432122545566258762|
prH sec |0010000121111122211101211110010211110000000000000000100111000000000001010000111255554432100000000111|
prE sec |0114421234356411112432113455431114788863100178998866111111356775211000011232110000012334466667752101 4|
prL sec |9774467543432455665455664433457764100025789820000037777765321236788887766667787633211212222222478763|
SUB sec |LLL...L.....E...LL....L.......LLL.EEEEE.LLLLEEEEEE.LLLLLLL...EEE.LLLLLLLLL...LLLLL........E.EEEE.LLLL.|


                101                VAR 2                                                 VAR 3...    200
AA      |GFILAGSSQKKV                              FCHTGLIDPGYRGEIKLIVLNTTKYNVTLFAGELRVSLFSFFFSTPIIY         |
PHD sec |EEEEEE                                      EEEEEEE     EEEEEEE        E    EEEEEEEEEE               |
Rel sec |8999943799999999999889986897511212665787457535999999658888512146722699988522476868998876888997999987|
prH sec |0000000000000000000000000001244332111000000000000000000011001100000011000000000000000121010010000001|
prE sec |8998863100000000000000000110000022477678762123689999872100024543113578988865421111000000000000000000|
prL sec |1000036788899999988888878887554331120137865200000027888864346775310001112456778789988878888888899888|
SUB sec |EEEEE..LLLLLLLLLLLLLLLLLLLLLLL.....EEEEE.LLL.EEEEEEELLLLLL....LL..EEEEEEEE...LLLLLLLLLLLLLLLLLLLLLLLL|


                201                VAR 3 cont.                                                       300
AA      |                                            DYDLLNRPQYSDDAGYDLYLQEDLMLFPQASTTVTIDSRV..PTTTKFFKP|
PHD sec |                                   E                EEE    EEE    EEEEEE E            EEEEE|
Rel sec |6777777876665666557866678767798321213576776677886332225688764457863241254232489955898712112466679883 2|
prH sec |1111000111222122211111111110001110000111221111123211100112320000001110000000000000000000112212000001|
prE sec |0000110000000000000001000010000334443211100000001222232100010000135643125552100268987534442100000135|
prL sec |77787788777777777778777877789855434567777777788755445678877666787532456633468996210014544567778888641|
SUB sec |LLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....LLLLLLLLLLLL....LLLLLLL..LLLL.....L.....LLLL.EEEEE......LLLLLLLL..|


                301                                                                                  400
AA      |VVFGRSGLATRGVVVDVVKWTHS.PLTLKIYNFTDNTLRYSAGTRICQVVFVHRRHFPSKLKH.FFTYINLNSKTSFYWANVSFV..DCQNDAYRSLVTL|
PHD sec |EEEE       EEEEE      EEEE       EEEEEEE                                                        EEEEE|
Rel sec |6864133322237998516789971479998347970355343011145543225876568989877788898888877888744667788983134522|
prH sec |0000000010000000000000000000000000000001234334322232111110000000011100011000100012211111100000000000|
prE sec |7876433333467998642100014688888621014667521013466655431011220000000000011110000003566754|
prL sec |1123465555531001247889975210013689843222554431000112478766788888877888888888778887666777888864322451|
SUB sec |EEE........EEEEE.LLLLLLL..EEEEE.LLL..EE........EE....LLLLLLLLLLLLLLLLLLLLLLLLLLLLLL..LLLLLLLL....E..|


                401           432
AA      |PCQEDTDRGYRGDSGFGSSGMR.........|
PHD sec |                HHHHHH  |
Rel sec |5578999777787886887655634455269|
prH sec |0100001111000000111111156667520|
prE sec |2211000000001001000122110000100|
prL sec |6678899878888887887766663332379|
SUB sec |LLLLLLLLLLLLLLLLLLLLLLLL...HH.LL|
```

Figure 4.14 Typical output from ProteinPredict program

The 8 sequence alignment generated by SAGA in Figure 4.8 was used as input data to the ProteinPredict program. This bypassed the MaxHolm alignment method which is the program default. The output data are arranged with respect to the EHV-2 sequence, corresponding to the first entry in the SAGA alignment. β-strands over 3 residues in length are highlighted in yellow and variable regions 2 and 3 are highlighted in blue. Further annotations from the ProteinPredict program are as follows:

AA - amino acid sequence of first entry in lineup (in this case EHV-2)
PHD sec - Profile network prediction Heidelberg: H=helix, E=extended (sheet), blank=other (loop)
Rel sec - Reliability index of prediction (0-9)
prH - 'probability' for assigning helix
prE - 'probability' for assigning strand
prL - 'probability' for assigning loop
SUB sec - a subset of the prediction, for all residues with an expected average accuracy > 82%
For the latter subset the following symbols are used: L is loop and "." means that no prediction is made for this residue, as the reliability is less than 5.

4.4.4 Comparison of class I structure to class II predictions

Each data set obtained from the ProteinPredict program contained large amounts of data similar to that in Figure 4.14. All of these predictions were compared directly to the *E.coli* doublet sequence using a new alignment. This was achieved by aligning the class II sequences to the *E.coli* doublet by the standard Feng and Doolittle method. The position of the predicted strands was noted and transposed onto this new lineup for comparison. An example of such a comparison is shown in Figure 4.15.

```
         1                                                                                                    100
Ehv2    ..................................................................MTLRKQRQEIYYEFTSSSFEITSPPESSKITLTN
Hvs     ..................................................................MPYKVPEIYYRFEPQTFYITSPARASNLQLIN
Ebv     ..................................................................MEACPHIRYAFQNDKLLL.QQASVGRLTLVN
Ehv1    ..........MASVTNLVD..SIVVVECGERWRARAEAAGRLLVL.INNHTVELSGEHGSAGE.........FYSVLTDVGVRVACSS
Ehv4    ..........MASATNLAD..NIVVVECSNGWEARAEADGRLLVL.INNHTVELSAGLGSAGE.........FYSVLTDVGVRVACSS
Bhv1    ......MANSAAATTATMSGDRGILVVE......LNAEAAPWRLESCCEPDSLALWGPIAPAAKRDETAPSGSLLYSRLINLNMKAAAPG
Hsv1    ......MSQWGSGAILVQPDSLGRGYDGDWHTAVATRGGGVVQLNLVNRRAVAFMPKVSGDSGWAVGRVSLDLRMAMPA
Hsv2    ......MSQWGPRAILVQTDSTNRRNADGDWQAAVAIRGGGVVQLNMVNKRAVDFTPAECGDSEWAVGRVSLGLRMAMPR
Prv     .......MEESAGATSAQSAATSVSESPAEETILVCASEPVTVDGGRL......LVCRSPGPEGFY..KVPLGLKVALPT
Vzv     MNEAVIDPILETAVNTGDMFCSQTIPNRCLKDTILIEVQPEC....ADTLQCVLDDKVSRHQPLLLRNHKKLELPSEKSVTRGGFYMQQLELLVKSAPPN
PHSV-1                              EEEE   EEEEE   HHHHHHHH    EEEEEEE   EEEEEE        EEEEEEEEEEEEE
Ptotal           HHHH                EEEEEE              EEEEEE          EEEEE   EEE        EEEEE     EEEEE
Cryst                               EEEEE                               EEEEE   EEE        EEEEEEEEEEEEE
Ecoli   ..................MMKKIDVKILDPRVGKEFPLPT.....YATSGSAGLDLRACLNDAVEL.....APGDTTLVPTGLAIHIADAP
                 1                                            2          9           3

         101                                                                                                  200
Ehv2    LYPILVRPYVPAVIPLGIRIIYGNKGQGFILAGSSQKKVFCHTGLIDPGYRGEIKLIVLNTTTKYNVTLFAGELRVSLFSFFFSTPII.YDY..DLLNR..
Hvs     HNNILVKAGQVTIVSTGI.IFPKETSFAFILYGKSAKSIFCHTGLIDPGFQGELKLIVLNKTEDDITLFENDLRVSVTAFVYGVPKL.HDY..SDLCP..
Ebv     KTTILLRPMKTTTVDLGLYARPPE.GHGIMLWGSTSRPVTSHVGIIDPGYTGELRLILQNQRRYNSTLRPSELKIHLAAFRYATPQMEEDK..GPINH..
Ehv1    GYAIVLTQI.SGLLPVE....PEPGNFSNVTFPENSAKYYTAYGIVDSGYRGVVKAVQFAPGI.NTSVPPGQMSLGLVLVKLARKSIHVTS..IGSTRD.
Ehv4    GYAIVLAQI.SGLPHVG....REPGNFSNITFTGNLANYYTAYGIVDSGYRGVVKAVQFANGV.NTVVPPGCMSLGLVLVKLSTETINVTN..INLTEN.
Bhv1    GYAIIMSQMRSGDTHMP....RPPA...............VAVGIVDSGYSGILRAIVWAPES.AAAAPPAGLALRLTLARLTTTLPRLIA..VDDDANA
Hsv1    DFCAIIHAP...........ALASPGHHVILGLIDSGYRGTVMAVVVAPKR.TREFAPGTLRVDVTFLDILATPPALTE..PISLRQF
Hsv2    DFCAIIHAP...........AVSGPGHVMLGLIDSGYRGTVLAVVVAPNG.TRGFAPGALRVDVTFLDIRATPPTLTE..PSSLHRF
Prv     GYAMLVAQR...........G...GGRTTNGIVDAGFRGEVQAIVAPGRP.RAQFYCTPLR..............
Vzv     EYALLLIQCKDTAL..........ADEDNFFVANGVIDAGYRGVISALLYYRPG.VTVILPGHLTIYLFPVKLRQSRLLPKN..V......
PHSV-1  EEEEEEE                          EEEEEEE     EEEEEEE      EE    EEEEEEEEEE       EE     E
Ptotal  EEEEEE                   EEE            EEEEEEEEE   EEEEEEEE        E      EEEEEEEEEE
Cryst   EEEEEEEEHHHHHHH                    VLGNLVGLIDSDYQGQLMISVWNRGQDSFTIQPGERIAQMIFVPVVQAEFNLVEFDATDRGH
Ecoli   SLAAMMLPRSGLGHKHGI.................VLGNLVGLIDSDYQGQLMISVWNRGQDSFTIQPGERIAQMIFVPVVQAEFNLVE
             4                              5          6         10          8

         201                                                                                                  300
Ehv2    ..................................................PQYSDDAGYDLYLQEDLMLFPQASTTVTIDS..RV
Hvs     ..................................................PRYSKDAGFDLYLPTDVTVKPRVPNRYSVNI..CC
Ebv     ..................................................PQYPGDVGLDVSLPKDLALFPHQTVSVTLTV..P.
Ehv1    GRTSEA.........NL.FYD............YFAPKRVEDAGYDISAPEDATIDPDESHFVDLP.I.VF
Ehv4    GRSPRV.........NI.FYD............YFAPKRDEDAGYDISAQTNATIEPDESYFVELP.I.VF
Bhv1    GTEAGV........EVPFFA............TFAPKRDEDAGYDIAMPYTAVLAPGENLHVRLP.V.AY
Hsv1    PQLAPPPPTGAGIREDPWLEGALGAPSVTTALPARRRGRSLVYAGELTPVQTEHGDGVREAIAFLPKREEDAGFDIVVRRPVTVPANGTTVVQPS.LRML
Hsv2    PQLAPSPL..AGLREDPWLDGALATAGGAVALPARRRGGSLVYAGELTPVQTTEHGDCVHEAPAFLPKREEDAGFDILIHRAVTVPANGATVIQPS.LRVL
Prv     ..LAPGIAT......DVPFFE............VFAPKRDEDAGYDIPCPRELVLPPGGAETVTLP...V
Vzv     ..LKHLDPIFKSIQVQPLSNSPSNYEKPVIPEFADISTVQQGQPLHRDSAEYHIDVPLTYKHIINPKRQEDAGYDICVPNYLYLKRNEFIKIVLPIIRDW
PHSV-1          HHHH                EEEEE                EEEEE                 E            EEEEE EEEE
Ptotal                                                                       EEE   EEE     EEEE
Cryst             EEEEE                                                       EEEEE   EEE     EEEEEEEEEEE
Ecoli   GGFGHSGRQMMKKIDVKILDPRVGKEFFLPTYAT................................SGSAGLDLRACLNDAVELAPGDTTLVPTGLAI
             1                                                            2          9           3

         301                                                                                                  400
Ehv2    PTTTKFFKPVVFGRSGLATR.....GVVVDVVKWTHSPL.TLKIYNFTDNTLRYSAGTRICQVVFVHRRHFPSKLKHFFTYINLNSKTSFYWANVSFVDC
Hvs     PAQLKSYKPVLFGRSGLAAK.....GLTIKVSRW.QNQL.QIIFYNYTKSQITYTARTRIAQVVFMHKKHLPTTLTRLKPTMHLSENIKYSWARVSFQDI
Ebv     PPSIPHHRPTIFGRSGLAMQ.....GILVKPCRWRRGGV.DVSLTNFSDQTVFLNKYRRFCQLVYLHKHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEV
Ehv1    ANSNPAVTPCIFGRSSMNRR.....GLIVLPTRWVAGRTCCFFILNVNKYPVSITKGQRVAQLLLTEDIDDA....LIPPT.VNYDNPFPTYSPS..EST
Ehv4    SSSNPAVTPCIFGRSSMNRR.....GLIVLPTRWVTGRTCCFFILNINKYPVYITKGQRVAQLVLTEDIDEA....LIPTN.VNYNTPFPTYSPT..GAV
Bhv1    AADAHAAAPYVFGRSSCNLR.....GLIVLPTAWPPGEPCRFVLRNVTQEPLVAAAGQRVAQLLLLARRLE.....WLPSG.LNDREPFPTSPRA..APP
Hsv1    HADAGPAACYVLGRSSLNAR.....GLLVVPTRWLPGHVCAFVVYNLTGVPVTLEAGAKVAQLLVAGADALP....WIPPDNFHGTKALRNYPRGVPDST
Hsv2    RAADGPEACYVLGRSSLNAR.....GLLVMPTRWPSGHACAFVVCNLTGVPVTLQAGSKVAQLLVAGTHALP....WIPPDNIHEDGAFRAYPRGVPDAT
Prv     HRTDGRHWAYVFGRSSLNLR.....GIVVFPTPWESG.PCRFRIQNRGAHPVTLESGQRVAQLV.....LT....REPLGWITGRSPFPATPR.....
Vzv     DLQHPSINAYIFGRSSKSRS.....GIIVCPTAWPAGEHCKFYVYNLTGDDIRIKTGDRLAQVLLIDHNTQI....HLK....HNVLSNIAFPYAIRG
PHSV-1  E         EEEEE       EEEEEE          EEEEEEEEE       EEEE   HHHHHHHHHH
Ptotal          EEEEE          EEEEEE          EEEEEE         EEEE   EEEEEEEE
Cryst   EE       EEEEEEEEHHHHHHH  EEEEEEEEE      EEEEEEEE     EEEE     EEEEEEE
Ecoli   HIADPSLAAMMLPRSGLGHKHGIVLGNLVGLIDSDYQGQLMISVWNRGQDSFTIQPGERIAQMIFVPV.................................
             4          5          6         10          7

         401                        445
Ehv2    QNDAYRSLVTLPCQEDTDRG..YRGDSGFGSSGMR.........
Hvs     KTFPVQDEKLYSSSKDTSDSQMSRGDAGLGSSGLM.........
Ebv     PSLAM......GDSGLSEALEGRQGRGFGSSGQ.........
Ehv1    KAPQSPVLWKFTTDFDREAPSSLRADGGFGSTGL.........
Ehv4    KHNPTPILWKFTEAFDHDAPSSARSEGGFGSTGL.........
Bhv1    APGAPRLRWRRVADLAAAVPPSAGRPRGFGSTGL.........
Hsv1    AEPRNPPLLVFTNEFDAEAPPSERGTGGFGSTGI.........
Hsv2    ATPRDPPILVFTNEFDADAPPSKRGAGGFGSTGI.........
Prv     APMQHRPAWLFARDF..VAPSSARGARGFGSTGL.........
Vzv     KCGIPGVQWYFTKTLDLIATPSERGTRGFGSTDKETNDVDFLLKH
PHSV-1           HHH
Ptotal          EEE                    HHHHHH
Cryst   EEEEEEEE
Ecoli   ...VQAEFNLVEDFDATD....RGEGGFGHSG............
             8
```
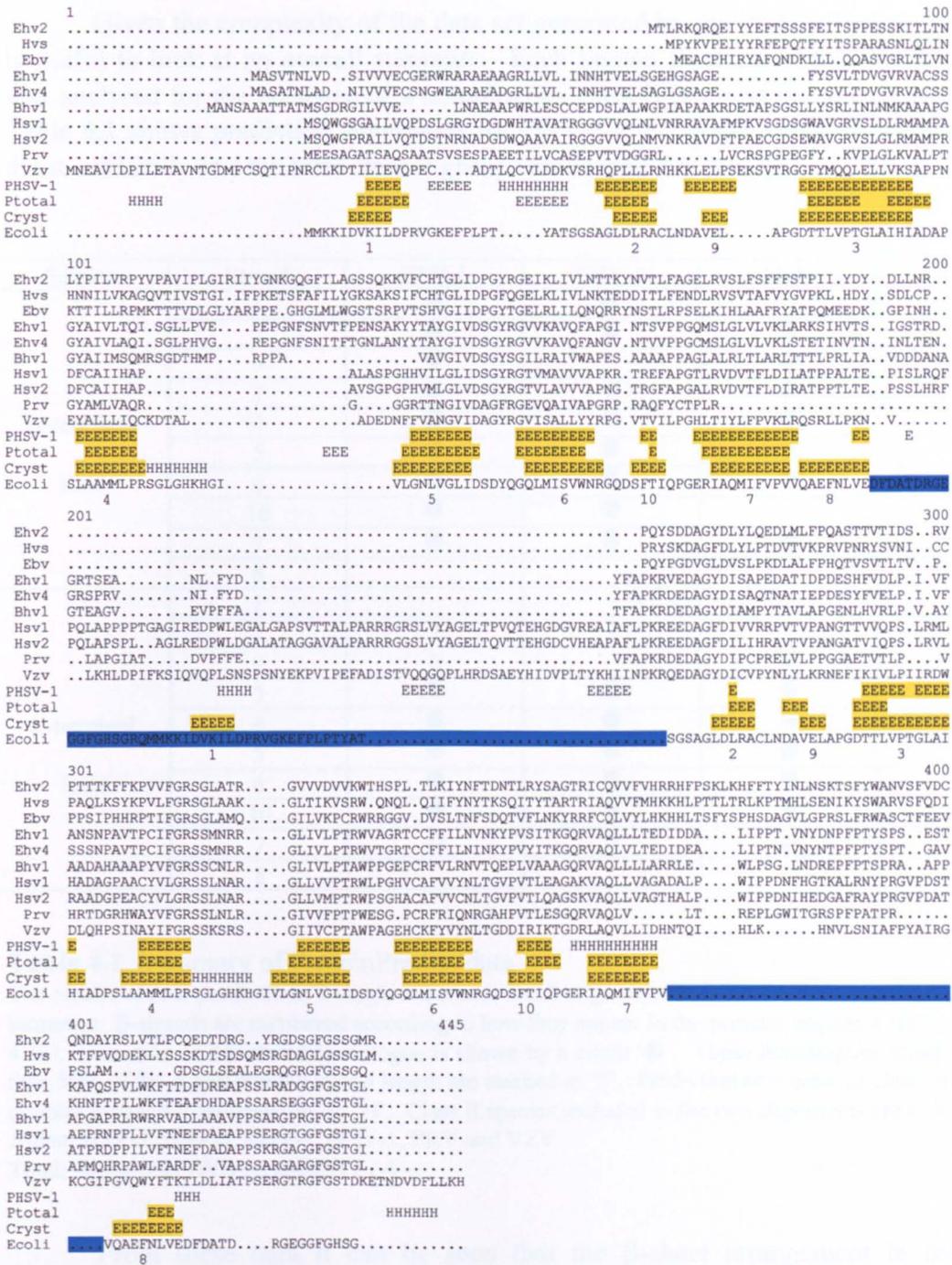
Figure 4.15 Comparison of class II predicted secondary structure to the *E.coli* doublet
The class II sequences shown were aligned with the *E.coli* doublet sequence using the GCG Pileup program. Three lines of secondary structure information are sandwiched between the class II group and the *E.coli* doublet. Secondary structures are indicated as an 'E' for extended β-strand and 'H' for helix. 'PHSV-1' indicated the secondary structure prediction using the HSV-1 sequence as single input data. 'Ptotal' indicates a prediction using a lineup of all available class II sequences as input data. 'Cryst' indicates the solved secondary structure in the *E.coli* crystal structure. Homologous β-strands are highlighted in yellow and numbered. The joining region and extension region (between motifs 4 and 5) are highlighted in blue.

Given the complexity of the data set generated by the ProteinPredict program it is useful to look at an overall summary. Each known β-strand in the *E.coli* structure was analysed for the presence of a homologous class II β-strand in various predictions. Table 4.1 shows prediction summaries for HSV-1 and EHV-1 single input sequences, the α subfamily and the total class II group (see legend for details).

Section	Strand	HSV-1	EHV-1	Alpha	Total
N-terminal Half	1	●	●		
	2	●	●	●	●
	9	●		●	
	3	●	●	●	●
	4	●	●	●	●
	5	●	●	●	●
	6	●	●	●	●
	10	●	●		
	7	●	●	●	●
	8				
C-terminal Half	1				
	2	?	?	?	?
	9	?		●	●
	3	●	●	●	●
	4	●	●	●	?
	5	●	●	●	●
	6	●	●	●	●
	10	●	●	●	●
	7	H	H	H	●
	8			?	?

Table 4.1 Summary of ProteinPredict data

The results of four predictions, 2 single sequences and 2 alignments, were compared to the *E.coli* doublet sequence. β-strands are numbered according to how they appear in the primary sequence (same as Figure 4.13). Detection of a β-strand homologue is shown by a circle '●'. Those homologues which have less than 50% of the *E.coli* doublet strand length are marked as '?'. Prediction of a helix in place of an *E.coli* doublet β-strand is indicated by an 'H'. Class II species included in the two alignments are as follows:
Alpha: HSV-1, HSV-2, EHV-1, BHV-1, PRV and VZV
Total: Alpha plus HVS (γ) and EBV (γ)

From these data it can be seen that the β-sheet arrangement in the *E.coli* dUTPase is predicted to be approximately duplicated in the class II dUTPases. This should be taken in the context of the alignment between the two classes represented in the previous section (Figure 4.10). There are two obvious gapped regions which are indicated on this figure and can also been seen highlighted in blue in the sequence comparison Figure 4.15. The latter of these gapped regions lies between motifs 4 and 5 corresponding to the region between β-strands 7 and 8. This extension is present in all the class II species analysed. It is clear that there is some divergence between the classes in this area and this is reflected in the low β-strand homologue predictions for strands 7 and 8. This is a key feature of the class II model discussed below.

Similarly, the region representing the join between the two class I subunits to form the monomeric class II molecule is extended with respect to the *E.coli* doublet. This region spans β-strands 8 from the N-terminal half to strand 2 from the C-terminal half. Comparison of class II sequences alone has demonstrated that this region is highly variable between species (Section 4.3). Again this is reflected in the absence of β-strand homologues predicted for this region. This is discussed with reference to the model below.

4.4.5 Proposed class II dUTPase structural arrangement

Several key observations were assembled to construct the class II structural model:

- the class I group are trimers while the class II group are monomers of approximately double the protein chain length
- each half of the class II molecule has a congruent β-strand arrangement to the class I subunit (by secondary structure prediction)
- a variable extended joining region exists between the two halves of the class II monomer
- a relatively conserved additional chain length exists between motifs 4 and 5 in the class II molecule
- each of the three class I active site regions is formed by motifs donated by all three subunits of the trimer
- the class II monomer has only one copy of each motif with which to form an active site.

In order to model the class II structure all of the above points must be accommodated in order to arrive at a structure which has the potential to be active, monomeric and have evolved through intragenic duplication. It is proposed that the class II active site will be constructed by the condensation of the five motif regions and as such will mirror the class I active site in terms of arrangement and enzymatic mechanism. Figure 4.16 shows a diagrammatic representation of the class II model.

This model was constructed using the Glaxo Wellcome Swiss-PDB Viewer. The starting point in the modelling process was the *E.coli* dUTPase structure complexed with dUDP (a). The trimer is shown with the 3 subunits coloured blue, green and yellow and rendered as ribbons. The dUDP molecules are coloured red and for clarity rendered as Van der Waals surfaces. Three dUDP molecules occupy the three active sites situated at the subunit interfaces. The visible motif regions 1-4 are highlighted in pink for the top-right active site. The motif 3 loop is clearly visible and is donated by the blue subunit. Motifs 1, 2 and 4, donated by the green subunit, are just visible behind the dUDP molecule. The C-terminal 16 residues containing motif 5 are

not visible in the *E.coli* crystal structure but extend from the yellow subunit tail. This tail can be seen pointing in the direction of the active site and is proposed to fold akin to the human dUTPase structure where motif 5 interacts with the substrate molecule and caps the active site. The model is progressed in four stages.

Diagram (b) shows the *E.coli* structure with the yellow subunit removed. This dimeric structure represents the approximate chain length equivalent of the class II dUTPases. The blue subunit represents the N-terminal half of the class II molecule while the green subunit represents the C-terminal half. The four motifs which constitute the active site in diagram (a) are conserved in structure (b) shown here interacting with a single dUDP molecule.

In the class I enzymes the active site is composed of motifs donated by all three subunits. A model is proposed to generate a homologous active site using only the two subunits available in structure (b). Motifs 1, 2 and 4 from the green subunit plus motif 3 from the blue subunit are forming an active site pocket mirroring the trimeric ancestor. In order to complete the active site, motif 5 is proposed to be donated from the extended arm of the green subunit. Structure (c) shows how this can be easily accomplished. To position motif 5 the extended arm is allowed to rotate by approximately 90° in the plane of the paper. The green subunit arm now occupies a homologous position to the yellow subunit arm in the original trimer.

This rearrangement is accomplished with minimal disruption to the overall structure and generates the possibility to condense all 5 motifs in a monomeric molecule. All that is required is an extension of this C-terminal arm to position motif 5 over the active site. As seen in the primary sequence studies, all class II dUTPases possess an extra chain length between motif 4 and 5 as compared to the class I group. This additional chain, of approximately 30-35 residues, is conserved exclusively within the class II group and is now proposed to allow correct positioning of motif 5. This extra chain is demonstrated in structure (d) where motif 5 is depicted in pink at the end of the C-terminal arm.

The only remaining rearrangement necessary is the connection of the two subunits to form a single chain. The last visible C-terminal residue of the blue subunit and the N-terminus of the green subunit are highlighted in yellow. Structure (e) shows the joining of these ends to complete the transformation in a monomer. A small loop is shown for this joining region but there is a large chain length variability between the class II species. This structure is therefore more representative of the shorter species such as EBV, EHV-2 and PRV. To arrive at a model more closely representing the

longer species like 1RN8 and 3DTU this loop would tend to be positioned in the
region of the active region on the opposite side from the ligand, so an extension
could be effectively positioned without disrupting this region or interfering with
the associated ligand-cofactor.

This observation, therefore gives at the very least a rationalisation to
a graphical, pseudo-physical, structural embodiment of the data trends. Modelling
the data in a structural manner as shown in (a) suggests a convincing rationale to
... model structure the data trends. ... of the
structural ... but ... found in this context can be explained by the presence
of this secondary structure prediction in the previous regions. The current model
does not highlight an extended sequence found between monomers ... of the class II
dUTPase. This is discussed in detail in Section 4.5.

4.4.6   Discussion

... sequence data have the ... relations between the
between ... molecules. In the ... the data has been
utilised ... subunit duplication with ... circumvent the
absence of ... structure data for the class II ... that permit further
investigation at ... level. Utilising sequence ... and secondary structure
prediction methods ... possible to continue the ... of structure
relationships. Residues ... secondary structure prediction were ... the construction
of class II model which provides insight into the structure of ... these
molecules.

Secondary structure prediction was carried out along the dUTPase
program. This allowed multiple ... models to be used to predict possible
class II secondary structure. ... prediction were used to determine new possible
effects. The first conserved structure family of predicted proteins ... the class
II group. Alignment of these proteins ... and another sequence ... and as
a high degree of complementarity ... as ... of the class II ... class I
structural relative of the classes ...

A model was constructed ... the ... data with secondary
structure prediction and deconstruction ... the model would portray the class
II monomer. The variable length inhibitor ... the ... helices in the ... the
molecule plus the extended region for the ... embodiment of each of this in ...
structurally. The model satisfies all the ... the ... an antagonist modelling,
creating an ... monomeric dUTPase from an ancestral protein.

Figure 4.16 Modelling of the class II dUTPase structure from a class I template
The figures are hybrid models based on the original *E.coli* dUTPase 3D coordinates (a and b) plus
regions of diagrammatic predictions (c-e). See main text for details.

longer species like HSV-1 and VZV this loop would have to be extended. Given the location of the joining region on the opposite side from the active site, an extended loop could be efficiently accommodated without disrupting the active site or interfering with the conserved β-sheet arrangement.

This model satisfies the criteria given at the beginning of this section and gives a graphical interpretation of the structural consequences of an intragenic duplication of the class I dUTPase. The final model shown in (e) represents a monomer formed directly by the joining of two trimer subunits. It therefore possess a copy of the subunit structure in each half of the molecule and in this context can be seen as a visualisation of the secondary structure predictions in the previous section. For clarity this model does not highlight the extended sequence found between motifs 4 and 5 in the class II dUTPases. This is discussed in detail in Section 4.5.

4.4.6 Discussion

DNA and protein sequence data have the ability to highlight relationships between distantly related molecules. In the case of dUTPases these data has been utilised to propose the intragenic duplication which links the two distinct classes. The absence of any crystallographic data for the class II dUTPases prevented further investigation at the structural level. Utilising sequence analysis with computer prediction methods it was possible to continue the investigation into interclass relationships. Results from secondary structure prediction work allowed the generation of class II model which provides insight into the structural evolution of these molecules.

Secondary structure prediction was carried out using the ProteinPredict program. This allowed multiple sequence alignments to be used to predict an overall class II secondary structure. Various alignments were used to decrease any biasing effects. The first observation was the large quantity of predicted β-strands in the class II group. Alignment of these predictions with the *E.coli* doublet sequence demonstrated a high degree of congruency. It appears that each half of the class II molecule is a structural relative of the class I subunit structure.

A model was constructed accommodating all previous data plus the secondary structure predictions and demonstrates how the class I trimer could evolve into the class II monomer. The variable length joining region between the two halves of the class II molecule plus the extended region found between motifs 4 and 5 are explained structurally. The model satisfies all the criteria necessary for an intragenic duplication creating an active monomeric dUTPase from an ancestral trimer.

4.5   Hydrophobic modelling of HSV-1 dUTPase

4.5.1   Introduction

To test the class II model further an investigation into molecular surfaces was carried out. It was proposed that if the class II dUTPases assembled as predicted from the model then specific hydrophobic regions would be conserved between the two groups. Experiments were carried out to test this conservation firstly on the internal hydrophobic regions and secondly on the subunit interacting regions. The methodology was to identify conserved regions of hydrophobicity within the class I group, generate a data set of conserved residues and use this to probe the class II molecules.

4.5.2   Identification of class I hydrophobic regions

Several hydrophobic regions are proposed to be conserved throughout the class I dUTPases. This assumption is based on the extremely conserved class I subunit and trimer structures identified in previous sections. Two regions of potential hydrophobic interaction were investigated: the subunit core and the trimer core.

The subunit secondary structure is made up mainly of β-strands (Figure 4.13) and it is clear that these structural elements contribute to the folding of a stable tertiary structure. It is proposed that in addition to this mechanism there is a hydrophobic bias in the subunit core generating energetically favourable conditions for folding. It is also proposed that a similar hydrophobic bias exists in the trimer core and increases the efficiency of subunit assembly. This was examined using the *E.coli* crystal structure to identify hydrophobic residues within the subunit and trimer cores.

There are various definitions for measuring hydrophobicity based either on amino acid positions in known 3D structures (Janin, 1979; Rose *et al.*, 1985) or the specific physicochemical properties of their side chains (Wolfenden *et al.*, 1981; Kyte & Doolittle, 1982). The residues chosen here represent a compromise between these different scales. Four scales were used (two from each measurement method) and residues were chosen that were found to be the most hydrophobic in all four. The hydrophobic subset used for this investigation was therefore Ala, Cys, Leu, Gly, Ile, Met, Phe and Val.
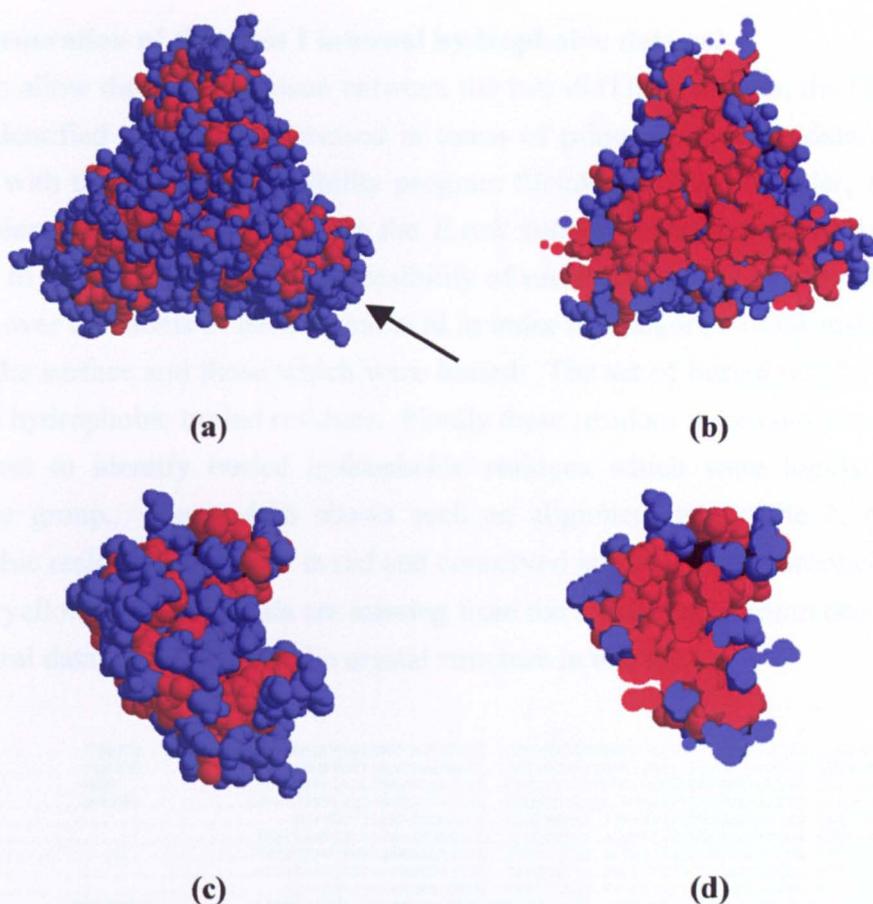
Figure 4.17 Distribution of hydrophobic residues in the *E.coli* trimer and subunit
Residues based on the defined hydrophobic subset are coloured in red. All other residues are coloured blue. (a) and (b) show the *E.coli* trimer as a surface view and a 50% slice through the middle respectively. (c) and (d) show the subunit as a surface view and a 50% slice through the middle respectively. The subunit is orientated with the C-terminal tail pointing away from the viewer. The angle of this view relative to the trimer is shown by an arrow in (a).

Molecular viewing programs permit efficient identification of internal hydrophobic residues. Each residue was coloured based on the above hydrophobic subset and slices were taken through the structure. Figure 4.17 gives a general indication of the distribution of hydrophobic residues in the *E.coli* trimer and subunit structures. The trimer is made up of a mixed outer surface with a strongly hydrophobic core (red). The subunit structure shows a similar pattern again with a mixed outer surface and a hydrophobic core. The subunit however has two regions of surface hydrophobicity which lie at the subunit-subunit interfaces and are shown in Figure 4.20.

4.5.3 Generation of the class I internal hydrophobic data set

To allow direct comparison between the two dUTPase classes, the hydrophobic regions identified had to be expressed in terms of primary sequence data. This was achieved with the surface accessibility program Sleuth (Dudek & Ponder, 1995). All water molecules were removed from the *E.coli* subunit coordinates and the program was used to calculate the surface accessibility of each atom in the structure. This was averaged over the atoms of each amino acid in order to categorise those residues which were on the surface and those which were buried. The set of buried residues was then refined to hydrophobic buried residues. Finally these residues were compared in a class I alignment to identify buried hydrophobic residues which were highly conserved within the group. Figure 4.18 shows such an alignment where the *E.coli* internal hydrophobic residues are shown in red and conserved inter-species hydrophobic loci are shown in yellow. Note that data are missing from the C-terminal portion due to the lack of structural data available from the crystal structure in this area.

```
                                            *            *   *           * *            *    * * *    * ***     *
Mpmv    ..............SLWGSQ..........LCSSQQKQPISKLTRATPGSAGLDL..CSTSHTVLTPEMGPQALSTGIYGPLPPNTFG.LILGRSSI.T
Srv1    ..............SLWGGQ..........LCSSQQKQPISKLTRATPGSAGLDL..SSTSHTVLTPEMGPQALSTGIYGPLPPNTFG.LILGRSSI.T
Jsrv    ..............MLWG...........NTAGSKRTIADLCRATRGSAGLDL..CATSYTVLTPEMGVQTLATGVFGPLPPGTVG.LLLGRSSA.S
Mmtv    ..............GVKGSG..........LNPEAPFFPIHDLPRGTPGSAGLDL..SSQKDLILSLEDGVSLVPTLVKGTLPEGTTG.LIIGRSS..N
Fiv     ...........................MIIEGDGILDKRSEDAGYDL..LAAKEIHLL.PGEVKVIPTGVKLMLPKGYWG.LIIGKSSI.G
Omvs    .........................SEIFLAREGSGILPKRAEDAGYDL..ICPQEVCIP.AGQVRKIPINLRINLKEDQWA.MVGTKSSF.A
Visn    .........................SEIFLAKEGRGILQKRAEDAGYDL..ICPQEISIP.AGQVKRIAIDLKINLKKDQWA.MIGTKSSF..
Caev    .........................SEIFLAKEGEGILPKREEDAGYDL..ICPEEVTIE.PGQVKCIPIELRLNLKKSQWA.MIATKSSM.A
Eiav    ........................MLAYQGTQIKEKRDEDAGFDL..CVPYDIMIP.VSDTKIIPTDVKIQVPPNSFG.WVTGKSSM..
Coxb    .........MTHSVQLK..........ILDKRLGSEFPLPAYATTGSAGLDLRACLDDEPLKIE.PDETCLISTGLAIYLGHSNVAATILPRSGL.G
Ecol    .........MMKKIDVK..........ILDPRVGKEFPLPTYATSSGSAGLDRRACLNDAVELA.PGDTTLVPTGLAIHIADPSLAAMMLPRSGL.G
Vaccc   ...........MFNMNINSP..........VRFVKETNRAKSPTRQSPGAAGYDL..YSAYDYTIP.PGERQLIKTDISMSMPKICYG.RIAPRSGL.S
Varv    ...........MFNMNINSP..........VRFVKETNRAKSPTRQSPYAAGYDL..YSAYDYTIP.PGERQLIKTDISMSMPKFCYG.RIAPRSGL.S
Human   ..............MQ..........LRFARLSEHATAPTRGSARAAGYDL..YSAYDYTIP.PMEKAVVKTDIQIALPSGCYG.RVAPRSGL.A
Orfv    ...........MEFCHTET..........LQVVRLSQNATIPARGSPGAAGLDL..CSAYDCVIP.SHCSRVVFTDLLIKPPSGCYG.RIAPRSG..A
Fadv    ......MDPFGSSSVPPCSTSDLPEPKLYFVRLSPHAVPPVRATHGAAGYDL..FSAYDIKVP.ARGRALVPTDLVFQFPPGCYG.RIAPRSGL.A
Calb    ............MTSEDQSLKKQKLESTQSLKVYLRSPKGKVPTKGSALAAGYDL..YSAEAATIP.AHGQGLVSTDISIIVPIGTYG.RVAPRSGL.A
Scer    .............MTATSDKV..........LNIQLRSASATVPTKGSATAAGYDI..YASQDTITIP.MAQGQMVSTDISFTVPVGTYG.RIAPRSGL.A
Lyces   MAENQINSPEITEPSPKVQKLDHPENGNVPFFRVKKLSENAVLPSRASSLAAGYDL..SSAAETKVP.ARGKALVPTDLSIAVPQGTYA.RIAPRSGL.A
Ccv     MDFNTDRGSAIPTGDDRGYQCGVIGDSVRFSVFTGYDAADVSIPKISSPGSAGFDL..SVLEDREFI.RGCHYRLPTGLAIAVPRGYVG.IITPRSS..Q
                                                                     ----                                        ----
                                                                   Motif 1                                     Motif 2

        **        * *          * * *            *   *   * ** **** *
Mpmv    MK.GLQVYP..GVIDNDYTGEIKIMA.KAVNNIVTVSQGNRIAQLILLPLIETDNKV.......QQPYRGQGSFGSSDI...................
Srv1    IK.GLQVYP..GVIDNDYTGEIKIMA.KAVNNIVTVPQGNRIAQLILLPLIETDNKV.......QQPYRGQGSFGSSDIYW................
Jsrv    LK.GILIHP..GVIDSDYTGEIKILA.SAPNKIIVINAGQRIAQLLLVPLVIQGKTI.......NRDRQDKGFGSSDAYW................
Mmtv    YKKGLEVLP..GVIDSDFQGEIKVMV.KAAKNAVIIHKGERIAQLILLPYLKLPNPVI......KEERGSGFGGPSPVHHW...............
Fiv     SK.GLDVLG..GVIDEGYRGEIGVIMINVSRKSITLMERQKIAQLIILPCKH.EVLEQG..KVVMDSERGDNGYGSTGVFSSW.............
Omvs    SK.GVFVQG..GIIDSGYQGIIQVVVYNSNDKEVIIPQGRKFAQLILMPLIH.EDLEAW..GETRRTERGNQGFGSTGAY..............
Visn    ANKGVFVQG..GIIDSGYQGTIQVVIYNSNNKEVVIPQGRKFAQLILMPLIH.EELEPW..GETRKTERGEQGFGSTGMYW.............
Caev    AK.GVFTQG..GIIDSGYQGQIQVIMYNSNKIAVVIPQGRKFAQLILMDKKH.GKLEPW..GESRKTERGEKGFGSTGMY..............
Eiav    AKQGLLING..GIIDEGYTGEIQVICTNIGKSNIKLIEGQKFAQLIILQHHS.NSRQPW..DENKISQRGDKGFGSTGVFW.............
Coxb    HKHGIVLGNLVGLIDSDYQGPLMVSCWNRGKEPYTINPGDRIALVVLPILK.AQFAVV..EEFELTERGAGGFGSSGQN............
Ecol    HKHGIVLGNLVGLIDSDYQGQLMISVWNRGQDSFTKQPGERIAQMIFVPVVQ.AEFNLV..EDFDATDRGEGGFGHSGRQ............
Vaccc   LK.GIDIGG..GVIDEDYRGNIGVILINNGKCTFNVNTGDRIAQLIYQRIYY.PELEEV..QSLDSTNRDGDQGFGSTGLR............
Varv    LK.GIDIGG..GVIDEDYRGNIGVILINNGKYTFNVNTGDRIAQLIYQRIYY.PELKEV..QSLDSTDRGDQGFGSTGLR............
Human   AKHFIDVGA..GVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIAQLICERIFY.PEIEEA..QALDDTERGSGGFGSTGKN............
Orfv    VKHFIDVGA..GVIDEDYRGNVGVVLFNFGNSDFEVKKGDRIAQLICERISC.PAVQEV..NCLDNTDRGDSGFGSTGSGACGGRDTAWYIS...
Fadv    AKFFIDVGA..GVIDPDYRGNVSVVLFNFSESSFNIRRGDRVAQLILERIMV.PELSEL..TQLGETDRGASGFGFTGMGAVDRNQRSVLEWLTPGSR
Calb    VKHGISTGA..GVIDADYRGEVKVVLFNHSEKDFEIKEGDRIAQLVLEQIVN.ADIKEISLEELDNTERGEGGFGSTGKN............
Scer    VKNGIQTGA..GVVDRDYTGEVKVVLFNHSQRDFAIKKGDRIAQLILEKIVDDAQIVVV..DSLEESARGAGGFGSTGV............
Lyces   WKYSIDVGA..GVIDADYRGPVGVVLFNHSEVDFEVKVGDRIAQLIVQKIVT.PEVEQV..DDLDSTVRGSGGFGSTGV............
Ccv     AKNFVST....GIIDSDFRGHIHIMV..SAIADFSVKKNQRIAQLVVTPCLTQSEVVPY..ETLERTRRGTGGFGSSGQ.............
        ---------       --------              ----------
        Motif 3              Motif 4                  Motif 5
```

Figure 4.18 Conservation of internal hydrophobic loci within the class I group
Internal hydrophobic residues in the *E.coli* structure are highlighted in red. Conserved hydrophobic residues in the other class I species are highlighted in yellow. Strongly conserved hydrophobic loci are marked with an asterisk (*).

It can be seen that internal hydrophobic loci are highly conserved within the class I sequences. This is especially interesting since many of these loci lie outside the conserved motif regions. Motifs 1 and 4 however contain a large percentage of internal hydrophobic residues. This is consistent with motif 4 having only one residue in direct contact with dUDP in the co-crystal (Q119) and motif 1 having no direct substrate interactions. I propose that these internal residues are conserved for a structural role rather than substrate binding or catalysis. Given their relative position it is likely that these motif residues stabilise the condensation of motifs from different subunits. Since the class I dUTPases are the only trimeric enzymes known to date in which the active site is made up from 3 different subunits it is not surprising that internal residues are conserved to maintain this unusual coordination.

Despite the variation in inter-motif primary sequences there is a high degree of hydrophobic conservation. This is consistent with the homologous nature of the class I structures. The most highly conserved loci are highlighted on the alignment with an asterisk (Figure 4.18) and it is these positions that were used to probe the class II molecules.

4.5.4   Comparison of class I and II subunit hydrophobic cores

The internal hydrophobic loci identified in the class I group were compared directly to the class II group with the standard *E.coli* doublet alignment used for secondary structure comparisons. A diagrammatic representation of this alignment is shown in Figure 4.19. The positions of the hydrophobic residues was obtained by superimposing the data form Figure 4.18 onto the alignment shown in Figure 4.15.

E. coli doublet sequence

Alignment of 10 class II sequences
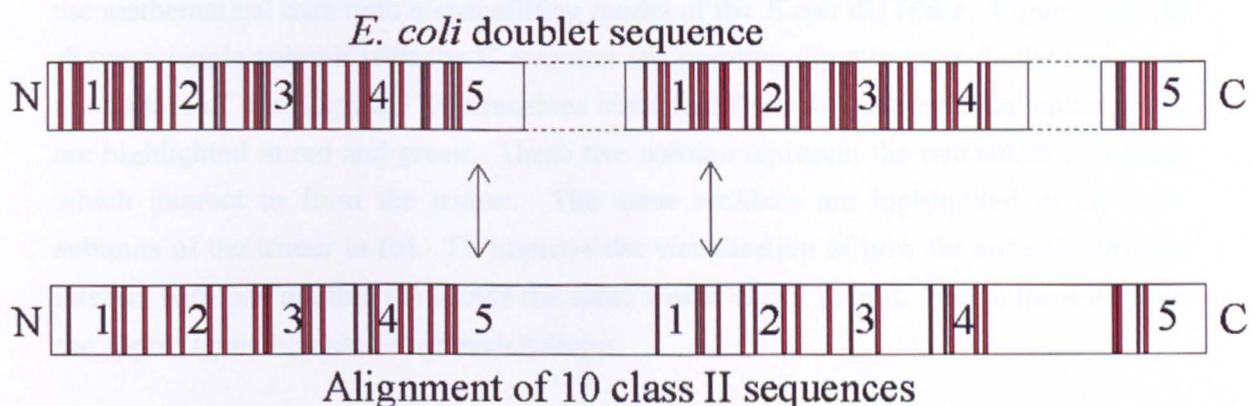
Figure 4.19 Comparison of the class I hydrophobic loci to the class II sequences
A diagrammatic representation of the hydrophobic sequence alignment is shown (not to exact scale). The *E.coli* doublet sequence is represented as a box with the position of each motif shown. The 31 class I conserved hydrophobic loci (identified with an asterisk in Figure 4.18) are represented by red bars. These were aligned with 10 class II sequences and the position of congruent hydrophobic loci is shown in the bottom diagram (22/31 in the N-terminal half and 20/31 in the C-terminal half).

A positive hit was defined as an *E.coli* locus which had 5 or more conserved hydrophobic loci out of the ten species in the class II lineup. There was 31 *E.coli* loci aligned with each half of the class II lineup. In the C-terminal half there was 20 positive hits (64%) while in the N-terminal half there was 22 positive hits (71%). This yields an average of over 67% positive hits over the entire class II lineup.

This gives an indication that internal hydrophobic loci are well conserved between class I and II dUTPases. More significantly it demonstrates that each half of the class II structure has a similar hydrophobic distribution to the class I subunit structure. It follows that each half of the class II molecule has an internal hydrophobic core corresponding to the class I subunits core. This supports the class II model supplying further evidence that the class II monomer approximates a condensation of two class I subunit structures.

4.5.5   Identification of the class I interface hydrophobic data set

It is not only the subunit core which has an internal bias for hydrophobic residues but also the trimer itself (Figure 4.17b). The internal core of the trimer is partly composed of hydrophobic subunit surfaces. These surfaces can be readily identified mathematically by defining their accessibility in two coordinate data sets. First the surface is calculated for a single subunit. The surface is then calculated from the entire trimer. Subtraction of the subunit surfaces from the trimer surface yields the internal subunit interfaces. It is these surfaces which allow subunit-subunit interaction in the formation of the trimer.

To allow graphical interpretation of these interfaces it was useful to transpose the mathematical data onto a spacefilling model of the *E.coli* dUTPase. Figure 4.20 (a) shows a single subunit with the C-terminal tail pointing directly towards the viewer at the bottom of the diagram. The residues identified from the mathematical calculations are highlighted in red and green. These two colours represent the two subunit surfaces which interact to form the trimer. The same residues are highlighted in all three subunits of the trimer in (b). To improve the visualisation of how the subunit surfaces interact with one another (c) shows the same trimer sliced in half. White lines indicate the approximate boundaries of each subunit.

(a)

(b)

(c)

(d)
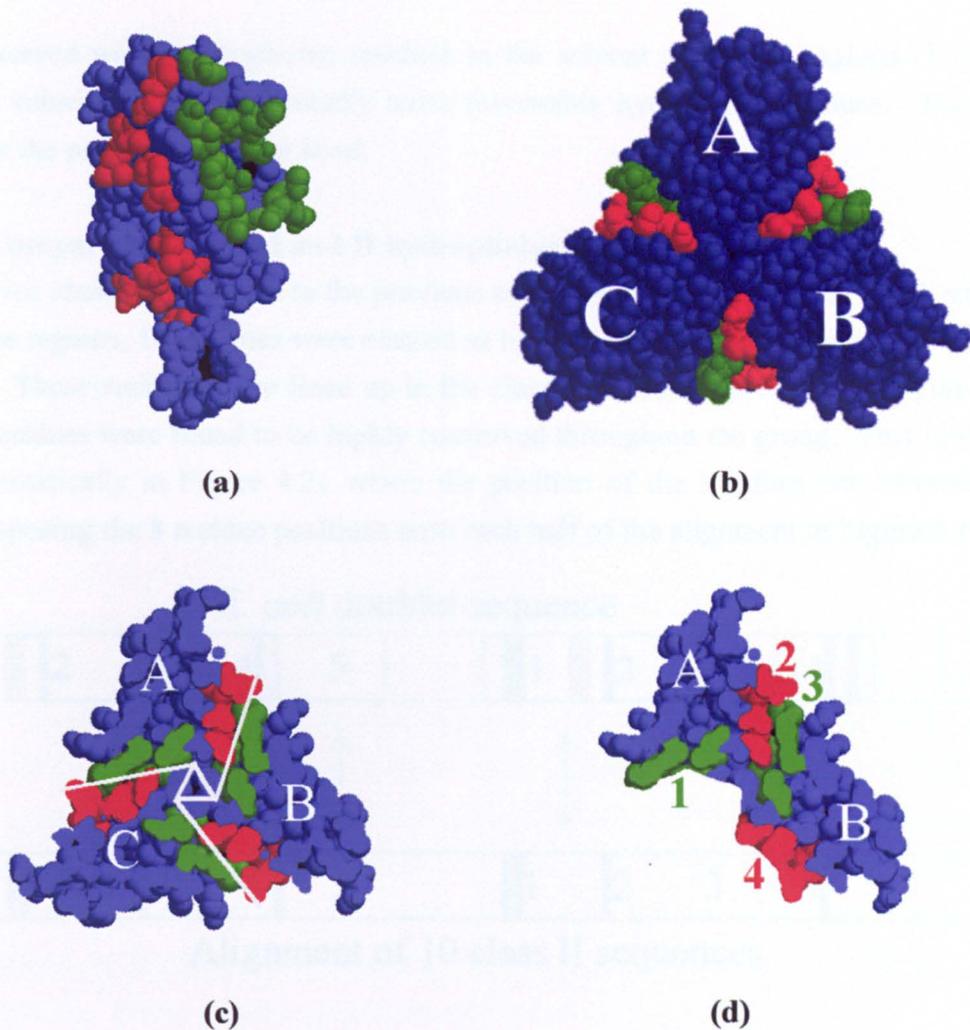
Figure 4.20 Identification of *E.coli* subunit interfaces

Spacefilling diagrams are shown for the *E.coli* subunit (a), trimer (b) and trimer 50% slice (c). The class II model is shown in (d). Interacting surfaces are shown in red and green. Diagrams were created with RasMol. See text for details.

It can be seen that each subunit possesses two major interacting surfaces coloured red and green. What is interesting is the fate of these surfaces in the class II molecule. Referring back to the model in Figure 4.16, the class II molecule is composed of two class I subunits joined in one chain. This model is represented as a spacefilling diagram in Figure 4.20 (d). One of the three trimer subunits has been removed to represent the class I monomer. Subunit 'A' represents the N-terminal half of the protein while subunit 'B' represents the C-terminal half.

What becomes apparent is that while three subunit interfaces are required to generate the trimer, only one subunit interface is required to generate the monomer. The N-terminal half requires only the red surface (2) while the C-terminal half requires only the green surface (3). The remaining two surfaces (1 and 4) are now in solvent contact. It is proposed that hydrophobic residues in the interacting region (2 and 3) will

be conserved while hydrophobic residues in the solvent accessible regions (1 and 4) will be substituted by energetically more favourable hydrophilic residues. This was tested at the primary sequence level.

4.5.6 Comparison of class I and II hydrophobic subunit interfaces

An identical approach to the previous section was used. From the *E.coli* subunit interface regions, 10 residues were classed as hydrophobic using the previously defined subset. These residues were lined up in the class I alignment as before in Figure 4.18 and 8 residues were found to be highly conserved throughout the group. This is shown diagrammatically in Figure 4.21 where the position of the residues was obtained by superimposing the 8 residue positions onto each half of the alignment in Figure 4.15.
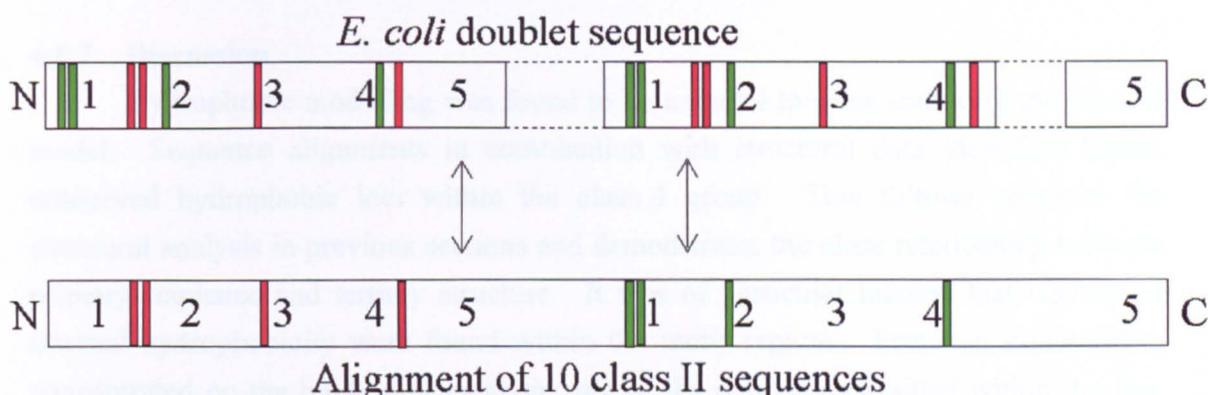
Figure 4.21 Comparison of hydrophobic interfaces at the primary sequence level
Alignment of the class I sequences identified eight highly conserved hydrophobic loci at the subunit interfaces. Four loci were found for each subunit surface and are coloured either red or green for comparison to Figure 4.20. The diagrammatic alignment shows that only loci from the red surface are conserved in the N-terminal half of the class II sequences while only loci from the green surface are conserved in the C-terminal half. This is proposed to indicate the conservation of hydrophobic surfaces 2 and 3 in Figure 4.20(d) and the loss of hydrophobic surfaces 1 and 4.

Four residues were identified on the red surface and four on the green surface. These residues were highlighted in the *E.coli* doublet sequence and aligned with the class II dUTPase sequences as before. A positive hit was defined as an *E.coli* hydrophobic locus which had 6 or more homologous loci out of the ten species in the class II lineup. The following results refer to Figures 4.20(d) and 4.21.

Surface	Colour in figure	Donor subunit	Class II region	Conservation
1	Green	A	N-terminal half	0/4 = 0%
2	Red	A	N-terminal half	4/4 = 100%
3	Green	B	C-terminal half	4/4 = 100%
4	Red	B	C-terminal half	0/4 = 0%

The data shown above support the theory that only those hydrophobic loci which are situated on the class II subunit interface (2 and 3) are conserved between the classes. The remaining surfaces (1 and 4) are solvent exposed in the class II model and are not necessary for structural interactions. Caution must be used with this interpretation of the results given the small data set available: only 8 hydrophobic loci are highly conserved throughout the class I group. In order to put these data in perspective the probability of an *E.coli* probe locus matching a random hydrophobic locus in the class II alignment was calculated. From 388 loci in the alignment 85 were classed as hydrophobic (conserved between at least 6 species). The chance of matching a hydrophobic locus by chance is just below 22%. This was deemed a reasonable background error.

4.5.7 Discussion

Hydrophobic modelling was found to be a useful tool for analysing the class II model. Sequence alignments in combination with structural data identified highly conserved hydrophobic loci within the class I group. This follows on from the structural analysis in previous sections and demonstrates the close relationship between primary sequence and tertiary structure. It was of particular interest that regions of internal hydrophobicity were found within the motif regions. Previous studies have concentrated on the binding or catalytic role of the residues contained within the five motif regions. These studies support an additional structural role for the motifs in the consolidation of three subunit surfaces to form each active site.

It is clear that there are interacting surfaces between the main body of the trimer and the extended tail regions. Small alterations in the position of the tail would rearrange hydrophobic loci in the primary sequence to a high degree and this was judged to be incompatible with the methods used. Only the globular regions of the subunits were therefore analysed. Comparison of the internal hydrophobic residues between the classes demonstrated a high degree of congruency. This supports the view that the class II molecules approximates two class I subunits joined in a single chain. It appears that the hydrophobic core supporting each subunit in the trimer has been conserved in duplicate in the class II enzymes.

Analysis of subunits within the trimer identified two surfaces containing conserved hydrophobic residues. Using these loci to probe the class II molecules a distribution pattern was revealed which again supports the class II model. These studies have the same fundamental constraints as the secondary structure predictions, namely the accuracy of inter-class primary sequence alignments. Although this analysis uses the same underlying comparisons as the secondary structure predictions the methods used are sufficiently divergent to strengthen support for the class II model.

4.6   Structural evolution of the class II dUTPases

4.6.1   Introduction

Previous sections have addressed the relationships between the class I and II dUTPases both at the primary sequence level and at the structural level. Although secondary structure prediction and hydrophobic modelling support the class II model it is also necessary to assess evolutionary relationships. The class II model is based on the intragenic duplication model proposed by D.J. McGeoch (1990b) plus the structural modelling data presented here. It is clear from the model that not only must the class I subunit gene be duplicated but also there must be a structural rearrangement to produce an active enzyme. It is therefore possible that a functional evolutionary intermediate existed between the class II monomer and the class I trimeric ancestor.

The class II model shown in Figure 4.16 demonstrates that a single active site cavity can be formed from the joining of two subunit structures. This cavity produced from motif 3 in the N-terminus and motifs 1, 2, and 4 in the C-terminus constitutes the majority of a functional active site. Now that structural data has become available for the human dUTPase C-terminal region it seems likely that motif 5 is necessary for catalytic function in the class I enzymes. A homologous situation for the class II enzymes is supported by HSV-1 mutagenesis data in the following chapter. Based on this finding an intragenic duplication cannot, by itself, create an active class II monomer. As can be seen in the class II model (Figure 4.16) an extension chain is necessary between motifs 4 and 5 to allow the correct positioning of motif 5 over the active site. The result of this structural constraint on the evolution of the class II dUTPases was investigated.

4.6.2 Maintenance of enzyme function during class II evolution

Chapter 1 introduced the role of the dUTPase enzyme demonstrating its importance in a variety of biological life cycles (Section 1.4.4). Clearly an intermediate between the class I and class II structures which lacks function could compromise an organism that is required to generate dUTPase activity. In short, the evolution of the class II dUTPase must be realised without the loss of dUTPase function at any stage. A model was generated which successfully meets this criterion.

Figure 4.22 demonstrates the genetic rearrangements necessary to generate a class II dUTPase. Each stage is set out with a genetic arrangement and a corresponding structure in order that all intermediates may be easily visualised. The evolutionary ancestor of the class II molecule is proposed to be structurally homologous to the present class I dUTPases.
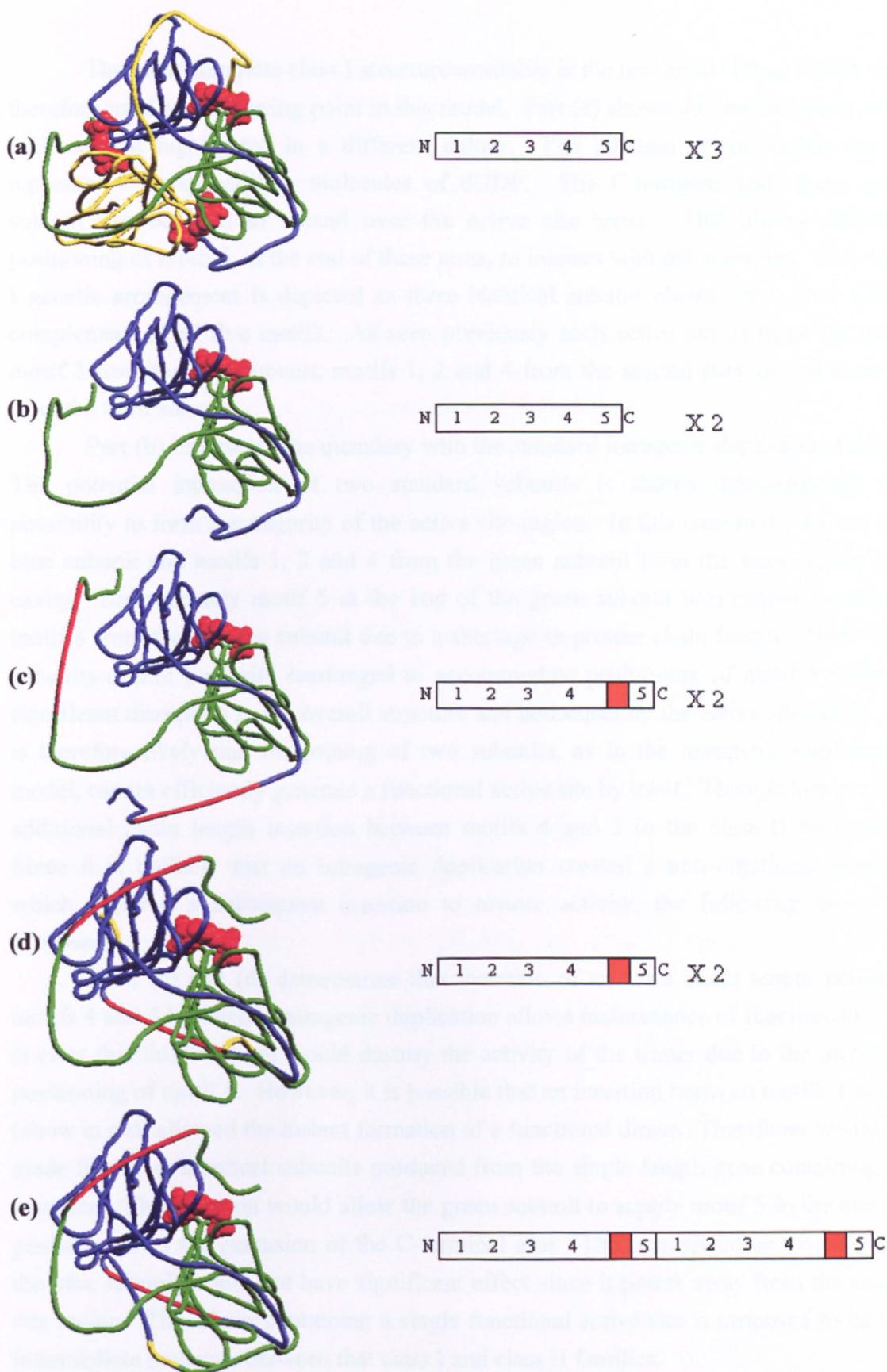
Figure 4.22 Structural and genetic evolution of the class II dUTPase
The left hand diagrams represent the structural consequences of the corresponding genetic rearrangements depicted on the right. The quantity of each protein chain required to construct the structure is indicated to the right of diagrams (a-d). The final diagram (e) represents the class II monomer as a single chain structure. Details are given in text.

The most complete class I structure available is the human dUTPase which was therefore used as the starting point in this model. Part (a) shows the human trimer with each subunit highlighted in a different colour. The position of the active site is represented by spacefilling molecules of dUDP. The C-terminal arms from each subunit can be seen to extend over the active site areas. This allows efficient positioning of motif 5, at the end of these arms, to interact with the substrate. The class I genetic arrangement is depicted as three identical subunit chains, each with a full complement of all five motifs. As seen previously each active site is made up from motif 3 from the first subunit, motifs 1, 2 and 4 from the second subunit and motif 5 from the third subunit.

Part (b) represents the quandary with the standard intragenic duplication theory. The potential interaction of two standard subunits is shown demonstrating the possibility to form the majority of the active site region. In this case motif 3 from the blue subunit and motifs 1, 2 and 4 from the green subunit form the basic active site cavity. Unfortunately motif 5 at the end of the green subunit arm cannot substitute motif 5 from the missing subunit due to a shortage in protein chain length. These two subunits cannot be easily rearranged to accommodate positioning of motif 5 without significant disruption to the overall structure and consequently the active site cavity. It is therefore likely that the joining of two subunits, as in the intragenic duplication model, cannot efficiently generate a functional active site by itself. There is however an additional chain length insertion between motifs 4 and 5 in the class II molecules. Since it is unlikely that an intragenic duplication created a non-functional enzyme which required a subsequent insertion to restore activity, the following theory is proposed.

Parts (c) and (d) demonstrate that insertion of an extra chain length between motifs 4 and 5 before the intragenic duplication allows maintenance of functionality. It is clear that this insertion would destroy the activity of the trimer due to the incorrect positioning of motif 5. However, it is possible that an insertion between motifs 4 and 5 (show in red) allowed the instant formation of a functional dimer. This dimer would be made from two identical subunits produced from the single length gene containing an insertion. The insertion would allow the green subunit to supply motif 5 in the correct position due to the extension of the C-terminal arm. The corresponding extension in the blue subunit would not have significant effect since it points away from the active site region. This dimer containing a single functional active site is proposed to be the intermediate molecule between that class I and class II families.

An intragenic duplication at this stage effectively converts the functional dimer into a functional double length monomer. This is shown in part (e) where the joining region has been teased out for clarity. Minimal structural rearrangement is necessary to

link the two subunits since additional chain length has already been supplied by the duplication of the insertion region. Conservation pressure is no longer maintained for the additional copies of the motifs which do not form part of an active site. In terms of primary sequence analysis these extra motifs are seen to be lost although a class II specific motif has replaced the C-terminal copy of motif 3.

4.6.3 Discussion

It order for the class II model to be viable, a mechanism had to be demonstrated showing the genetic and structural evolution from the class I ancestor. This model was constrained by the fact that any structural rearrangement from intragenic duplication must result in a functional dUTPase. In the proposed model an insertion between motifs 4 and 5 destroys trimer activity but allows the assembly of a functional dimer. This dimer structure is proposed to be the functional intermediate between the class I and class II families.

The reason why a dimeric dUTPase has not been discovered may be due to the efficiency of subunit assembly. As shown from the evolutionary model, the genetic rearrangement necessary to generate a double length monomer from the dimer intermediate is a simple intragenic duplication. This allows a single gene to produce a single protein chain without the need for subsequent subunit assembly. It may be that this method of enzyme construction is energetically more favourable since the production of single subunits does not guarantee functionality. The question may then be asked: why is the trimeric version seen so widely in nature? Subunit assembly in the trimer generates three functional active sites whereas dimer assembly only generates a single active site. Clearly a trimeric dUTPase has a lower energetic cost to the organism per active site.

The final question is the most difficult to answer: why does the herpesvirus family utilise a double length monomer and not the class I trimer? From an energetic standpoint it appears that the trimer is more efficient. The trimer produces 3 active sites per three subunits compared to the monomers equivalent of one active site per two subunits. It may be that subunit assembly is so inefficient in the herpesvirus host environment that production of a single folding protein allows large scale enzyme production in a shorter time period. It may be that given the complex nature of the trimer organisation with all three subunits contributing to each active site that even a slight error in one of the subunits folding renders the entire trimer useless. There may be another reason however. The next section investigates the possibility that the class II monomer produces a secondary element which may endow a secondary accessory function.

4.7 Investigation of a class II specific motif

4.7.1 Introduction

Sequence analysis of the class II dUTPases revealed a C-terminal conserved region occupying the corresponding position of the class I motif 3. This conserved region was named motif X and represents the only major region of primary sequence conservation specific to the class II dUTPases. Although no function has been assigned to motif X as yet, its potential significance is discussed below.

4.7.2 Sequence analysis of the class II motif X

The motif X region of the class II dUTPases occupies the corresponding position of motif 3 in the class I group. In the intragenic duplication model the original copy of motif 3 is lost as it no longer constitutes part of the class II active site. This is best visualised in Figure 4.10 (Section 4.3.3). One of the intriguing properties of motif X is that it is found not only in the α- and γ-herpesvirus dUTPases but also in the β-herpesvirus dUTPase homologues. These homologues share little sequence conservation with the class II dUTPases and possess no convincing copies of motifs 1-5. Although there are small areas of local conservation (typically no more than two consecutive residues) these homologues are not thought to be functional dUTPases. Their classification in the Swiss-Prot database as putative dUTPases is based on the congruent position of their ORF in the β-herpesvirus genome compared to α- and γ-herpesvirus. Figure 4.23 shows the motif X region in both α- and γ-herpesvirus aligned with four β-herpesvirus species.

```
SS Predict  E      EEEEEE          EEEEEE       EEEE    EEEEE
(γ)  EBV    GRSGLAMQGILVKPCRWRRGGV.DVSLTNFSDQTVFLNKYRRFCQLV
(γ)  EHV-2  GRSGLATRGVVVDVVKWTHSPL.TLKIYNFTDNTLRYSAGTRICQVV
(γ)  HVS    GRSGLAAKGLTIKVSRW.QNQL.QIIFYNYTKSQITYTARTRIAQVV
(α)  HSV-1  GRSSLNARGLLVVPTRWLPGHVCAFVVYNLTGVPVTLEAGAKVAQLL
(α)  HSV-2  GRSSLNARGLLVMPTRWPSGHACAFVVCNLTGVPVTLQAGSKVAQLL
(α)  EHV-1  GRSSMNRRGLIVLPTRWVAGRTCCFFILNVNKYPVSITKGQRVAQLL
(α)  EHV-4  GRSSMNRRGLIVLPTRWVTGRTCCFFILNINKYPVYITKGQRVAQLL
(α)  BHV-1  GRSSCNLRGLIVLPTAWPPGEPCRFVLRNVTQEPLVAAAGQRVAQLL
(α)  PRV    GRSSLNLRGIVVFPTPWESG.PCRFRIQNRGAHPVTLESGQRVAQLV
(α)  VZV    GRSSKSRSGIIVCPTAWPAGEHCKFYVYNLTGDDIRIKTGDRLAQVL
               Motif 2          Motif X              Motif 4
(β)  HCMV   GVRQFSQSDLIIRPTIWLPGTAAGVTVVNTSHTTVCISPHTTVAKAV
(β)  MCMV   PCRHLATKRVLLDPTVWRPNSLAVLRVLNASDEHVDLEAGMAMAKII
(β)  HHV-6  PSKEIAKLGLLIETYIWNKDTIPSIKIFNSTRKTIYIPTGICIARII
(β)  HHV-7  ANKEILCHGLVVETNIWLKNTTPSVKIFNPTSQRIFVQAGICIATII
```

Figure 4.23 Class II alignment showing relative position of motif X
Ten class II dUTPases from the α- and γ-herpesvirus are aligned on the top of the diagram with four β-herpesvirus dUTPase homologues on the bottom. The position of motif X (yellow) is shown relative to motifs 2 and 4 (cyan). Residues completely conserved in the α- and γ- herpesviruses are indicated in green. Note that the β-herpesvirus species do not possess full copies of the flanking motifs 2 and 4 although there are small regions of sequence conservation. Secondary structure predictions from the ProteinPredict program are shown at the top of the alignment ('E' represents a predicted β-strand).

The motif X region contains several highly conserved residues (highlighted in green) which are also reasonably well conserved in the β-herpesvirus homologues. The majority of the other non-conserved loci in each motif X have largely similar residue substitutions. This is true of both the α- and γ-herpesvirus dUTPases and also the β-herpesvirus dUTPase homologues.

4.7.3   Secondary structure prediction of the class II motif X

The fact that the class I dUTPases do not possess a motif X is intriguing. It opens up the possibility that the class II dUTPases share a new structure which may endow them with a novel secondary function. This theory is compounded by the finding that a group of β-herpesvirus proteins share only this motif X region and not the other 5 motifs which constitute the dUTPase active site. It is conceivable that these β-herpesvirus proteins share only a function supported by motif X.

Experimental work is required to test the function of the β-herpesvirus proteins and any potential role of the conserved motif X region. In the class II dUTPases, motif X can be mapped on various models to yield structural predictions. Although these predictions cannot identify functional significance they do reveal some interesting points. Several prediction methods were employed including ProteinPredict, Threader and Swiss-Modeller.

The class II motif X region was initially analysed with the ProteinPredict program during the study of the class II β-strand predictions (Sections 4.4.3 & 4.4.4). The secondary structure predictions for this region are shown in Figure 4.23 at the top of the class II alignment. Motif X is predicted to contain two β-strands with the conserved Trp residues close to the middle of the joining sequence. This is interesting because it resembles the class I motif 3 structure which is composed similarly of two β-strands with a highly conserved Tyr in the middle of the joining sequence. It may not be surprising that motif X is similar to motif 3 in overall structure given that motif 3 originally occupied this position after the initial intragenic duplication. To further test this theory, more detailed modelling approaches were used.

4.7.4   Structural fold recognition by protein sequence threading

The program Threader (Jones *et al.*, 1992) was used to predict the local structure of the class II motif X region. Protein threading employs a different methodology than the previously used ProteinPredict program. Basically a library of unique protein folds is obtained from the database of known protein structures. 266 folds were present in the library used for these predictions. These folds are defined as 3D chains and the original primary sequences are ignored. The test sequence, in this case motif X, is then optimally fitted or 'threaded' onto each 3D fold in the library. The

energy of each possible fit is then calculated by the addition of each pairwise interaction between the test sequence and each known fold. These energy or 'threading' values are then ranked in order. The library fold with the lowest energy (easiest fit) is predicted to be the closest match to the test sequence. Threader gives each fold a pairwise energy score (Z) which is based on the following scale:

Z - Score	Explanation
Z < -3.5	Very significant – probably a correct prediction
Z < -3.0	Significant – good chance of being correct
-2.7 < Z < -3.0	Borderline – possibly correct
-2.0 < Z < -2.7	Poor – needs other confirmation
Z > -2.0	Very poor – probably no suitable folds in the library

Table 4.2 Interpretation of Threader Z-scores
Each structure which is threaded against the test structure is given a Z-score which relates to the significance of the match. Taken from the Threader user notes (D.T. Jones, 1994).

At the time of using Threader there were no dUTPase structures present in the fold library. Both the *E.coli* and HSV-1 motif 3 regions were both analysed by Threader in order to determine how closely they matched the known *E.coli* structure. Threader predicted a loop structure for motif 3 with two flanking β-strands which is essentially the correct structure for the class I motif 3 region. The closest match was a β-loop-β region of a fatty acid binding protein (1MDC) of the Tobacco Hornworm (*Manduca sexta*). The Z-score for the 1MDC loop as a structural homologue motif 3 was –3.69. This score is extremely low and represents a highly significant Threader prediction. This was encouraging since the program did not have a dUTPase motif 3 loop in its fold library.

(a)   **(b)**
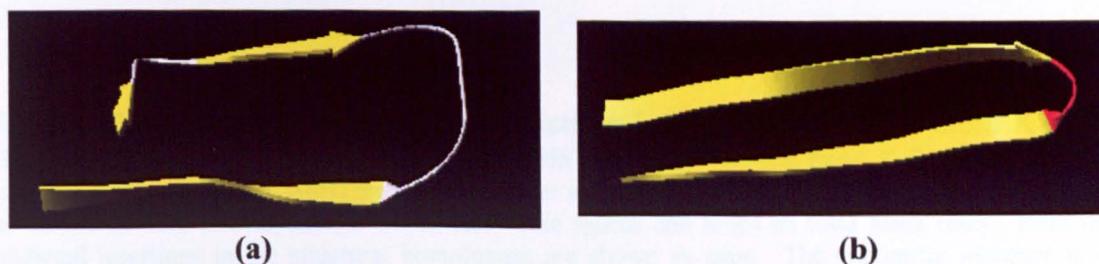
Figure 4.24 Comparison of *E.coli* motif 3 structure to Threader prediction
The structure of the *E.coli* motif 3 structure is shown in (a). Figure (b) shows the Threader predicted structural homologue, a β-loop-β structure in the fatty acid binding protein 1MDC. β-strands are shown as yellow arrows. The red region in the 1MDC structure indicates where Threader predicts an insertion of two residues. This insertion would increase the similarity between (a) and (b).

Note that the 1MDC structural homologue is predicted to have a two residue insertion in the loop region (shown in red in Figure 4.24). This insertion would essentially lever the two β-strands apart and generate a much more homologous structure to motif 3 than that shown in the figure. Although there is no functional relationship between motif 3 and the predicted homologue, the structural relationship is clear.

Since reasonable data were obtained from the motif 3 predictions, Threader was subsequently used to predict the structure of the motif X region of HSV-1. The closest related fold to the motif X test sequence was again 1MDC. The Z-score for this structural homologue was –2.84. This score is inferior to the prediction for the motif 3 loop and represents only borderline significance. However, when alignments are made of the top ten Z-scores for both HSV-1 motif 3 and motif X, the consensus predictions are highly similar. This is seen in Figure 4.25 where the best threaded homologues are given in descending order from top to bottom.

(a) Motif 3   (b) Motif X
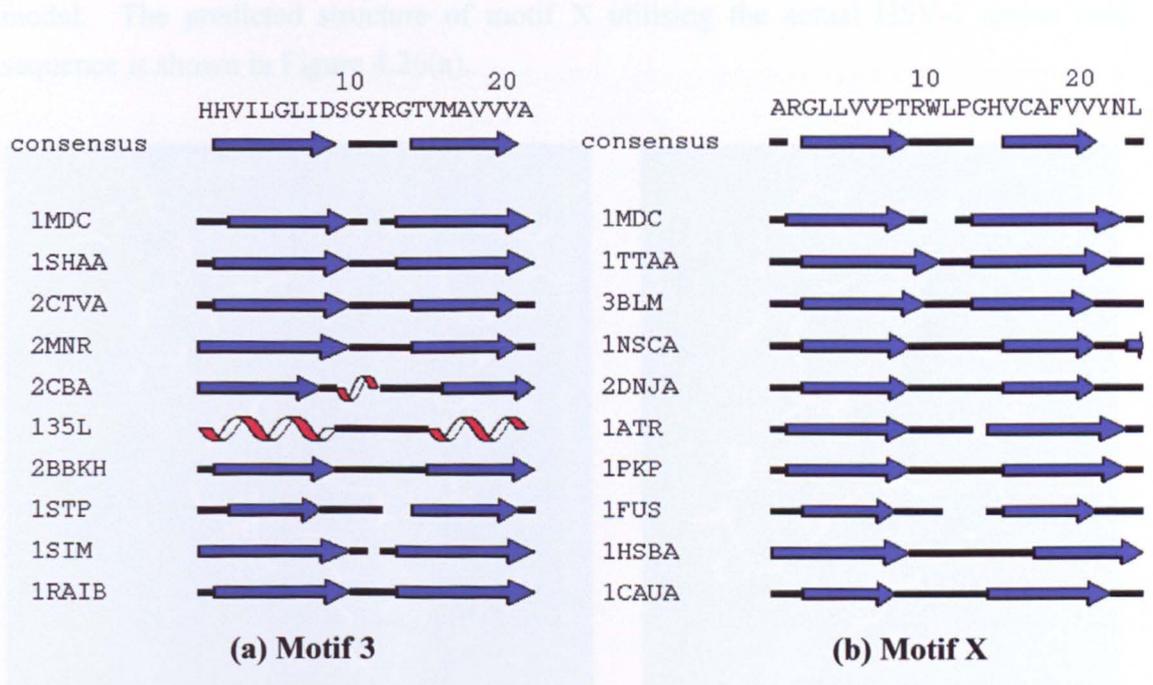
Figure 4.25 Comparison of Threader predictions for HSV-1 motif 3 and motif X
The ten best Threader predictions are shown for HSV-1 motif 3 (a) and motif X (b). The structural homologues (PDB codes at right) are ranked in order with the best (lowest) Z-score at the top. β-strands are shown as blue arrows, helices as red and white spirals and loops as solid black lines. Note that predicted insertions in the structural homologues are shown as gaps. The consensus sequence is an unweighted compilation of the top ten homologous folds.

Although the best Threader prediction for HSV-1 motif X is not as good as motif 3, the comparison of the consensus results plus the previously determined ProteinPredict results suggests that the two structures may be related. In order to

generate a 3D structure based on the primary sequence (as opposed to a structural homologue) a new modelling approach was used.

4.7.5 Three dimensional modelling of HSV-1 motif X

The previous data demonstrate the possibility that the class II motif X may have evolved from the ancestral motif 3 and has retained a similar basic structure. In order to achieve a more direct relationship between the HSV-1 primary sequence and the predicted structure a local model was created for motif X. This work was done in collaboration with N. Arbuckle in the MRC Virology Unit. The HSV-1 motif X sequence was used to probe the PDB database for primary sequence homologues using the TBLASTX program (Altschul *et al.*, 1990). The top five structures corresponding to the primary sequence matches were used as input to the modelling program, ProMod (Peitsch, 1995; Peitsch, 1996). This program basically creates an average structure based on the structures with the best primary sequence similarity and produces a 3D model. The predicted structure of motif X utilising the actual HSV-1 amino acid sequence is shown in Figure 4.26(a).

(a)   (b)
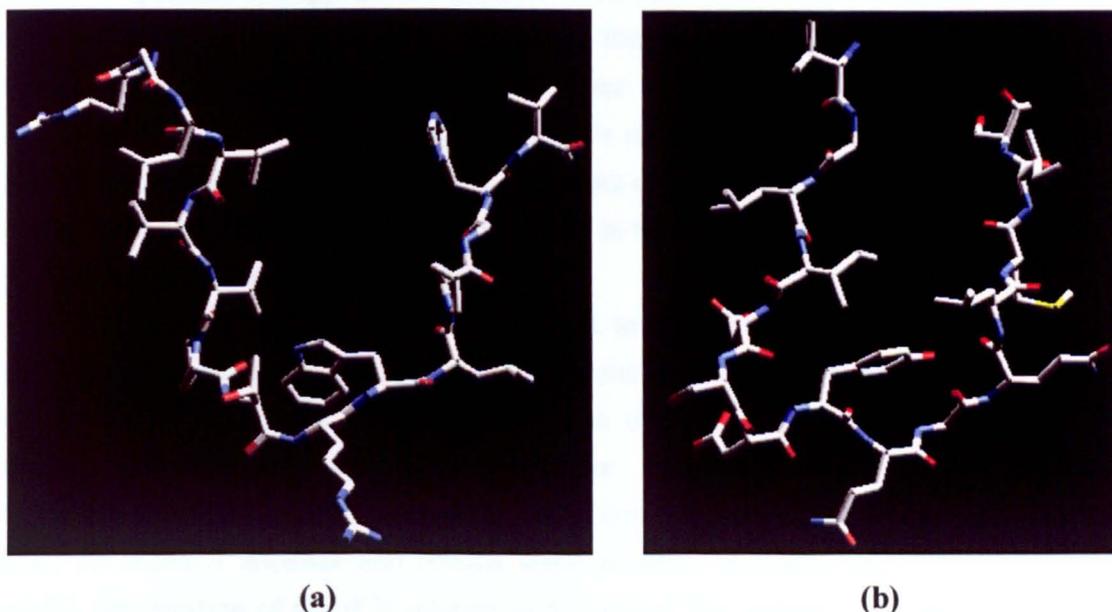
Figure 4.26 Modelling of HSV-1 motif X using the Swiss-Model ProMod program
Figure (a) shows the predicted structure of the HSV-1 motif X based on known structures with primary sequence similarity. Figure (b) shows the structure of the *E.coli* motif 3 loop for comparison.

Comparison of the predicted structure of the HSV-1 motif X to the structure of *E.coli* motif 3 reveals several similarities. The overall fold is generally homologous although the arms of the motif X loop (which form two anti-parallel β-strands in motif

3) are spread further apart. Notably there is a Trp at the base of motif X occupying a similar position to the highly conserved Tyr in motif 3.

The limitations of this type of modelling are clear. The prediction relies on the PDB database containing structures with primary sequence which closely match the test sequence. This modelling approach also assumes that similar primary sequences have similar folds which is not always the case. Surrounding protein interactions can influence local folding and cannot be predicted from this type of program. Despite these reservations it is encouraging to see that three different methods, ProteinPredict, Threader and Swiss-Modeller, are all in general agreement on the overall structure of the motif X region.

4.7.6 Discussion

Interest in the motif X region of the class II dUTPases was generated when primary sequence analysis by D.J. McGeoch at the MRC Virology Unit revealed sequence conservation outwith the known dUTPases. Several β-herpesvirus ORFs which occupy congruent genome positions to the α- and γ-herpesvirus dUTPases were shown to possess a copy of the class II motif X. These β-herpesvirus ORFs are unlikely to encode functional dUTPase given that they do not possess any of the five motifs which constitute the standard dUTPase active site. The fact that the only conserved region in the β-herpesvirus ORFs is motif X, lays open the possibility that this motif encodes a specific function. In terms of the class II dUTPases this function would be additional to dUTPase activity while in the β-herpesvirus homologues it could potentially constitute the sole function.

Three different approaches were used to predict the structure of the class II motif X region. Looking at the overall consensus between the ProteinPredict, Threader and Swiss-Modeller data there is evidence to support an anti-parallel β-strand loop structure with similarity to the motif 3 structure. This does not give a direct indication of potential function but it does support the theory that motif X structure originated from the motif 3 ancestor and retains some of these characteristics. In the class II model, the position of motif X relative to the rest of the protein is on the opposite side to the dUTPase active site. This is shown in the later Figure 5.17(a) which highlights motif X in the class II model. If this prediction is correct then it would be possible for motif X to possess a binding or catalytic function for example, without disrupting dUTPase activity.

The significance of an additional structure similar to motif 3 in the class II dUTPases is not clear. What is apparent is that the motif 3 β-hairpin structure is the basis for nucleotide recognition in the dUTPase active site allowing discrimination

between sugar and base. It is even possible that motif X retains this discriminatory function and is itself involved in nucleotide binding although this remains to be tested.

A secondary function supplied by motif X would be consistent with the class II evolutionary model proposed in the previous section (4.6). From an energetic perspective the class I dUTPases contain three active sites from three subunit chains. The class II dUTPase have only one active site for the equivalent of two subunit chains. This apparent inefficiency in the amount of protein synthesis required per active site would be redressed if the class II enzyme had an additional function besides dUTPase activity. These theories require to be tested experimentally and it is probable that not only the class II dUTPases need to be investigated but also the β-herpesvirus dUTPase homologues. Given that this motif X is specific to herpesviruses and therefore represents a potential drug target, the grounds for expanding research in this area are clear.

Chapter 5 - RESULTS OF EXPERIMENTAL WORK

5.1   Recombinant expression of HSV-1 dUTPase

5.1.1   Introduction

In order to perform detailed analysis of the HSV-1 dUTPase it was necessary to devise a system capable of generating a reliable supply of active enzyme. The major considerations were the ease of manipulation of the ORF, the quality and quantity of over-expressed enzyme and the potential for rapid purification. Several bacterial expression systems were investigated for their ability to meet these criteria.

5.1.2   The Kodak FLAG Biosystem
5.1.2.1 Overall Strategy

The Kodak FLAG system was employed to provide high expression levels combined with efficient detection and purification of the recombinant dUTPase. This system offered two main advantages compared to standard bacterial expression systems. The ORF could be cloned into the pFLAG vector resulting in a fusion protein with the OmpA signal peptide for secretion to the periplasmic space and the FLAG marker peptide for detection and purification. Generation of pure native enzyme could then be achieved with affinity purification using murine anti-FLAG IgG monoclonal antibodies followed by enterokinase digestion.

OmpA	FLAG	HSV – 1 dUTPase
Signal peptide permits secretion	FLAG peptide permits affinity column purification	FLAG peptide is cleaved to generate intact, native dUTPase
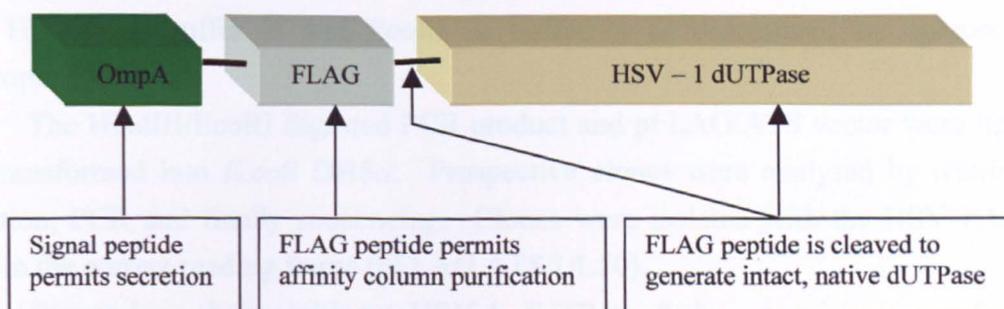
Figure 5.1 The IBI FLAG™ system components

The OmpA signal peptide - a 21 amino acid fusion peptide to allow secretion to the periplasmic space and resultant cleavage as it crosses the inner cytoplasmic membrane. Benefits included isolation of the recombinant protein from the cytoplasmic proteases and an efficient extraction protocol (Movra *et al.*, 1980).

The FLAG marker - a hydrophilic, 8 amino acid fusion peptide (DYKDDDDK) with a high surface probability representing good accessibility on the protein surface. The FLAG marker would allow column affinity purification using the anti-FLAG monoclonal antibody covalently linked to an agarose support.

The *tac* promoter and *lac*I repressor - the *tac* promoter is a combination of the *trp* promoter and the *lac* operator to allow strong controllable expression under the *lac*I repressor (Amann *et al.*, 1983).

The strategy was to insert the HSV-1 UL50 ORF into the MCS of the pFLAG.ATS vector to express a fusion protein with an N-terminal OmpA signal sequence and the FLAG octapeptide tag for binding anti-FLAG antibodies.

5.1.2.2 Results using the Kodak IBI System

The UL50 ORF was successfully amplified from purified HSV-1 (strain 17+) DNA by PCR using primers (F1/F2) designed to generate a 5' HindIII site and a 3' EcoRI site with 8bp overhangs.

```
F1.   5' GTG TGT GTA AGC TTA GTC AGT GGG GAT CCG GGG CGA TC 3'
                      HindIII

F2.   5' GTG TGT GTG AAT TCT CTA AAT ACC GGT AGA ACC AAA ACC 3'
                      EcoRI
```

Figure 5.2 Oligonucleotides synthesised for IBI FLAG system
Oligonucleotides F1 and F2 were designed to amplify HSV-1 UL50 by PCR generating a 5' HindIII site and a 3' EcoRI site. Note that each primer is flanked by a $(GT)_3$ tail to allow efficient restriction digestion of the fragment.

The PCR reaction was optimised until a sufficient quantity of product could be obtained at 25 cycles without contaminating products. Titration of $MgCl_2$ was found to have a large effect with a 2mM concentration providing at least 2X the product quantity than the standard 5mM concentration. The PCR product was gel purified and digested separately with HindIII and EcoRI. The pFLAG.ATS vector was successfully digested with HindIII in buffer B and EcoRI in buffer H as determined by agarose gel electrophoresis.

The HindIII/EcoRI digested PCR product and pFLAG.ATS vector were ligated and transformed into *E.coli* DH5α. Prospective clones were analysed by restriction digestion, PCR and finally sequencing. Clones were isolated with the HSV-1 UL50 ORF in the correct reading frame (pFLAG.ATS/UL50).

Expression of recombinant HSV-1 dUTPase fusion protein was performed according to Section 3.2.1. Plasmid pFLAG.ATS/UL50 was transformed into *E.coli* strain DH5α and cultures were grown up. Bacteria transformed with the host plasmid, pFLAG.ATS, were used as a control. Protein expression was induced by the addition of IPTG and extraction was performed according to Section 3.2.2.

Expression of the dUTPase-FLAG fusion protein was monitored by SDS-PAGE analysis of the bacterial protein extracts. A fusion protein of the predicted size was observed amongst the background of bacterial proteins (Figure 5.3 A). Further analysis of the bacterial extracts confirmed that this fusion protein was found exclusively in the insoluble fraction (Figure 5.3 B). Various changes in the expression conditions were

tested, aimed at increasing the solubility of the recombinant protein. The bacterial culture temperature was lowered down as far as 15°C in an attempt to moderate protein expression but there was no increase in the solubility of the recombinant dUTPase. Various experiments were performed changing conditions such as temperature (37C°, 30°C, 26°C or 15°C), induction time (1-20hr), IPTG concentration (0.5-0.05mM), bacterial media (L-broth or 2YT-broth) and bacterial host strain (DH5α or BW313) but the recombinant dUTPase protein remained insoluble.

Lane	Sample
1	High MW Markers
2	Crude extract Clone 1 pre-induction
3	Crude Extract Clone 1 post induction
4	Crude extract Clone 2 pre-induction
5	Crude Extract Clone 2 post induction
6	Control vector only post induction

Lane	Sample
1	High MW Markers
2	Insoluble periplasmic fraction
3	Soluble periplasmic fraction
4	Insoluble whole cell extract
5	Soluble whole cell extract

Figure 5.3 Expression and solubility of FLAG-tagged recombinant HSV-1 dUTPase
Gel A shows the expression of 2 pET3a/UL50 positive clones tagged with the OmpA signal sequence and the FLAG tag as is Figure 5.1. Comparison of pre- and post-IPTG induced expression (Lanes 2-5) demonstrates the appearance of a novel band of 40kDa corresponding to the tagged HSV-1 dUTPase (black arrows). Lane 6 shows the IPTG induced host plasmid control, pET3a, indicating the novel band is produced by the inserted UL50 gene.
Gel B shows the fractionation of a crude extract of Clone 1 as prepared according to Section 3.2.2. Comparison of the periplasmic fractions (Lanes 2 and 3) indicates that the tagged HSV-1 dUTPase is transported to the periplasmic space but remains insoluble. Comparison of the whole cell fractions (Lanes 4 and 5) confirms that the recombinant dUTPase is insoluble.

5.1.3   The pET T7 expression system

5.1.3.1 Overall Strategy

The plasmid pET-3a/UL50 was obtained as a kind gift from Dr. O. Björnberg, Lund University, Sweden. This construct contained the UL50 ORF in the pET-3a vector to allow expression under the control of the T7 promoter. This system was shown to allow efficient expression of active soluble HSV-1 dUTPase by Björnberg *et al.* (1993).

The UL50 gene was subcloned from pET-3a/UL50 into the more recently developed pET-23a vector (Novagen, Inc.). The strategy was to maintain the successful expression obtained in the pET-3a/UL50 and in addition supply an f1 origin to facilitate mutagenesis and sequencing in the same plasmid.

5.1.3.2 Subcloning of UL50

The pET-3a/UL50 construct was transformed into *E.coli* DH5α for the propagation of plasmid DNA. The plasmid was harvested using the Promega Wizard system (Section 3.1.11) to obtain a large quantity of purified pET-3a/UL50 DNA. The fragment containing the UL50 gene was excised by double restriction digests using HindIII and XbaI and gel purified (Sections 3.1.4 and 3.15). The pET23a plasmid was harvested from transformed *E.coli* DH5α and digested with HindIII and XbaI as before. The DNA fragment containing the UL50 gene was ligated into the pET-23a vector and transformed into *E.coli* DH5α. Potential clones were screened by restriction analysis and agarose gel electrophoresis.
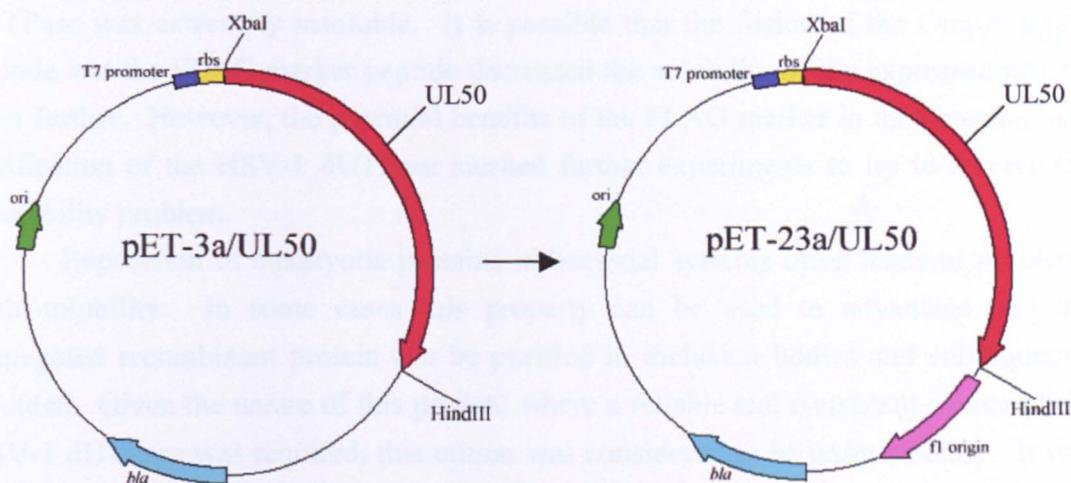
Figure 5.4 Plasmid map of pET-3a/UL50 and pET-23a/UL50
The pET-3a/UL50 vector contains the T7 promoter and a ribosome binding site (rbs). Plasmid selection in bacteria is achieved with the *bla* gene expressing ampicillin. The UL50 ORF was transferred to the newer plasmid pET-23a using the Xba I and Hind III restriction sites.

5.1.3.3 Expression and extraction of HSV-1 dUTPase

The method for expression of the pET23a/UL50 construct and extraction of HSV-1 dUTPase was performed essentially according to O. Björnberg *et al.* (1993) (see Sections 3.2.1 and 3.2.2). There were two main differences from the methods used in the Kodak IBI system. Lysis of the induced bacteria was performed by freeze/thawing in the presence of detergent as opposed to the addition of lysozyme. Detergent (Triton X-100 or polyoxyethylene-10-tridecyl ether) was used at a concentration of 0.1% in the extraction buffer and at 0.01% in all subsequent purification and storage buffers. Experiments had previously shown that although the addition of detergent did not increase the release of protein from the bacteria, it did result in an increased recovery of recombinant dUTPase activity (Björnberg *et al.*, 1993).

Extracts were analysed for the expression of HSV-1 dUTPase from the pET23a/UL50 construct by SDS-PAGE (Section 3.2.3). Soluble HSV-1 dUTPase was detected and shown to possess activity although there was still a significant amount of insoluble enzyme. This was in agreement with previous experiments with the donor construct, pET3a/UL50, which still allowed successful purification (Björnberg *et al.*, 1993). With a source of active recombinant dUTPase in the vector of choice it was decided to continue to the purification stage.

5.1.4   Discussion

The original Kodak IBI bacterial expression system was chosen before experiments by O. Björnberg *et al.* (1993) had shown that bacterially expressed HSV-1 dUTPase was extremely insoluble. It is possible that the fusion of the OmpA signal peptide and the FLAG marker peptide decreased the solubility of the expressed protein even further. However, the potential benefits of the FLAG marker in the detection and purification of the HSV-1 dUTPase merited further experiments to try to resolve the insolubility problem.

Expression of eukaryotic proteins in bacterial systems often leads to problems with solubility. In some cases this property can be used to advantage and the aggregated recombinant protein can be purified in inclusion bodies and subsequently refolded. Given the nature of this project, where a reliable and consistent source of the HSV-1 dUTPase was required, this option was considered to be unsatisfactory. It was reasoned that there was potential for site directed mutations to alter the efficiency of the protein refolding. In such a case, it would be extremely difficult to determine whether a mutation was modifying enzyme activity by interfering with active site interactions or as a consequence of altered refolding.

It is known that a reduction in the bacterial culture temperature during expression of recombinant proteins increases solubility in some cases. It is presumed

that by slowing the metabolism of the bacterial cell, and consequently the expression of the recombinant protein, correct folding is more likely to be achieved. This is exemplified in the bacterial expression of vaccinia virus ribonucleotide reductase (RR) (Slabaugh *et al.*, 1993). At 37°C bacterial expression of vaccinia virus RR large subunit yielded a completely insoluble product. Reduction of temperature to 15°C and IPTG concentration down from 0.4mM to 0.05mM produced up to 70% of the recombinant RR subunit in a soluble form. Attempts at reducing the expression of HSV-1 dUTPase by lowering temperature, reducing IPTG concentration and using less rich media (L-broth instead of 2YT-broth) did not increase the solubility of the enzyme.

It was known at this point that HSV-1 dUTPase had been successfully expressed in the pET system as an untagged protein (Björnberg *et al.*, 1993). In this study it was discovered that the method of extraction was crucial to the recovery of active enzyme. For example Yeda pressing, which was successfully used to extract overproduced *E.coli* dUTPase (Hoffman *et al.*, 1987), was found to be unsatisfactory at extracting HSV-1 dUTPase. A considerable quantity of protein was obtained by this method but the activity was very low. This problem was partially overcome by adding detergent to the extraction buffer. Following this example, detergent was used when lysing bacteria expressing the HSV-1 dUTPase FLAG fusion protein, but again, there was no increase in solubility.

It was decided in the light of the solubility problem that a new expression system should be tried. Various systems were considered but it seemed reasonable to choose the pET system which had been shown to produce partially soluble, untagged, HSV-1 dUTPase. The expressing construct, pET3a/UL50, was kindly donated by Dr O. Björnberg, Lund University, Sweden. This construct was transformed into *E.coli* strain BL21 and found to produce active dUTPase. Unfortunately this construct did not contain an f1 origin of replication and therefore the ORF would have to be shuttled between two vectors, one for mutagenesis and one for expression. To circumvent this problem, the UL50 ORF was successfully subcloned into the newer vector, pET23a, which contained an f1 origin of replication.

The reason why HSV-1 dUTPase is more soluble in the pET system is probably twofold. Firstly, the enzyme is expressed in its native form with no additional amino acid residues. Secondly, from previous studies on the extraction of HSV-1 dUTPase, it is likely that the method of extraction plays a large role in the recovery of soluble enzyme. In the pET system, freeze/thawing of the bacterial cells disrupts the cell membrane thus releasing the indigenous lysozyme present in the *E.coli* strain BL21(DE3)pLysS. This technique may be instrumental in the release of a higher proportion of soluble enzyme.

The use of the pET system allowed the production of active, soluble HSV-1 dUTPase. However, the production of untagged enzyme, one of the functions of the system which is necessary for its effectiveness, created a more complex problem for purification. The original plan for purification, based on affinity for a fusion tag, could not be applied therefore the native enzyme had to be purified on the basis of its specific biochemical characteristics. Although this was a major disadvantage when purifying a large number of enzyme constructs it allowed greater inference into the effect of site-directed mutations in the context of the native enzyme. In short, there is a trade-off between preserving the protein close to the natural condition and devising a system whereby the recombinant protein can be easily manipulated and examined.

The transfer of the HSV-1 ORF into the pET23a vector combines the benefits of a soluble expression system with the potential for generating mutations in the same construct without subcloning.

5.2   Purification of HSV-1 dUTPase

5.2.1   Introduction

The aim of this stage of work was to develop a reliable and consistent method applicable for the production of wild type and mutant HSV-1 dUTPase in a pure state. Initially purification of the HSV-1 dUTPase was based on a two stage system employing phosphocellulose and Mono S cation exchange columns. This protocol was later refined to allow a one step Mono S purification utilising the selective solubility of the HSV-1 dUTPase in the presence of chelators.

5.2.2   FPLC purification by cation exchange

Phosphocellulose followed by Mono S cation exchange column chromatography had been previously shown to allow successful purification of HSV-1 dUTPase expressed from the pET3a/UL50 construct dUTPase (Björnberg et al., 1993). Initial experiments demonstrated that this protocol was also efficacious at purification from my own construct, pET23a/UL50 (Section 3.2.4). However, consistent with the Swedish study, the end product of this purification was not homogeneous and was degraded over time even in the presence of protease inhibitors.

5.2.3   Selective buffer extraction

A new purification protocol was investigated based on the selective solubility of HSV-1 dUTPase in various buffers (A.-C. Bergman, personal communication). It was found that a high degree of purification could be achieved by altering the initial extraction buffers used immediately after recombinant expression. Following bacterial

lysis, cells were spun as normal and then resuspended in modified extraction buffer. This buffer differed from the standard pET extraction buffer (Section 2.2.4) by the exclusion of the chelators EDTA and EGTA normally present each at 1mM. This had the effect of partitioning up to 90% of the HSV-1 dUTPase to the pellet. Re-extraction of this pellet was then performed with the standard pET extraction buffer containing 1mM EDTA and 1mM EGTA. The sample was allowed to re-extract overnight at 4°C followed by centrifugation. The supernatant contained about 40% of the HSV-1 dUTPase from the pellet in an active and almost homogeneous state.

This method was employed in preference to the original extraction method. Further purification was deemed necessary and column chromatography was used a final step. Several extracts were tested using various column and buffer combinations. Mono S chromatography was found to be the most reliable and consistent method using standard buffers containing 0.01% detergent and 1mM EDTA and EGTA. Many of the mutant enzymes were repurified using this superior method.
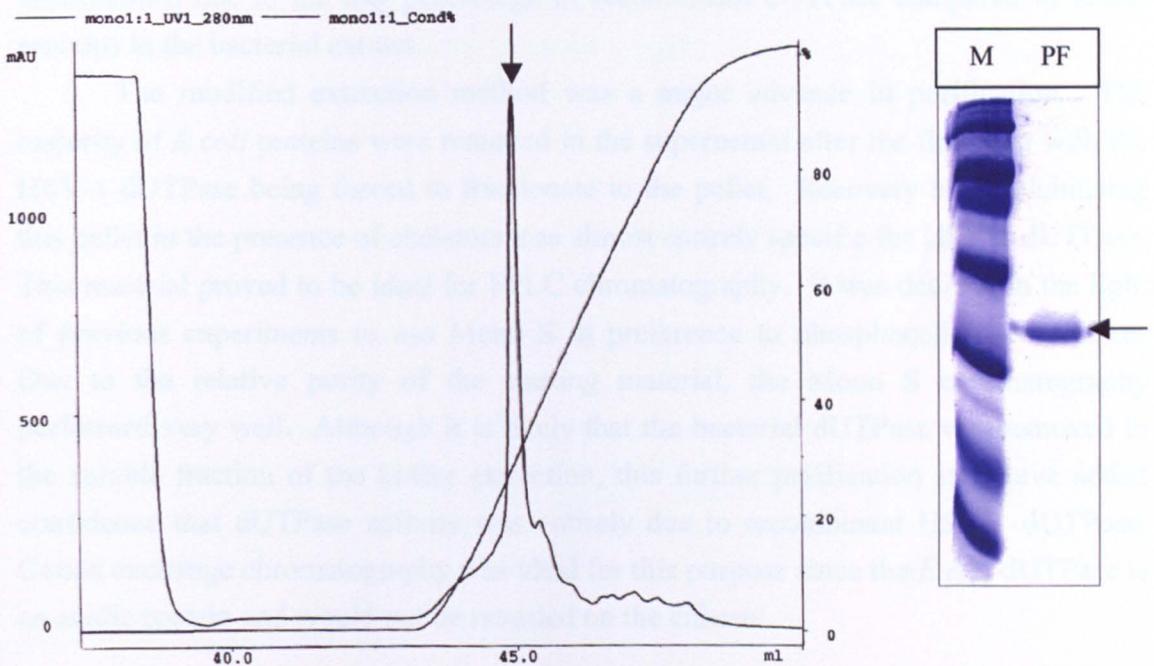
Figure 5.5 FPLC and SDS-PAGE of recombinant HSV-1 dUTPase
The graph above is the gradient section of an FPLC trace taken directly from the Unicorn operating system of the Pharmacia FPLC machine. It shows an example of the final Mono S chromatography purification stage. The starting material is semi-purified HSV-1 dUTPase from the modified buffer extraction method detailed above. Cleared supernatant was injected into a prewashed Mono S HR 5/5 column and washed at a flow rate of 1ml/min. A NaCl gradient from 0.1M to 1M was applied to the column at the same flow rate. Fractions were collected and analysed by SDS-PAGE and dUTPase assays.
        The SDS-PAGE gel (M=MW markers, PF=peak fraction) shows the peak fraction (black arrow on trace) as a single band which corresponded to the peak dUTPase activity. The gel shows the absence of the low MW impurities found in the previous phosphocellulose/Mono S purification system.

5.2.4 Discussion

Initial purification based directly on the method of O. Björnberg *et al.* (1993) proved to be problematic at producing homogeneous HSV-1 dUTPase from the pET bacterial expression system. Low molecular weight bands could be visualised by SDS-PAGE and, as reported in the Swedish study, degradation products could also be seen over time. A further problem was the variability of the phosphocellulose columns. These were made up freshly in the lab and although every effort was taken to ensure consistent handling it was likely that inconsistencies in the packing of the phosphocellulose matrix caused overall variability. In practical terms this meant that these runs could not be automated and that there was a large variability in enzyme recovery and purity. For consistent purification in order to allow direct comparison of different mutant enzyme preparations this system was deemed unacceptable. To get round this problem a one stage Mono S purification was attempted. It was hoped that the pre-packed Mono S columns would allow more consistent purification. This was unsuccessful due to the low percentage of recombinant dUTPase compared to *E.coli* proteins in the bacterial extract.

The modified extraction method was a major advance in purification. The majority of *E.coli* proteins were removed in the supernatant after the first spin with the HSV-1 dUTPase being forced to fractionate to the pellet. Recovery by resolubilising this pellet in the presence of chelators was almost entirely specific for HSV-1 dUTPase. This material proved to be ideal for FPLC chromatography. It was decided in the light of previous experiments to use Mono S in preference to phosphocellulose columns. Due to the relative purity of the starting material, the Mono S chromatography performed very well. Although it is likely that the bacterial dUTPase was removed in the soluble fraction of the buffer extraction, this further purification step gave added confidence that dUTPase activity was entirely due to recombinant HSV-1 dUTPase. Cation exchange chromatography was ideal for this purpose since the *E.coli* dUTPase is an acidic protein and would not be retarded on the column.

5.3   Site-directed mutagenesis of HSV-1 dUTPase

5.3.1   Introduction

The method of Kunkel *et al.* (1987) was chosen to perform site-directed mutagenesis. This method was considered practical to allow a large number of mutants to be generated from a single batch of HSV-1 UL50 ssDNA template. Each specific mutation required a single oligonucleotide to be synthesised thus providing an efficient, low cost system. A substantial amount of time was dedicated to optimising the mutagenesis system so that any future mutations could be made quickly.

5.3.2   Mutagenesis rationale

Initially the tyrosine at position 100 in the motif 3 region of the HSV-1 dUTPase was chosen to mutate to an alanine (3Y100A). This was an ideal first target because this residue appears to be of major importance in the *E.coli* enzyme (Vertessy *et al.*, 1995) and is highly conserved in both class I and class II dUTPases (Sections 4.2 & 4.3). The DNA sequence change required to make this mutant involved the loss of a unique KpnI restriction site which allowed fast screening of potential mutated plasmids by restriction analysis.

Wild type UL50   5'.....TCG GGG TAC CGC GGA..... 3'

                                                    ↓

3Y100A mutant   5'.....TCG GGG GCT CGC GGA..... 3'
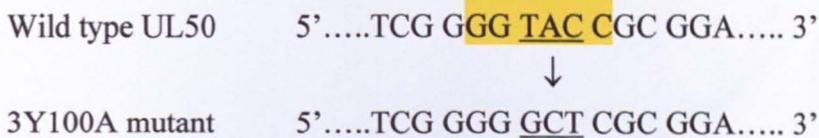
Figure 5.6 Position of 3Y100A mutation in relation to wt KpnI restriction site
The changed nucleotides are underlined and the KpnI restriction site is coloured yellow (recognition site GGTAC*C where the star indicates the cutting position). Mutation of the wt UL50 coding sequence from TAC (tyrosine) to GCT (alanine) disrupted the KpnI site allowing for fast mutation screening. First nucleotide shown is 292 bases from the 5' end of UL50.

5.3.3   Mutagenesis results for 3Y100A

Kunkel mutagenesis was performed according to Section 3.1.14. It was found that the preparation of high purity uracil enriched ssDNA template was a key factor for efficient mutagenesis. The mutagenesis reactions were eventually refined to yield over 10% positive mutant colonies from those picked and analysed. This was an acceptable level to allow generation of approximately 20 mutant constructs. Figure 5.7 shows restriction digest analysis of ten plasmids generated by the mini-prep method, digested with KpnI and visualised by gel electrophoresis (Sections 3.1.10 & 3.1.3). Note that the wt plasmid, pET23a/UL50, contains only one KpnI site which lies in the motif 3 region of the UL50 gene. Efficient restriction with this enzyme therefore produces linear DNA from the circular plasmid which can be visualised as a single band.

Plasmids which have lost the KpnI site remain intact and are visualised as multiple bands corresponding to the various forms of plasmid DNA.

Figure 5.7 Scanning for 3Y100A mutation by restriction digestion
10 potential colonies from the Kunkel method were grown up and mini-prepped (Section 3.1.10). Standard restriction digests were performed with KpnI restriction enzyme (Section 3.1.5) and products visualised by agarose gel electrophoresis. Lanes 5 and 7 show undigested plasmid as a result of site-directed mutagenesis within the KpnI site. All other lanes represent wt plasmids linearised by digestion with KpnI.

To confirm that this was indeed the correct mutation the area in the region of the KpnI site was sequenced. Figure 5.8 shows the nucleotide sequence of the wild type HSV-1 UL50 gene with the corresponding nucleotide changes in the 3Y100A mutant.

Figure 5.8 Sequencing of the wt HSV-1 UL50 and corresponding mutant 3Y100A
Figure shows the result of sequencing the complementary DNA strand. Note the base changes from GTA (reverse complement TAC) coding for tyrosine to AGC (reverse complement GCT) coding for alanine (highlighted in yellow).

5.3.4 Large scale mutagenesis of HSV-1 UL50

The targets for mutagenesis were based primarily on two criteria: those residues which were highly conserved between class I and class II dUTPases (Section 4.3) and specific areas of interest within the HSV-1 dUTPase tertiary model (Section 4.4.5).

All the oligonucleotides were designed to yield a single amino acid change in the HSV-1 dUTPase protein. Each one was designed to give a specific codon change with the exception of primer 15 which was synthesised to give a mixed population. This primer was included to test the efficiency of generating multiple mutants at a specific locus with a single mixed primer synthesis. Table 5.1 shows the mutant oligonucleotides which were synthesised for the specific mutagenesis of the HSV-1 UL50 ORF.

No	Mutant	Nucleotide Change	Oligonucleotide Sequence
1	3Y100A	TAC→GCT	5' CAT AAC GGT TCC GCG **AGC** CCC CGA GTC GAT AAG 3'
2	3Y100S	TAC→AGT	5' CAT AAC GGT TCC GCG **ACT** CCC CGA GTC GAT AAG 3'
3	3Y100F	TAC→TTT	5' CAT AAC CGT TCC GCG **AAA** CCC CGA GTC GAT AAG 3'
4	3D97A	GAC→GCC	5' TCC GCG GTA CCC CGA **GGC** GAT AAG ACC CAG TAT 3'
5	3D97E	GAC→GAA	5' TCC GCG GTA CCC CGA **TTC** GAT AAG ACC CAG TAT 3'
6	XW275A	TGG→GCG	5' TAC GTG CCC GGG GAG **CGC** GCG CGT AGG AAC GAC 3'
7	XW275F	TGG→TTT	5' TAC GTG CCC GGG GAG **AAA** GCG CGT AGG AAC GAC 3'
8	XT273S	ACG→TCC	5' CCC GGG GAG CCA GCG **GGA** AGG AAC GAC CAG GAG 3'
9	XT273A	ACG→GCG	5' CCC GGG GAG CCA GCG **CGC** AGG AAC GAC CAG GAG 3'
10	4Q303N	CAG→AAC	5' CCC CGC AAC CAG GAG **GTT** GGC GAC CTT GGC GCC 3'
11	4Q303D	CAG→GAC	5' CCC CGC AAC CAG GAG **GTC** GGC GAC CTT GGC GCC 3'
12	4Q303E	CAG→GAG	5' CCC CGC AAC CAG GAG **CTC** GGC GAC CTT GGC GCC 3'
13	4Q303T	CAG→ACG	5' CCC CGC AAC CAG GAG **CGT** GGC GAC CTT GGC GCC 3'
14	NC76A	TGC→GCC	5' GGC GTG AAT AAT CGC **GGC** AAA GTC AGC CGG CAT 3'
15	4Q303L	CAG→TTG	5' CCC CGC AAC CAG GAG **SNN** GGC GAC CTT GGC GCC 3'
16	5F366Y	TTT→TAT	5' AAT ACC GGT AGA ACC **ATA** ACC CCC GGT CCC GCG 3'
17	5F366A	TTT→GCT	5' AAT ACC CGT AGA ACC **AGC** ACC CCC GGT CCC GCG 3'
18	4K300A	AAG→GCG	5' CAG GAG CTG GGC GAC **CGC** GGC GCC GGC CTC GAG 3'
19	2R260A	CGG→GCA	5' GGC GTT GAG CGA CGA **TGC** CCC CAA CAC ATA GCA 3'
20	1D222A	GAC→GCC	5' GCG ACG GAC GAC AAT **GGC** GAA ACC GGC ATC CTC 3'

Table 5.1 Oligonucleotides synthesised for site-directed mutagenesis of HSV-1 UL50
Mutagenic primers are the reverse complement of the 5'→3' UL50 nucleotide sequence due to the direction of the F1 origin in pET23a. Primer 15 was synthesised to produce a mixed population of the following bases: S=G/C, N=T/A/G/C. The first mutant sequenced from this primer was 4Q303L (CAG→TTG).

Due to the nature of the Kunkel mutagenesis system it was decided to fully sequence all of the UL50 mutant constructs. It was essential to have confidence that there was no secondary mutations anywhere in the HSV-1 dUTPase coding region. A primer set was generated to cover the entire UL50 gene allowing good overlap between each successive sequence reading. These oligonucleotides are shown in Table 5.2.

Code Number	Distance from UL50 ATG start codon (bps)	Direction	Oligonucleotide sequence
P1	98	R	5' GCC CCC GCG AGT AGC GAG GG 3'
P2	31	F	5' TCC AGC CGG ACA GCT TGG GT 3'
P3	131	F	5' CGC GGT GGC TTT TAT GCC GA 3'
P4	221	F	5' CTT TTG CGC GAT TAT TCA CG 3'
P5	331	F	5' CTA AAA GGA CGC GGG AAT TT 3'
P6	421	F	5' CGA TTT CCC TGC GGC AGT TC 3'
P7	518	F	5' AAG CGT GAC TAC GGC CCT AC 3'
P8	630	F	5' TTC CTT CCA AAA CGC GAG GA 3'
P9	721	F	5' CAT CCC TCC GCA TGC TCC AC 3'
P10	846	F	5' TTT GTT GTT TAC AAC CTT AC 3'
P11	966	F	5' ACC AAA GCG CTT CGA AAC TA 3'
P12	1038	F	5' GTG TTT ACG AAC GAG TTT GA 3'
P13	424	R	5' ATC GGC TCG GTG AGG GCC GG 3'
P14	1030	R	5' GGC GGG TTC CTG GGT TCG GC 3'
P15	810	R	5' CGA CCA GGA GGC CGC GGG CG 3'

Table 5.2 Oligonucleotides synthesised for sequencing HSV-1 UL50

Oligonucleotides denoted with an 'F' were designed to sequence in the forward direction (5'→3') and 'R' in the reverse direction (3'→5'). Distances are calculated from the first nucleotide in the ATG start codon to the 5' terminal nucleotide of the primer. All primers are 20 nucleotides in length

This primer set was used extensively since the majority of subsequent mutations did not change the restriction profile of the UL50 gene and therefore had to be screened by sequencing. This led to the development of small volume single track sequencing. This method differed from the standard sequencing method (Section 3.1.13) in that reactions were performed in microwell plates using only 1/8th of the standard reaction volumes. The same single dideoxynucleotide was chosen to terminate each reaction based on one base change expected from the mutagenic oligonucleotide.
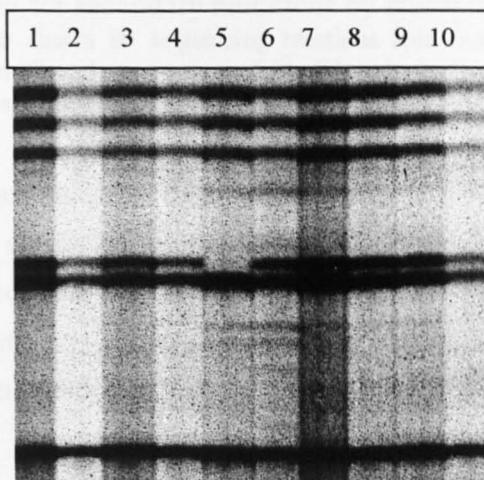
Figure 5.9 Mutation screening by single track sequencing

Prospective mutant plasmids were analysed by single track sequencing. This gel shows a typical screening for mutant 1D222A. The expected nucleotide changes in this mutant were GAC→GCC. Plasmids were sequenced in single reactions using ddATP as a chain terminator. Lane 5 represents a mutant with the expected substitution of an adenine nucleotide.

This allowed the rapid screening of potential mutant plasmids by sequencing while being very economical with reagents and Sequenase enzyme. It proved to be very useful since mutations could initially be picked out clearly by eye without the need to produce large numbers of linear nucleotide sequences. Figure 5.9 shows a typical example of this method.

Using this method it was possible to isolated plasmids positive for all of the 20 targets shown in Table 5.1. Each of these constructs was then fully sequenced using the primer set in Table 5.2.

5.3.5 Results of sequencing mutant constructs

Due to the large number of sequence reactions required to fully sequence all the mutant UL50 genes a similar method to the mutation screening was employed. For each primer, modified 1/8th volume reactions were set up for all 19 mutants plus the wt control. Narrow wells were used to allow 80 reactions to be run on each gel. The samples were loaded in groups of 20 with the same termination dideoxynucleotide (Figure 5.10). This allowed secondary mutations to be identified easily.

1-20 (T)	1-20 (C)	1-20 (G)	1-20 (A)
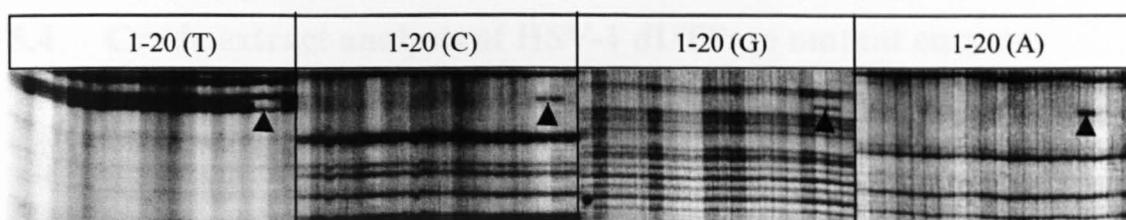
Figure 5.10 Screening for secondary mutations by multiple track sequencing
Section of autoradiograph shows 80 sequencing reactions split into 4 groups of 20. Each group represents 19 mutant plasmids and one wt sequenced with a single dideoxynucleotide. Mutations can be see as the absence and presence of single bases (arrows).

It was found at this stage that some constructs possessed additional mutations. These were generally single base mutations at sites distal from the annealing area of the mutagenic oligonucleotide. Six out of the nineteen starting mutants were found to contain additional mutations. It was decided to carry on to the analysis stage with the thirteen remaining constructs.

5.3.6 Discussion

The Kunkel method of site-directed mutagenesis proved to be useful at producing UL50 mutant constructs but did have some unexpected problems. It was hoped that this method would have the advantages of reproducibility and accuracy compared to other methods. Site-directed mutagenesis by PCR had also been investigated but was ruled out at an early stage due the possibility of polymerase errors

causing secondary mutations. In the light of the Kunkel mutagenesis results, with 6 out of 19 mutants synthesised with additional mutations, this rationale was unjustified. It was hoped that because each mutation was produced from the same batch of uracil enriched ssDNA, with the mutagenic oligonucleotide the only variable, that secondary mutations would be uncommon. It was expected that any additional mutations which did occur would be in the annealing region of the mutagenic primer caused by poor synthesis of the oligonucleotide. It was therefore surprising to find that all of the secondary mutations occurred at sites distal from the mutagenic oligonucleotide.

These mutations were either a single base change, single base insertion or single base deletion. The most likely explanation for such random mutations was T7-DNA polymerase errors during the *in vitro* second strand synthesis. It is unlikely that the uracil enriched ssDNA template was at fault since no two mutations occurred at the same site. The reason for the T7 DNA polymerase errors is unknown. What remains clear is that for any mutagenesis reaction using similar methods it is essential to perform complete sequencing of the gene of interest for every construct synthesised.

5.4   Crude extract analysis of HSV-1 dUTPase mutant enzymes

5.4.1   Introduction

Experiments were conducted to test the feasibility of analysing the dUTPase mutant enzymes in crude extracts. It was recognised that the bacterial host dUTPase would exhibit background activity in this type of analysis but at this stage it was reasonable to suggest that this could be measured effectively in control experiments.

5.4.2   Expression

The 16 sequenced UL50 mutant constructs were transformed into the *E.coli* BL21(DE3)pLysS strain allowing expression under the T7 promoter (see Section 5.1.4.4). 100ml cultures were grown up from glycerol stocks of transformed bacteria. These were induced and expressed according to Section 3.2.1(b). Lysis was performed by freeze thawing and extracts were assayed for dUTPase activity by the TLC method (Section 3.2.6).

5.4.3 Crude extract results

Initial expression experiments produced a range of dUTPase activity in the various mutants. It was shown that the specific activity calculated for individual mutant constructs was variable between assays and within a narrow range compared to the controls.
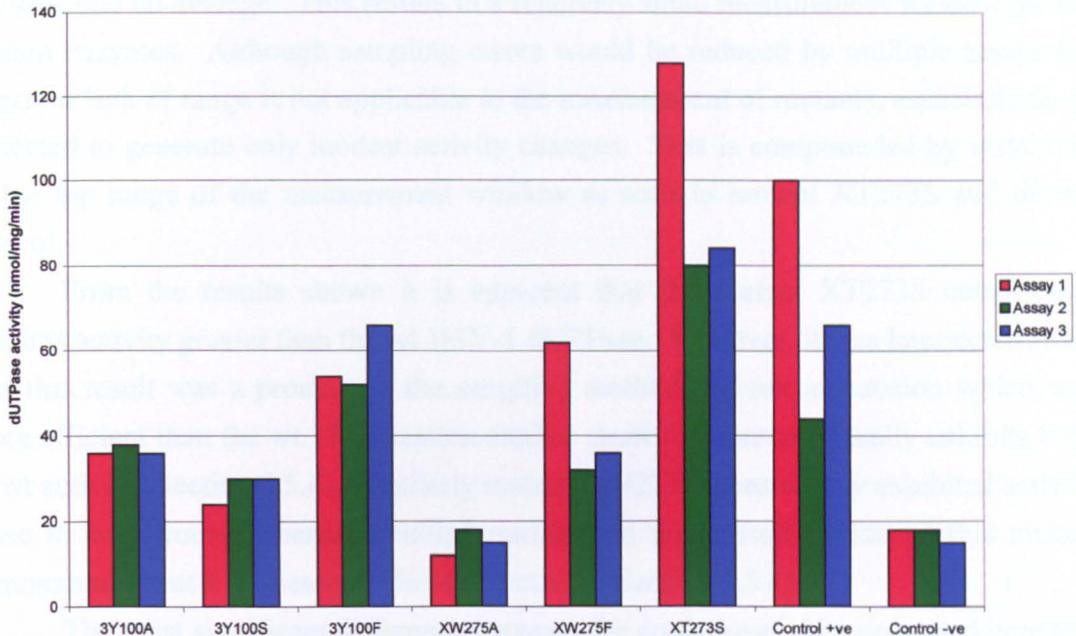
Figure 5.11 Comparison of dUTPase activity in crude extracts
Activity measurements are given for 6 mutant constructs plus a positive (pET23a/UL50) and negative (pET23a) control. Three readings are given for each sample representing three separate expression experiments. dUTPase activity is calculated as the amount of dUTP converted to dUMP in one minute per mg of total protein.

Attempts were made to produce consistent activity values for the mutants by regulating expression. Culture growth was followed closely and IPTG induction was performed only when exact OD_{600} readings were reached. Although some progress was made, it became clear that crude extract analysis was not an acceptable method to measure small differences between mutant dUTPase constructs.

Figure 5.11 shows data for three such expression experiments. The background activity from the negative plasmid only control was relatively consistent, however, cultures expressing recombinant HSV-1 dUTPase varied considerably between different expression experiments. The wt positive control could not be standardised by this method with activities ranging from 100nmol mg^{-1}min^{-1} (Assay 1) down to 44nmol mg^{-1}min^{-1} (Assay 2). There is a reasonable level of consistency in many of the mutants assayed such as 3Y100A and 3Y100S. Unfortunately these data are not sufficient to analyse the mutant enzymes with any degree of precision due to the small range between the negative and positive controls.

147

In every assay a control was set up containing all the reaction components except enzyme. This was used to calculate any background degradation of the dUTP substrate and was subtracted from the results shown. In this respect, the activity measured for the negative control is due entirely to the *E.coli* host dUTPase. This background is high compared to the positive control representing approximately 30% of the wt value on average. This results in a relatively small measurement window for the mutant enzymes. Although sampling errors would be reduced by multiple assays the apparent lack of range is not applicable to the measurement of mutants, especially those expected to generate only modest activity changes. This is compounded by instability in the top range of the measurement window as seen in mutant XT273S and the wt control.

From the results shown it is apparent that the mutant XT273S consistently exhibits activity greater than the wt HSV-1 dUTPase. However, it was later determined that this result was a product of the sampling method and not a mutation which was more efficient than the wt. Purification studies show this mutant actually exhibits 85% of wt activity (Section 5.5.4). Similarly mutant XW275A consistently exhibited activity close to background whereas detailed purification and quantification of this mutant demonstrated that it possesses 29% of wt activity (Section 5.5.4).

The most significant difference between the crude assay data presented here and the detailed analysis presented later in Section 5.5 was the method by which specific activity is calculated. In crude extract analysis, enzyme activity is measured in terms of total bacterial protein. In order to compare extracts in this way it must be presumed that the fraction of total protein which is HSV-1 dUTPase remains constant between expression experiments. Subsequent comparison of activities calculated from crude extracts compared to purified enzyme demonstrated that this was not the case.

5.4.4 Discussion

The literature contains many examples of experiments where the activity of mutant enzyme constructs expressed in *E.coli* are measured directly in crude extracts. The initial line of investigation in this study was to employ a similar methodology for the HSV-1 dUTPase mutants allowing a fast resolution of results. Although efforts were made to overcome the problems, clearly crude extract analysis is not compatible with this expression system.

It was noted that *E.coli* transformed with only the control plasmid, pET23a, grew faster than cultures expressing recombinant HSV-1 dUTPase from plasmid pET23/UL50. This gave a direct indication that recombinant expression was mildly toxic to the host bacteria. This could potentially create positive selection pressure for any bacteria harbouring plasmids which are deficient at expressing the recombinant

protein. It is likely in large bacterial cultures grown over long periods that these non- or low-expressing bacteria would outgrow the high expressing bacteria to some degree. This effect would be amplified by the use of starter cultures.

This problem arose due to the fact that even in the absence of IPTG, there was still a low level of recombinant expression. This leaky expression was caused in this system by low level expression of T7 RNA polymerase from the *lacUV5* promoter in the pLysS lysogen and subsequent transcription of the UL50 gene. This may explain why several rounds of selective single colony isolation, fresh transformation plus the discontinued use of overnight starter cultures improved culture growth consistency.

Cultures grown to saturation, such as overnight cultures, contain large amounts of β-lactamase from the *bla* gene in the pET plasmid (Studier *et al.*, 1990). If such a culture is then used to inoculate a larger flask, much of the ampicillin in the media can be destroyed. Since there is selection for bacteria carrying the pET plasmid only when there is a source of ampicillin there may be a resultant outgrowth of bacteria deficient in the pET plasmid completely.

It is likely that these effects contributed significantly to the variability seen in the crude extract analysis. Aside from expression itself, further effects would be incorporated due to the variation in soluble enzyme recovery. It is probable that this property resulted in the misleading mutant data shown in Figure 5.11. Mutant XT273S showed consistently higher activity than the wt enzyme. Subsequent purification and quantification of this mutant revealed that there was up to a 3 times greater recovery of soluble enzyme compared to the wt HSV-1 dUTPase. This suggests that the mutant dUTPase from this construct may be either expressed to a higher percentage of total bacterial protein or is more soluble than the wt enzyme. The activity result is therefore explained since crude extract comparison calculates activity in terms of total soluble protein. Likewise mutant XW275A appears to have a lower recovery than the wt enzyme and may be expressed to a lower degree or is more insoluble than the wt. Crude extract analysis is therefore measuring three distinct variables: enzyme activity, enzyme expression and enzyme solubility. To separate these three factors it was necessary to perform activity analysis on purified quantifiable enzyme.

Since purification of the wt HSV-1 dUTPase had been reasonably well optimised (Section 5.2) it was decided to employ the same methodology with a selection of mutant enzymes.

5.5 Detailed analysis of HSV-1 dUTPase mutant constructs

5.5.1 Introduction

In order to perform activity analysis in a more controlled environment, attempts were made to purify and quantify the recombinant HSV-1 dUTPase enzymes. Purification of the mutant enzymes was based on the methodology used for the wt enzyme (Section 5.2). These enzymes were then quantified by SDS-PAGE allowing accurate comparison between each mutant.

5.5.2 Purification of HSV-1 dUTPase mutants

To produce sufficient quantities of the mutant HSV-1 dUTPase constructs for purification, multiple 300ml cultures were grown up in 2l flasks. It was noticed that bacterial growth varied considerably for different mutant cultures inoculated at the same time and in the same incubator. After consultation with A.-C. Bergman on this matter it was discovered that three rounds of single colony selection of *E.coli* BL21(DE3)pLysS followed by fresh transformation with the appropriate mutant plasmid improved growth consistency. This method, although time consuming, was effective and therefore used for all subsequent expression experiments.

Initially extraction and purification was performed according to the method of O. Björnberg *et al.* (1993). The low purity of the recombinant enzymes purified in this way presented a major problem. SDS-PAGE revealed multiple bands and evidence of protease activity. The majority of mutants purified by this method were deemed unacceptable for activity analysis. As attempts were being made to improve the existing method, details of the new method became available. Five mutant enzymes plus a wt control were extracted and purified according to the modified method as detailed in Section 5.2.2. The combination of the highly selective chelator extraction plus Mono S chromatography allowed these samples to be purified to a high degree.

5.5.3 Quantification of HSV-1 dUTPase mutants

Correct measurement of specific enzyme activity was shown to be highly significant in this analysis. There were many variables which had to be considered. Firstly expression levels varied between mutants therefore total protein calculation could not be used. Standard Bradford assays of the purified sample could not be used due to the variability of the extraction, solubility and overall purification of the mutant extracts. In order to achieve direct comparison between purified mutant extracts it was decided to use SDS-PAGE visualisation. Purified samples were mixed with a standard quantity of BSA. Doubling dilutions of these solutions were loaded and run on SDS-PAGE gels. The proteins were visualised by staining the gels with Coomassie brilliant

blue R250 solution (Section 3.2.3). Purified enzyme was then directly compared to the BSA standards by eye and quantified on this basis (Figure 5.12).
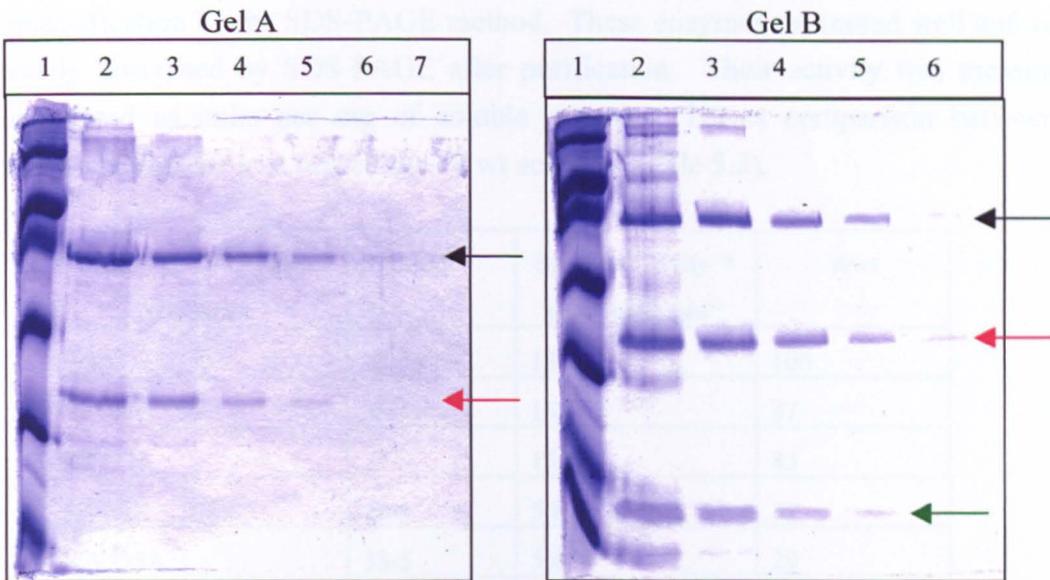
Figure 5.12 Quantification of HSV-1 dUTPase by SDS-PAGE

Gel A shows SDS-PAGE of HSV-1 dUTPase mutant 5F366Y compared to BSA standards. Lanes 2-7 show doubling dilutions of BSA standards (black arrow) compared to purified dUTPase (red arrow). Gel B shows quantification of mutant XT273S by the same method. The green arrow highlights additional bands in this extract indicating a reduction in purity compared to Gel A. This method allows the quantification of both the pure sample, Gel A, and the impure sample, Gel B, by allowing a direct visual comparison of the recombinant dUTPase against the BSA standards.

The SDS-PAGE method allowed direct quantification of soluble HSV-1 dUTPase. The two samples shown in Figure 5.12 have a large variation in overall purity. Determination of total protein content in these two extracts would not give a true indication of the amount of recombinant dUTPase present. This visual method allowed direct quantification and therefore correct calculation of specific activity even when samples had contaminating protein.

5.5.4 Specific activity of HSV-1 dUTPase mutants

The method described above for quantifying recombinant enzyme requires the production of a good quantity of pure samples. Unfortunately with many of the enzyme extracts this was difficult to achieve. The majority of mutants purified by the original two-stage column method could not be quantified by the SDS-PAGE method. Although enzyme activity was measurable and could be calculated in terms of activity per mg of protein, these values were regarded as unreliable. As mentioned previously, the relative amount of recombinant HSV-1 dUTPase compared to total protein was variable between extracts. The new method for purification based on selective buffer

solubility and Mono S chromatography provided enzyme quality much more applicable to this quantification method.

Six HSV-1 dUTPase samples were in a state pure enough to allow quantification by the SDS-PAGE method. These enzymes expressed well and could be easily visualised by SDS-PAGE after purification. Their activity was measured and expressed as units per mg of soluble enzyme. Direct comparison between these enzymes is given as a percentage of wt activity (Table 5.3).

HSV-1 dUTPase construct	Code	Specific activity * μmol mg^{-1} min^{-1}	%wt
WT	18-56	18.7	100
NC76A	36-11	16.3	87
XT273S	27-5	15.9	85
5F366Y	28-5	5.8	31
XW275A	33-5	5.4	29
5F366A	34-6	0.17	0.9

Table 5.3 Activity of purified HSV-1 dUTPase mutants compared to wt
Activities were calculated by measuring the conversion of [^{3}H]dUTP to [^{3}H]dUMP (Section 3.2.6) and quantification of soluble enzyme was performed as in Figure 5.12. In the nomenclature of NC76A, "N" refers to Non-motif since this mutation lies out with any of the conserved motif regions.
* Note specific activity is measured as the amount of dUMP formed per mg of recombinant HSV-1 dUTPase per minute.

The activities of these mutants occupies a wide range from close to wt levels down to less than 1% of wt. Mutant NC76A lies out with the expected active site region and was used as a control while the remaining four mutants were designed to affect what were thought to be key residues in the HSV-1 dUTPase enzyme. These mutations are described graphically in the following Section 5.8.

5.5.5 Discussion

Efficient separation of the mutant dUTPase enzymes from the bacterial host enzyme was achieved by the purification method. *E.coli* dUTPase is highly soluble in the extraction buffers used and would be separated by the chelator fractionation. Further confidence is gained from the Mono S purification stage where the *E.coli* dUTPase would not be retarded on the column. Activities calculated can therefore be attributed directly to the recombinant enzymes.

As proposed with the crude extract results, expression levels between individual mutants varied considerably. In order to study these enzyme preparations on comparable level it was necessary to devise a method of quantification. SDS-PAGE

visualisation allowed not only a direct comparison of the enzymes as a single band, it also provided an indication of purity directly before activity was measured. This later point was of obvious importance since purified enzyme could fall out of solution on storage.

5.6 Truncation of the HSV-1 dUTPase C-terminal region

5.6.1 Introduction

Truncation of the HSV-1 C-terminal region was used as a quick method to test the potential function of the conserved motif 5 region. Data from these experiments preceded the site directed mutagenesis studies and gave the first indication of the importance of the C-terminal tail in the class II dUTPases. A summary of the methods and results is given.

5.6.2 Truncation of HSV-1 UL50 by PCR

Truncation of the C-terminal portion of HSV-1 dUTPase and subsequent deletion of the motif 5 region was accomplished by PCR mutagenesis. Four oligonucleotide primers were designed to amplify two truncated forms of the UL50 ORF (-16 and -32 $\alpha\alpha$) and one full length control.
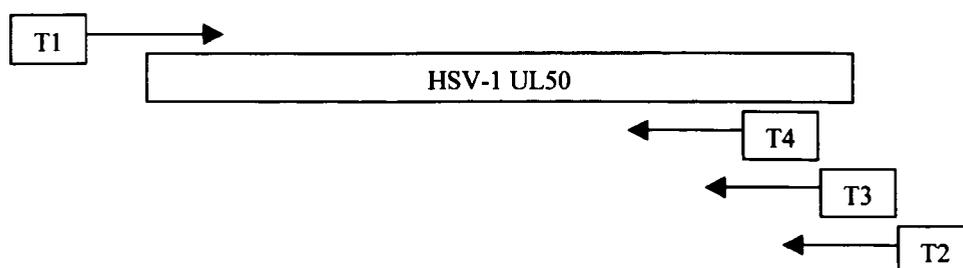

```
T1.   5' (GT)3 CAT ATG AGT CAG TGG GGA TCC GGG GCG 3'
                NdeI

T2.   5' (GT)3 AAG CTT CTA AAT ACC GGT AGA ACC AAA ACC 3'
                HindIII

T3.   5' (GT)3 AAG CTT CTA CTC CGC GTC AAA CTC GTT CGT AAA 3'
                HindIII

T4.   5' (GT)3 AAG CTT CTA TTC GGC GGT TGA GTC CGG AAC ACC 3'
                HindIII
```

Figure 5.13 Oligonucleotide design for HSV-1 UL50 truncations
Four oligonucleotides were designed with (GT)₃ tails for efficient restriction digestion. T1 corresponds to the 5' end of UL50 and incorporates an NdeI site (underlined) at the start ATG codon. T2 corresponds to the 3' end of UL50 with the natural stop codon (bold) followed by a HindIII site. T3 and T4 correspond to the deletion of 16 and 32 C-terminal residues respectively by the insertion of an engineered TAG stop codon into the ORF (reverse complement CTA shown in bold). These oligonucleotides were also flanked by a HindIII restriction site.

PCR products were successfully obtained for the primer pairs T1/T2 full length UL50 (Trunc-0), T1/T3 deletion of 16 residues (Trunc-16) and T1/T4 deletion of 32 residues (Trunc-32). These were digested with NdeI and HindIII, ligated into the vector pET23a and transformed into *E.coli* DHα. Prospective colonies were grown up and plasmids harvested. Restriction digest analysis confirmed that the UL50 fragments had been inserted correctly.

Trunc-16   .....APPSERGTGGFGSTGI (C-term)

Trunc-32   .....PRNPPLLVFTNEFDAEAPPSERGTGGFGSTGI (C-term)

Figure 5.14 Regions deleted by PCR truncations

Figure shows the residues deleted from the HSV-1 dUTPase by the insertion of an engineered stop codon into the UL50 ORF. The position of motif 5 is shown in yellow.

5.6.3   Expression and analysis of truncated constructs

The truncated constructs plus the wt PCR control were transformed into BL21(DE3)pLysS bacteria and expressed according to Section 3.2.1(b). To determine the initial activity of these mutants, dUTPase assays were performed on crude extracts of the two truncated enzymes plus the wt PCR control. No measurable activity was found in either of the truncated proteins Trunc-16 or Trunc-32. The wt HSV-1 dUTPase generated by this PCR method, Trunc-0, exhibited activity corresponding to the original pET23a.

Extracts of Trunc-0 and Trunc-16 were semi-purified using the selective buffer method (Section 5.2.3) and visualised by SDS-PAGE. Bands were identified corresponding to the full length HSV-1 dUTPase for Trunc-0 plus a smaller band corresponding to the Trunc-16 construct.

The position of truncation, Trunc-16, can be visualised in the human dUTPase structure in Section 5.8 (Figure 5.16a). The 11 terminal residues are shown as a wireframe representation and constitute the homologous area that is deleted in the shorter of the two truncations (5 C-terminal residues not visible in human structure). Using this structure as a model for the HSV-1 dUTPase, removal of this portion of the enzyme can be seen to expose the substrate to the solvent. It is likely that a catalytic environment cannot be achieved without motif 5 and therefore accounts for the complete activity loss seen in both truncated HSV-1 dUTPases.

5.6.4   Discussion

Truncation of the C-terminal region was performed before data were available for the motif 5 mutants. Analysis of one such mutant, 5F366A, in Section 5.5.4 demonstrated that it possessed less than 1% of wt activity. Given that this effect was

realised by the substitution of a single amino acid in motif 5, it is not surprising that deletion of the entire motif 5 region resulted in the expression of truncated proteins with no detectable activity.

Although these truncations proved to be crude in the light of the single residue mutations, they were advantageous in highlighting the importance of motif 5 at an early stage. The PCR method employed to generate these truncations by deletion of coding nucleotides at the terminal end of UL50 proved to be very efficient. Using the same strategy it would be of interest to generate more subtle deletions of the C-terminal region ranging from 1 to 15 C-terminal residues. This would be a quick way to investigate how much of the C-terminal region was required for efficient binding and hydrolysis of dUTP. It is likely that a specific truncation would give partial activity. This would be a very interesting mutation to obtain kinetic data by stopped-flow analysis in order to determine the role of this region in binding and/or catalysis. Further experiments were explored with the truncated proteins generated in this study and are detailed in the following Section 5.7.

5.7 Analysis of the C-terminal region with oligopeptides

5.7.1 Introduction
A facility in the MRC Virology Unit existed for the synthesis and purification of oligopeptides. It was decided to utilise this technology to generate peptides for the analysis of the C-terminal region of HSV-1 dUTPase. Previous studies by SDM and PCR truncation had demonstrated that this region was important by modifying the activity of the enzyme (Sections 5.5 & 5.6).

The rationale behind the use of peptides was based on the model created for the HSV-1 dUTPase detailed in Chapter 4. It was envisaged that the C-terminal region acts as a flexible clamp, only adopting defined tertiary structure when the substrate is situated in the binding site. If this was indeed the case it is possible that short peptides corresponding to the C-terminal tail could disrupt this clamping action by competing for the site normally occupied by the tail. It was envisaged that soluble peptides might interact at the active site region containing bound dUTP thus preventing the C-terminal arm adopting the tertiary structure necessary to complete hydrolysis of the dUTP.

5.7.2 Design and synthesis of C-terminal peptides
Five peptides were synthesised corresponding to the C-terminal region of HSV-1 dUTPase. Each peptide was 16 residues in length and overlapped the adjacent peptide by 8 residues (Figure 5.15). The peptides were analysed by HPLC to determine

their purity. Peptides 1, 2 and 4 had initial purities of 92%, 95% and 96% respectively. Peptide 3 (60%) and 5 (69%) were re-purified to over 90% by HPLC. Peptide 3 was found to be insoluble in water and although it could be successfully solubilised in ammonia it was not used in this study. Lysophilised stocks of the other 4 peptides were frozen and reconstituted in water at a concentration of 1mg/ml before use.

	C-terminal region of HSV-1 dUTPase		C
	Peptide 5	APPSE**RGTGGFGS**TGI	
	Peptide 4	FTNEFDAEAPPSE**RGT**	
	Peptide 3	PRNPPLLVFTNEFDAE	
	Peptide 2	GVPDSTAEPRNPPLLV	
	Peptide 1	KALRNYPRGVPDSTAE	
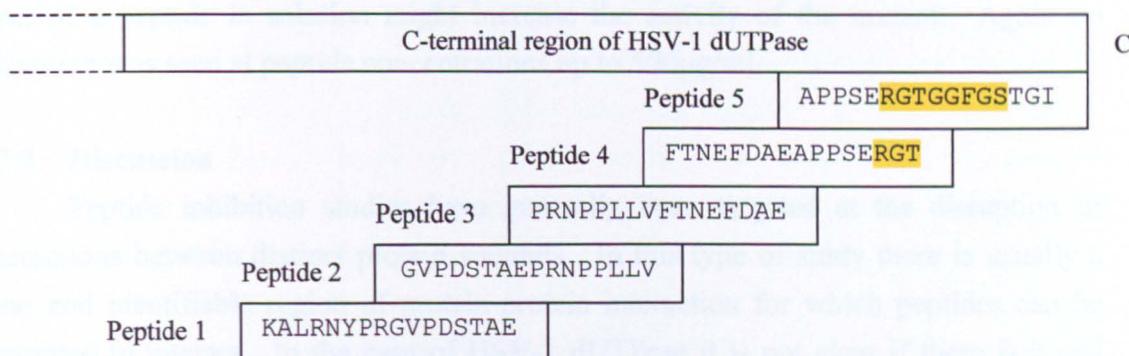
Figure 5.15 Oligopeptide design for HSV-1 UL50 C-terminal region
Five oligopeptides were synthesised corresponding to the primary sequence of the HSV-1 dUTPase C-terminal region. Peptides were 16 residues in length and overlapped the adjacent peptide by 8 amino acids. The motif 5 region is highlighted in yellow showing that the entire motif is present in Peptide 5.

5.7.3 Analysis of peptides with HSV-1 dUTPase

Purified wt HSV-1 dUTPase was used to determine the effect of the generated peptides. Several approaches were used in an attempt to modify enzyme activity by the addition of the peptides in solution. Initially wt enzyme was assayed by standard methods (Section 3.2.6) to give a stable activity of 50% conversion of dUTP to dUMP in 10min at 37°C. The assay was then performed with single peptides added to a final concentration between 0.5-50 μg/ml. No effect was found in these experiments.

Experiments were then performed at much increased peptide concentrations. Purified wt enzyme was diluted to give dUTP conversion values in the region of 10-15%. Single peptides were then added at a concentration of 800μg/ml. Control experiments were carried out without peptide and also with the substitution of peptide with BSA at 800μg/ml. Again no effect was detected with the peptides at this higher concentration.

5.7.4 Analysis of peptides with HSV-1 dUTPase C-terminal truncations

An experiment was designed to test the effect of peptides in truncated versions of the HSV-1 wt. The rationale behind this was based on supplying the truncated enzyme with peptides in solution corresponding to the deleted sections. It was envisaged that if the peptides could mimic the C-terminal section of the HSV-1 dUTPase in solution then they might restore a low level of enzyme activity. Various

concentrations of the mutant Trunc-16 and the peptides were tested but there was no detectable restoration of enzyme activity.

Peptide 5, containing the motif 5 region, was tested in similar experiments combined in solution with mutant 5F366A. This mutant had been purified to such a degree that although it only exhibited 1% of wt activity this could be easily measured in standard reactions. It was thought that the addition of a correct version of motif 5 in the form of a peptide in solution might increase the activity of the mutant. Again no activation was seen at peptide concentrations up to 500μg/ml.

5.7.5 Discussion

Peptide inhibition studies have generally been targeted at the disruption of interactions between distinct protein subunits. In this type of study there is usually a clear and identifiable region of protein-protein interaction for which peptides can be generated to interact. In the case of HSV-1 dUTPase it is not clear if there is a real protein-protein or protein-substrate interaction between the C-terminal region and the active site area. However the crystal structure of the human dUTPase showing the C-terminal arm clamping down on the substrate in the active site merited these peptide experiments. Given the high homology of motif 5 between the human and HSV-1 dUTPases, and indeed all the other class I and II dUTPases, it is not unreasonable to predict a similar mechanism.

The peptides used in these experiments were all 16 amino acids in length. It was hoped that this would allow high quality synthesis yet still provide an adequate length for the formation of a tertiary conformation. From the *E.coli* crystal structure data it is presumed that the C-terminal 16 residues are disordered and only form a defined structure when dUTP is bound in the active site. Since the reaction products of the hydrolysis of dUTP could not leave the active site until this arm becomes disordered again, it is likely that the formation of a tertiary clamping structure is an extremely fleeting event. If a similar case exists in the HSV-1 dUTPase then the probability of a peptide ordering itself over the newly bound substrate before the C-terminal arm has a chance to clamp may be extremely small. This would essentially mean that any modification of enzyme activity would be undetectable by standard methods.

It is unfortunate that these peptide studies did not produce any insight into the mechanism of the HSV-1 C-terminal region especially since previous experiments demonstrated the importance of this region in catalysis. It may prove worthwhile to test these peptides using the more rigorous methods of stopped-flow analysis.

5.8   Interpretation of HSV-1 dUTPase mutant data

5.8.1   Introduction

The results from Chapter 4 on the modelling of the class II dUTPases allow the interpretation of the mutagenesis data in a structural context. The HSV-1 mutant dUTPase data given in the previous sections stem from two distinct target regions: the C-terminal tail containing motif 5 and the class II specific motif X. The specific activity data allow a basic indication of the importance of the mutated residues relative to the reduction in activity compared to the wild type enzyme. It does not however offer any insight into the mechanism of this effect. To address this, the class II modelling data were amalgamated with the experimental mutagenesis data. This section is divided between the motif 5 mutants and the motif X mutants.

5.8.2   Interpretation of the motif 5 mutagenesis data

Chapter 4 demonstrated the likelihood of the class II active site closely mirroring the class I active site. Based on this assumption the class I structural data were used as a basic model for the class II active site. Previous to the publication of the human dUTPase structure there was no data available for the C-terminal region encompassing motif 5. Its role in the function of dUTPase was not clear and the HSV-1 mutagenesis target Phe366 was chosen based on its high sequence conservation between the two classes. In the class I dUTPases the motif 5 Phe is conserved almost completely throughout the group (see class I alignment Figure 4.1). The only notable exception to this is the FIV dUTPase where the motif 5 Phe has been substituted for a Tyr. The motif 5 Phe is also highly conserved in the class II dUTPases. This is seen in the class II alignment shown in Figure 4.7. From the ten species shown only HVS does not carry a Phe in motif 5 where it is substituted by a Leu.

Based on the primary sequence data two mutations were selected at the HSV-1 Phe366 locus. A conservative mutation to Tyr (5F366Y) was made on the basis that this substitution already exists in nature in the class I FIV dUTPase. In order to assess the possible role of a the Phe aromatic ring structure in the functioning of the active site it was decided to remove the entire side chain by substitution to Ala (5F366A). To gain a visual perspective of the likely function of this residue the human dUTPase structure was used to represent the HSV-1 mutations. Figure 5.16 shows the position of the conserved Phe in relation dUDP bound in the active site.
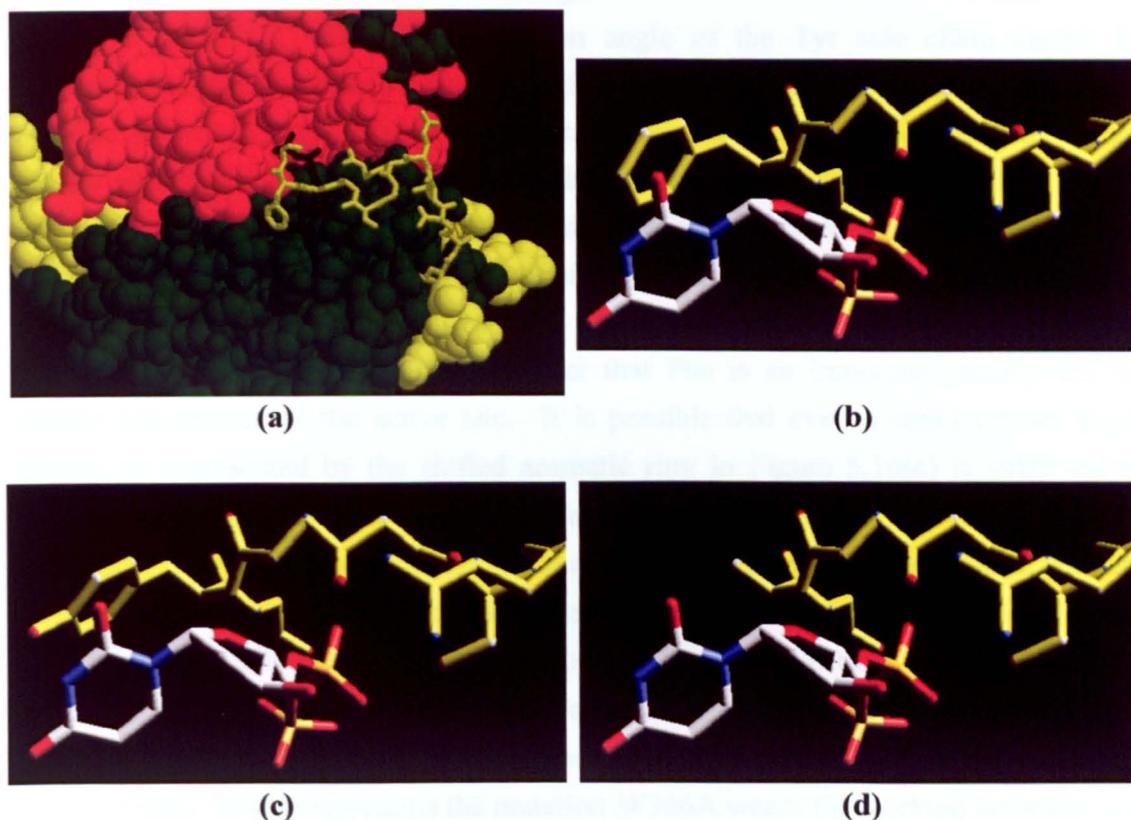
Figure 5.16 The human dUTPase active site as a model for HSV-1 mutagenesis

Four diagrams of the human dUTPase co-crystallised with dUDP. Figure (a) shows the position of the C-terminal tail in relation to the active dUDP. The 11 terminal residues are shown in wireframe representation. Figures (b, c and d) are all shown from the same perspective with the viewer looking out from the active site cavity towards the bound substrate. Figure (b) shows the wild type situation where the C-terminal tail (yellow) reaches over to cap the active site. The side chain of the motif 5 residue Phe135 (equivalent to HSV-1 Phe366) can be seen stacking over the uracil ring of dUDP. Figure (c) shows the mutation of this Phe to a Tyr (equivalent to HSV-1 mutation 5F366Y) and (d) to an Ala (equivalent to HSV-1 5F366A).

The human dUTPase structure is a useful model given the main section of the C-terminal tail, including motif 5, becomes ordered in the crystal structure with bound dUDP. The terminal 5 residues remain unordered but since the final residue, Gly136, constitutes the C-terminal span of motif 5, the structure encompasses the mutation sites. Figure 5.16(a) shows the position of the C-terminal tail relative to the substrate. Figure (b) shows the side chain of Phe135 (top left in yellow) acting as a 'Phe-lid' by stacking on the uracil ring of dUDP. It was proposed by C. Mol *et al.* (1996) that this conformation may hold the uracil in the appropriate position for catalysis while excluding water from the active site. In support of the importance of this interaction the same group reported that mutation of Phe135 to an Ala in the human dUTPase resulted in greatly reduced activity. Based on the class II model in Chapter 4 it is presumed that this conserved motif 5 Phe has a homologous function in the HSV-1 active site.

Figure 5.16(c) represents the mutagenic substitution of the HSV-1 motif 5 Phe to a Tyr (5F366Y). The precise torsion angle of the Tyr side chain cannot be determined in this model but it is proposed to stack on the uracil ring in a homologous position to the wt Phe. In physicochemical terms the difference between the wt and mutant side chains is the addition of a hydroxyl group in Tyr compared to Phe. Since the aromatic ring is conserved, only minimal disruption to enzyme activity was predicted by this substitution. It was therefore surprising that the 5F366Y mutation resulted in an activity drop of almost 70% compared to wt. There are several possible mechanisms for this reduction. It is clear that Phe is an important residue for the correct functioning of the active site. It is possible that even a small torsion angle change as represented by the shifted aromatic ring in Figure 5.16(c) is sufficient to incorrectly position the uracil ring of the substrate and therefore reduce catalytic efficiency. The additional hydroxyl group of Tyr may also create a steric clash with the surrounding cavity edge and disrupt ordering of the tail. Given that the hydroxyl group of Tyr is potentially reactive it is also possible that the hydroxyl may participate in hydrogen bonding and hinder the release of cleaved substrate. This latter point is of potential significance since the reaction product dUMP is a competitive inhibitor.

Figure 5.16(d) represents the mutation 5F366A where the stacking aromatic ring structure is removed completely. The resultant activity drop to under 1% of wt was expected in the light of the substantial activity reduction in the more conservative Tyr substitution. Loss of the aromatic ring most likely results in positional instability of the substrate due to the absence of the uracil moiety clamping mechanism. Figure 5.16(a) shows the position of the conserved Phe in the human active site viewed from the solvent. From this angle it can be visualised that removal of the bulky Phe side chain may also increase the likelihood of solvent penetration into the active site. In fact what is most interesting about this mutation is that it results in an enzyme with any activity at all. This supports the role of the conserved Phe as a secondary effector and not as part of the cleavage mechanism. Overall these data highlight the role of motif 5 in capping the active site and creating an efficient catalytic environment.

5.8.3 Interpretation of motif X mutagenesis data

Mutagenesis of the motif X region was based on primary sequence conservation within the class II dUTPases. The motif X alignment in Figure 4.23 demonstrates that HSV-1 residue Trp275 is totally conserved between the ten class II species shown. Since the functional significance of motif X is unknown it was decided to substitute Trp275, the largest conserved side chain in motif 5, to an Ala. The thinking behind this mutation was to disrupt any function of motif X while maintaining the correct conformation of the dUTPase active site region.

The other target in the motif X region was Thr273 which is conserved among seven of the ten class II species in alignment Figure 4.23. This residue is substituted by Val, Cys and Ser in EHV-2, EBV and HVS respectively. The mutation XT273S was chosen in order to substitute the conserved Thr273 with the residue Ser already found in nature. This target was selected in order to make a more subtle change to motif X. There was a potential hazard with making a large substitution, as in Trp275 to Ala, since the structural significance of the residue could not be determined. Motif X lies between motifs 4 and 5 and therefore any dramatic local structural change could potentially cause misfolding. Based on these assumptions it was decided to select one radical change (XW275A) and one conservative change (XT273S) for comparison.

Motif X is specific to the class II dUTPases and therefore cannot be visualised using the class I structures. In order to get some representation of the function of motif X and its relation to the mutagenesis data it is necessary to utilise the class II model. Figure 5.17(a) shows the predicted position of motif X (coloured red) in the class II model. Secondary structure predictions using the Threader program demonstrated that this motif region is structurally related to motif 3 (Section 4.7.4) and is shown as such in the model. Figure 5.17(b) gives a more detailed representation of the Swiss-Model ProMod prediction of the motif X loop region (Section 4.7.5) and the position of the two mutation targets XT273S and XW275A.

(a)   (b)
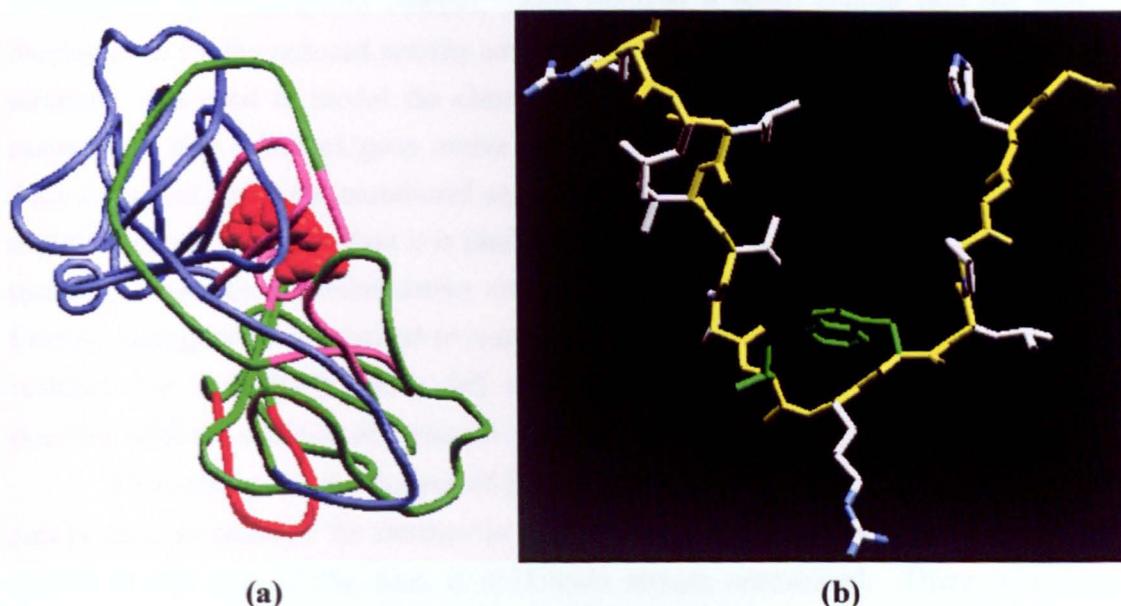
Figure 5.17 Graphical visualisation of the HSV-1 motif X mutations
Figure (a) shows the class II model with the position of motif X highlighted in red. Note that it is in the opposite side of the molecule from the active site region. Figure (b) shows the motif X region as predicted by the Swiss-Model ProMod program. The backbone is highlighted in yellow with the side chains of Thr273 and Trp275 highlighted in green.

The data gained from the Threader and Swiss-Model programs in Section 4.7 provide some evidence that the motif X region forms a loop structure. Using these

predictions directly the position of the mutations can be superimposed on a model loop. The Swiss-Model prediction for the motif X region is shown in Figure 5.17(b). In this model, residue Trp275 is situated at the bottom of the loop in the equivalent position to Tyr at the base of the motif 3 loop. The residue Thr273 would therefore occupy a position on one of the β-strands close to the bottom of the loop. As mentioned in Section 4.7, the predicted motif X loop is quite similar to the class I motif 3 loop (Figure 4.4) with a bulky ring side chain pointing inwards at the base of the loop.

The mutagenesis data demonstrate an activity reduction of 15% for XT273S and 70% for XW275A compared to wt HSV-1 dUTPase. This result is consistent with a dramatic residue substitution and a conservative substitution respectively. Interpretation of these results in the structural model is difficult. It is however interesting that the removal of the largest side chain in motif X results in an enzyme which still produces 30% of wt activity. Comparing this to the motif 5 mutations above suggests that motif X does not play any role in the conversion of dUTP to dUMP. The reason for the activity reductions may be due to minor alterations in the folding of the protein although further investigation is necessary.

5.8.4 Discussion

The data gained from the mutational analysis of the HSV-1 dUTPase were investigated in a structural context. This allowed a basic insight into the possible mechanisms for the reduced activity seen in these mutants. The class I human dUTPase structure was used to model the class II motif 5 mutations. This was based on the assumption that a homologous active site region is shared between the two classes. This is backed up by the mutational analysis of the highly conserved Phe residue in the motif 5 region. From the data it is likely that the HSV-1 C-terminal arm, encompassing motif 5, possesses a similar active site capping mechanism to the class I dUTPases. Further mutagenesis is required to test the functions of other highly conserved motif 5 residues but it is likely that a full understanding of this region will only become possible with the addition of a class II crystal structure.

This latter point is also true of the class II specific motif X. Although modelling can be used to enhance the mutagenic data, without a structure or defined function the details of this area of the class II dUTPases remain unresolved. There is not yet a sufficient quantity of mutagenesis data to determine if the motif X region is a separate functional locus from the dUTP hydrolysis site. The data show however that substitution of the highly conserved Trp residue, constituting the removal of the largest side chain in this motif, results in an active enzyme possessing 30% of wt activity. It is tempting to suggest therefore that the motif X region has a separate function, but again further investigation is required.

Chapter 6 - Final Discussion

The dUTPase of HSV-1 belongs to a distinct structural subset encoded specifically by the α- and γ-herpesviruses. Although four members from the class I dUTPase trimers have been crystallised there is no available structure for any member of the class II monomers. Attempts to crystallise the HSV-1 dUTPase have been severely compromised by the poor solubility of the recombinant enzyme expressed in *E.coli*. Several groups are now involved in recombinant expression and crystallisation trails of VZV and EBV dUTPases.

In order to investigate the HSV-1 dUTPase it was decided that a structural model was necessary. Since no structural data was available for the class II dUTPases the model was based on the evolutionary relationships with the class I enzymes. The model therefore encompasses the class I structures of *E.coli*, FIV, EIAV and human to arrive at a general structural arrangement for the class II dUTPases. Analysis of the class I enzymes allowed not only the characterisation of elements which produced overall tertiary structure but also the functional roles of specific residues at the active site. The structural similarity within a group containing dUTPases from such diverse species was startling especially since no other enzyme has yet been described with an active site consisting of regions from all three subunits of a trimer. It can be concluded that these enzymes have evolved to maximise structural integrity, catalytic efficiency and substrate specificity. The class I dUTPase therefore provides a strong basis for the modelling of the class II enzymes which are most likely variants which have diverged from an ancestor common to both classes.

The class II dUTPases proved to be substantially heterogeneous compared to the class I enzymes and the model had to be flexible enough to accommodate these differences. Analysis at the primary sequence level provided a method for directly comparing the two classes. This was achieved by generating a class I doublet sequence and applying different alignment methods. Secondary structure predictions revealed that the overall structure of the class II monomer approximated two copies of the class I subunit structure joined as a single protein chain. This led to the hypothesis that the intragenic duplication had effectively duplicated the class I subunit structure to produce a functionally active monomer. This model was supported by hydrophobic modelling data. An evolutionary pathway was described whereby dUTPase functionality could be maintained utilising a dimer intermediate. Although this model satisfies the criteria for assembly of a functional active site it does not directly address the question as to why the herpesviruses have adopted a different structural arrangement compared to the diverse class I species.

Investigation into a conserved region between the class II dUTPases and the β-herpesvirus homologues (motif X) provides a possible solution. From a general energetic point of view it appears that the provision of a trimer with three active sites is more favourable that a monomer representing the mass two subunits generating only one active site. Additionally it is noted that the β-herpesvirus do not encode their own dUTPase and are presumed to utilise the cellular class I enzyme. As previously discussed, the replicating environment is presumably critical in terms of the necessity for a virally encoded enzyme. What is interesting is that not only do the α- and γ-herpesviruses not utilise the cellular class I dUTPase, or encode a class I enzyme like several other virus families, but they encode a substantially variant form. The properties of this enzyme that make it evolutionarily more advantageous remain unclear. What is apparent is that the class II variant takes up double the genome space and is predicted to be less energetically favourable that the class I version.

It is possible that the class II dUTPases perform an unknown function specifically related to the replication of these herpesviruses. It is also possible that the motif X region in the β-herpesvirus dUTPase homologues represents a functional site in a protein without dUTPase activity. If this was indeed the case the class II molecule could potentially possess dual functionality, satisfying the conservation pressure inherent in the double length monomer. To investigate this theory further, this region was targeted for modelling and mutagenesis studies.

The absence of any class II structural data make local prediction difficult. The three methods used (secondary structure prediction, protein fold threading and primary sequence homology modelling) arrived at a general consensus but this is not enough in itself to predict potential function. Motif X was predicted to be structurally related to the ancestral motif 3 despite the almost complete lack of primary sequence similarity. In the dUTPase active site, the motif 3 loop is the basis for nucleotide binding and specificity. If there was an additional function attributable to motif X, nucleotide binding would be a potential candidate. Mutagenesis of this region modulated dUTPase activity relative to the severity of the residue substitution and although a structural prediction exists, the effect of these mutations on general folding were not determined. Without further mutagenesis data, and ideally a class II crystal structure, the disassociation of the motif X region from dUTPase activity cannot be substantiated.

Investigation of the replication machinery of HSV-1 has revealed many interactions between specific enzymes. An extension of this localisation is seen in bacteriophage T4 in which enzymes involved in nucleotide metabolism are linked in a dNTP synthesising complex which may in turn be linked directly to the enzymes localised at the DNA replication fork. The advantages of such co-localisation are clear, generating DNA building blocks for replicative enzymes in a confined physical space.

An evolutionary step forward could be envisaged as a single enzyme involved in nucleotide metabolism with dual functionality. Future work involving a more detailed analysis of the motif X region in the class II enzymes would prove interesting, especially if paralleled with investigation into the β-herpesvirus homologues. Even if the class II dUTPases do not posses a secondary function it would be interesting to reveal the function of the β-herpesvirus homologues.

The mutagenesis studies in this course of study highlight several potential areas of homologous function between the class I and class II enzymes. The class II model predicts an overall structural consensus for the active site regions between the two classes. Truncation of the C-terminal arm containing motif 5 demonstrated that this region was necessary for the function of the HSV-1 dUTPase. Site-directed mutagenesis allowed these data to be refined by identifying the importance of a key residue which is highly conserved between the class I and class II dUTPases. Phe 366, the HSV-1 equivalent of Phe 135 in the human dUTPase, is predicted to provide a homologous function to the class I enzymes. Even conservative substitution of the residue produced a dramatic reduction in catalytic activity. The position of this mutation, only six residues from the C-terminus, gives some confidence that it is not disrupting the overall folding of the enzyme. It is likely that this residue is conserved for its capping function, interacting directly with the uracil moiety of the bound substrate. It would be interesting to apply the circular dichroism (CD) techniques used by Vertessy *et al.* (1998) to investigate the conformational shift in the C-terminal arm of the *E.coli* dUTPase during substrate binding. Utilising CD techniques, conformational change of the HSV-1 C-terminal arm induced by substrate binding could be determined. Furthermore, using the HSV-1 mutants, it could be tested whether this mechanism was disrupted by mutation of a single motif 5 residue.

The existence of a substrate capping mechanism homologous to that revealed in the class I dUTPases prompted the investigation into peptide inhibitors. The potential for conformational disruption of the C-terminal arm structure was investigated. Short peptides corresponding to the HSV-1 C-terminal sequence were combined in solution with HSV-1 dUTPase under varying conditions. No inhibition was found, which may reflect the inability of the peptides to adopt a structural conformation capable of disrupting the capping mechanism. However, peptide inhibitors have traditionally been used to disrupt subunit interaction rather than conformational folding within a single protein species. There are energetic barriers which must be considered since the local protein chain is competing with a peptide in solution. In the absence of a class II structure it would be interesting to apply this experiment to the class I human dUTPase. The structural information available for the human dUTPase C-terminal arm would allow a more defined approach to the design of peptide inhibitors for this region.

The original experimental strategy combining site-directed mutagenesis with crude assay analysis proved to be inappropriate for the analysis of HSV-1 dUTPase. This substantially reduced the number of recombinant enzymes which could be tested. The original strategy was compromised by the inherent insolubility of the recombinant enzyme expressed in bacterial systems. This unfortunate property of the HSV-1 enzyme has now been well documented by a number of groups, not least those involved in crystallisation trials. It is possible that continued site-directed mutagenesis studies will eventually produce a recombinant mutant with a higher solubility although crystallisation of a native class II enzyme is more valued. It may be useful in parallel experiments to clone and express the HSV-2 dUTPase. Although this enzyme has high sequence conservation with the HSV-1 dUTPase, this does not preclude the possibility that it may be more soluble in bacterial expression systems. The advantages of a highly soluble class II dUTPase for general experimental analysis are clear, allowing easier expression, extraction, purification and kinetic analysis.

Overall, the analysis of the HSV-1 dUTPase proved to be a useful model for the class II enzymes. The modelling work has allowed a more in-depth evaluation of the evolutionary relationships between the two classes. Although the mutagenesis studies did not yield as much information as was initially anticipated, the common functionality predicted between the two dUTPase classes by structural modelling was supported.

References

Ablashi, D. V., Berneman, Z. N., Kramarsky, B., Whitman, J., Asano, Y. & Pearson, G. R. (1995). Human herpesvirus-7 (HHV-7) - Current Status. *Clinical and Diagnostic Virology* **4**, 1-13.

Ackermann, M., Braun, D. K., Pereira, L. & Roizman, B. (1984). Characterisation of herpes simplex virus 1 α Proteins 0, 4, and 27 with monoclonal antibodies. *Journal of Virology* **52**, 108-118.

Adams, R. L., Burdon, R. H., Campbell, A. M., Leader, D. P. & Smellie, R. M. (1981). Replication of DNA. In *The Biochemistry of the Nucleic Acids*, 9th edn. London, New York: Chapman and Hall.

Albrecht, J. C., Nicholas, J., Biller, D., Cameron, K. R., Biesinger, B., Newman, C., Wittmann, S., Craxton, M. A., Coleman, H., Fleckenstein, B. & Honess, R. W. (1992). Primary structure of the herpesvirus saimiri genome. *Journal of Virology* **66**, 5047-5058.

Alford, C. A., Stagno, S., Pass, R. F. & Britt, W. J. (1990). Congenital and perinatal cytomegalovirus infections. *Reviews of Infectious Diseases* **12**, 745- 753.

Altschul, S. F., Gisg, W., Miller, W., Myres, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Amann, E. J., Brosius, J. & Ptashne, M. (1983). Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *E.coli. Gene* **25**, 167-178.

Anagnostopoulos, I., Herbst, H., Niedobitek, G. & Stein, H. (1989). Demonstration of monoclonal EBV genomes in Hodgkin's disease and Ki-1 positive anaplastic large cell lymphoma by combined southern blot and *In situ* hybridisation. *Blood* **74**, 810-816.

Bachenheimer, S. L. & Roizman, B. (1972). Ribonucleic acid synthesis in cells infected with herpes simplex virus 1. Polyadenylic acid sequences in viral messenger ribonucleic acid. *Journal of Virology* **10**, 875-879.

Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S. & Barrell, B. G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207-211.

Baradaran, K., Dabrowski, C. E. & Schaffer, P. A. (1994). Transcriptional analysis of the region of the herpes simplex virus type-1 genome containing the UL8, UL9, and UL10 genes and identification of a novel delayed-early gene product, OBPC. *Journal of Virology* **68**, 4251-4261.

Barker, D. E. & Roizman, B. (1990). Identification of three genes nonessential for growth in cell culture near the right terminus of unique sequences of the long component of herpes simplex virus 1. *Virology* **177**, 684-691.

Barker, D. E. & Roizman, B. (1992). The unique sequence of the herpes simplex virus 1 L component contains an additional translated open reading frame designated UL49.5. *Journal of Virology* **66**, 562-566.

Barnard, E. C., Brown, G. & Stow, N. D. (1997). Deletion of the herpes simplex virus type 1 UL8 protein: effect on DNA synthesis and ability to interact with and influence the intracellular localization of the UL5 and UL52 proteins. *Virology* **237**, 97-106.

Barnett, B. C., Dolan, A., Telford, E. A. R., Davison, A. J. & McGeoch, D. J. (1992). A novel herpes simplex virus gene (UL49a) encodes a putative membrane protein with counterparts in other herpesviruses. *Journal of General Virology* **73**, 2167-2171.

Batterson, W., Furlong, D. & Roizman, B. (1983). Molecular genetics of herpes simplex virus. VII. Further characterisation of a temperature-sensitive mutant defective in release of viral DNA and in other stages of the viral reproductive cycle. *Journal of Virology* **45**, 397-407.

Baylis, S. A., Purifoy, D. J. & Littler, E. (1989). The characterization of the EBV alkaline deoxyribonuclease cloned and expressed in *E.coli*. *Nucleic Acids Research* **17**, 7609-7622.

Beck, W. R., Wright, G. E., Nusbaum, N. J., Chang, J. D. & Isselbacher, E. M. (1986). Enhancement of methotrexate cyto-toxicity by uracil analogs that inhibit deoxyuridine triphosphate nucleotidohydrolase (dUTPase) activity. *Advances in Experimental Medicine and Biology* **195**, 97-104.

Beck, W. S., Chang, J. D. & Vilpo, J. A. (1984). Uracil misincorporation into DNA and the cyto-toxicity of methotrexate - Importance of dUTPase levels. *Clinical Research* **32**, A550-A550.

Beck, W. S., Wright, G. E., Nusbaum, N. J. & Isselbacher, E. M. (1985). Enhancement of mtx cyto-toxicity my uracil analogs that inhibit cellular dUTPase activity. *Pediatric Research* **19**, 745-745.

Bennet, S. E., Schimerlik, M. I. & Mosbaugh, D. W. (1993). Kinetics of the uracil-DNA glycosylase/inhibitor protein association. *Journal of Biological Chemistry* **268**, 26879-26885.

Ben-Zeev, A. & Becker, Y. (1977). Requirements of host cell RNA polymerase II in the replication of herpes simplex virus in alpha-amanitin sensitive and resistant cell lines. *Virology* **76**, 246-253.

Bergman, A.-C. (1997). Viral dUTPases. In *Department of Biochemistry*. Sweden: Lund University.

Bergman, A. C., Bjornberg, O., Nord, J., Nyman, P. O. & Rosengren, A. M. (1994). The protein P-30, encoded at the *gag-pro* junction of mouse mammary tumour virus, is a dUTPase fused with nucleocapsid protein. *Virology* **204**, 420-424.

Bergman, A.-C., Nyman, P. O. & Larsson, G. (1997). Kinetic characterisation of dUTPase from herpes simplex virus type 1. *manuscript*.

Bernard, J. & Mercier, A. (1993). Sequence of 2 EcoRI fragments from salmonis herpesvirus-2 and comparison with ictalurid herpesvirus-1. *Archives of Virology* **132**, 437-442.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimaouchi, T. & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112**, 535-542.

Bertani, L. E., Haggmark, A. & Reichard, P. (1961). Synthesis of pyrimidine deoxyribonucleoside diphosphates with enzymes from *Escherichia coli*. *Journal of Biological Chemistry* **236**, 67-68.

Bischofberger, N. & Wagner, R. W. (1992). Antisense approaches to antiviral agents. *Seminars in Virology* **3**, 57-66.

Björnberg, O., Bergman, A.-C., Rosengren, A. M., Persson, R., Lehman, I. R., Nyman, P. O. (1993). dUTPase from herpes simplex virus type 1; Purification from infected green monkey kidney (Vero) cells and from overproducing *Escherichia coli* strain. *Protein expression and purification* **4**, 149-159.

Björnberg, O. & Nyman, P. O. (1996). The dUTPases from herpes simplex virus type 1 and mouse mammary tumor virus are less specific than the *Escherichia coli* enzyme. *Journal of General Virology* **77**, 3107-3111.

Blackburn, G. M. & Gait, M. J. (1995). Drug inhibition of nucleic acid biosynthesis. In *Nucleic Acids in Chemistry and Biology*, 2nd edn. Oxford: Oxford University Press.

Booy, F. P., Newcomb, W. W., Trus, B. L., Brown, J. C., Baker, T. S. & Steven, A. C. (1991). Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* **64**, 1007-1015.

Bossemeyer, D. (1994). The glycine-rich sequence of protein kinases: a multifunctional element. *TIBS* **19**, 201-205.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein dye binding. *Anal. Biochem*istry **72**, 248-254

Brandimarti, R., Huang, T. M., Roizman, B. & Campadelli-Fiume, G. (1994). Mapping of herpes simplex virus 1 genes with mutations which overcome host restrictions to infection. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5406-5410.

Britt, W. J. & Alford, C. A. (1993). Cytomegalovirus - Clinical symptoms associated with HCMV infection. In *Fields Virology*, 3rd edn. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott-Raven Publishers.

Broyles, S. S. (1993). Vaccinia virus encodes a functional dUTPase. *Virology* **195**, 863-865.

Buckmaster, A. E., Scott, S. D., Sanderson, M. J., Boursnell, M. E. G., Ross, N. L. J. & Binns, M. M. (1988). Gene sequence and mapping data from Mareks disease virus and herpesvirus of turkeys - Implications for herpesvirus classification. *Journal of General Virology* **69**, 2033-2042.

Calder, J. M. & Stow, N. D. (1990). Herpes simplex virus helicase-primase: the UL8 protein is not required for DNA-dependent ATPase and DNA helicase activities. *Nucleic Acids Research* **18**, 3573-3578.

Camacho, A., Arrebola, R., Pena-Diaz, J., Ruiz-Perez, L. M. & Gonzalez-Pacanowska, D. (1997). Description of a novel eukaryotic deoxyuridine 5'-triphosphate nucleotidohydrolase in *Leishmania major*. *Biochemistry Journal* **325**, 441-447.

Campbell, M. E. M, Palfreyman, J. W. & Preston, C. M. (1984). Identification of herpes simplex virus DNA sequences which encode a *trans*-acting polypeptide responsible for stimulation of immediate early transcription. *Journal of Molecular Biology* **180**, 1-19.

Canman, C. E., Lawrence, T. S., Shewach, D. S., Tang, H. Y. & Maybaum, J. (1993). Resistance to fluorodeoxyuridine-induced DNA damage and cytotoxicity correlates with an elevation of deoxyuridine triphosphatase activity and failure to accumulate deoxyuridine triphosphate. *Cancer Research* **53**, 5219-5224.

Canman, C. E., Radany, E. H., Parsels, L. A., Davis, M. A., Lawrence, T. S. & Maybaum, J. (1994). Induction of resistance to fluorodeoxyuridine cytotoxicity and DNA damage in human tumour cells by expression of *Escherichia coli* deoxyuridine triphosphatase. *Cancer Research* **54**, 2296-2298.

Caradonna, S. J. & Adamkiewicz, D. M. (1984). Purification and properties of the deoxyuridine triphosphate nucleotidohydrolase enzyme derived from HeLa S3 cells - Comparison to a distinct dUTP nucleotidohydrolase induced in herpes simplex virus-infected HeLa S3 cells. *Journal of Biological Chemistry* **259**, 5459-5464.

Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). Crystal structure of a dUTPase. *Nature* **355**, 740-743.

Challberg, M. D. (1986). A method for identifying the viral genes required for herpesvirus DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 9094-9098.

Challberg, M. D. (1991). Herpes simplex virus DNA replication. *Seminars in Virology* **2**, 247-256.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865-1869.

Chang, Y. & Moore, P. S. (1996). Kaposi's sarcoma (KS)-associated herpesvirus and its role in KS. *Infectious Agents and Disease-Reviews Issues and Commentary* **5**, 215-222.

Chatis, P. A. & Crumpacker, C. S. (1992). Resistance of herpesviruses to antiviral drugs. *Antimicrobial Agents and Chemotherapy* **36**, 1589-1595.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M. & Barrell, B. G. (1990). Analysis of the protein coding content of the sequence of human cytomegalovirus strain AD169. *Current Topics in Microbiology and Immunology* **154**, 125-169.

Chee, M. S., Lawrence, G. L. & Barrell, B. G. (1989). Alpha-, beta- and gammaherpesviruses encode a putative phosphotransferase. *Journal of General Virology* **70**, 1151-1160.

Cheng, Y.-C., Chen, J. Y., Hoffmann, P. J. & Glaser, R. (1980). Studies on the activity of DNase associated with the replication of the Epstein-Barr virus. *Virology* **100**, 334-338.

Chou, J. & Roizman, B. (1986). The terminal a sequence of the herpes simplex virus genome contains the promoter of a gene located in the repeat sequences of the L-component. *Journal of Virology* **57**, 629-637.

Chou, J. & Roizman, B. (1990). The herpes simplex virus 1 gene for ICP34.5, which maps in inverted repeats, is conserved in several limited passage isolates but not in strain 17syn+. *Journal of Virology* **64**, 1014-1020.

Clements, J. B., Watson, R. J. & Wilkie, N. M. (1977). Temporal regulation of herpes simplex virus type 1 transcription: location of transcripts in the viral genome. *Cell* **12**, 275-285.

Cohen, D. M. (1992). Molecular aspects of anti-herpesvirus drugs. *Seminars in Virology* **3**, 3-12.

Cohen, G. H. (1972). Ribonucleotide reductase activity of synchronised KB cells infected with HSV. *Journal of Virology* **9**, 408-418.

Cohen, S. S. (1968). From viral pyrimidines to virus-induced acquisition of function. In *Virus induced enzymes*, pp. 26-58. New York: Columbia University Press.

Conner, J., Furlong, J., Murray, J., Meighan, M., Cross, A., Marsden, H. & Clements, J. B. (1993). Herpes simplex virus type-1 ribonucleotide reductase large subunit: regions of the protein essential for subunit interaction and dimerization. *Biochemistry* **32**, 13673-13680.

Cordonnier, A., Casella, J.-F. & Heidmann, T. (1995). Isolation of novel human endogenous retrovirus-like elements with foamy virus related *pol* sequence. *Journal of Virology* **69**, 5890-5897.

Corey, L. & Spear, P. G. (1986). Infections with herpes simplex viruse 1. *New England Journal of Medicine* **314**, 686-691.

Costa, R. H., Draper, K. G., Kelly, T. J. & Wagner, E. K. (1985). An unusual spliced HSV type 1 transcript with a sequence homology to Epstein-Barr virus DNA. *Journal of Virology* **54**, 317-328.

Costanzo, F., Campadelli-Fiume, G., Foa-Tomas, L. & Cassi, E. (1977). Evidence that herpes simplex DNA is transcribed by cellular RNA polymerase II. *Journal of Virology* **21**, 996-1001.

Crumpacker, C. S., Schnipper, L. E., Marlow, S. I., Kowalsky, P. N., Hershey, B. J. & Levin, M. J. (1982). Resistance of antiviral drugs of herpes simplex virus isolated from a patient treated with acyclovir. *New England Journal of Medicine* **306**, 343-346.

Crute, J. W., Tsurumi, T., Zhu, L., Weller, S. K., Olivo, P. D., Challberg, M. D., Mocarrski, E. S. & Lehman, I. R. (1989). Herpes simplex virus 1 helicase-primase: a complex of three herpes virus encoded gene products. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2186-2189.

Curtin, N. J., Harris, A. L. & Aherne, G. W. (1991). Mechanism of cell death following thymidylate synthase inhibition: 2'-deoxyuridine-5'-triphosphate accumulation, DNA damage, and growth inhibition following exposure to CB3717 and dipyridamole. *Cancer Research* **51**, 2346-2352.

Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. & Neuhard, J. (1992). Characterisation of the *Escherichia coli* codba operon encoding cytosine permease and cytosine deaminase. *Molecular Microbiology* **6**, 1335-1344.

Darby, G., Sailsbury, H. & Field, H. J. (1981). Altered substrate specificity of herpes simplex virus thymidine kinase confers acyclovir-resistance. *Nature* **289**, 81-83.

Davison, A.J. (1984). Structure of the genome termini of varicella-zoster virus. *Journal of General Virology* **65**, 1969-1977.

Davison, A. J. (1998). The genome of salmonid herpesvirus 1. *Journal of virology* **72**, 1974-1982.

Davison, A. J. (1992). Channel catfish virus - a new type of herpesvirus. *Virology* **186**, 9-14.

Davison, A. J. & Scott, J. E. (1986). The complete DNA sequence of varicella-zoster virus. *Journal of General Virology* **67**, 1759-1816.

Davison, A. J. & Taylor, P. (1987). Genetic relations between varicella-zoster virus and Epstein-Barr virus. *Journal of General Virology* **68**, 1067-1079.

Dolan, A., McKie, E., MacLean, A. R. & McGeoch, D. J. (1992). Status of the ICP34.5 gene in herpes simplex virus type-1 strain 17. *Journal of General Virology* **73**, 971-973.

Dolan, A., Jamieson, F. E., Cunningham, C., Barnett, B. C. & McGeoch, D. J. (1998). The genome sequence of herpes simplex virus type 2. *Journal of Virology* **72**, 2010-2021.

Dorsky, D. I. & Crumpacker, C. S. (1988). Expression of the herpes simplex virus type 1 DNA polymerase gene by *in vitro* translation and effects of gene deletions on activity. *Journal of Virology* **62**, 3224-3232.

Dotti, C. G. & Simons, K. (1990). Polarised sorting of viral glycoproteins to the axon and dendrites of hippocampal neurons in culture. *Cell* **62**, 63-72.

Dudek, M. J. & Ponder, J. W. (1995). Accurate modeling of the intramolecular electrostatic energy of proteins. *Journal of Computational Chemistry* **16**, 791-816.

Duker, N. J. & Grant, C. L. (1980). Alterations in the levels of deoxyuridine triphosphatase, uracil-DNA glycosylase and AP endonuclease during the cell cycle. *Experimental Cell Research* **125**, 493-497.

Dunham, L. F. & Price, A. R. (1974). Deoxythymidine triphosphate-deoxyuridine triphosphate nucleotidohydrolase induced by *Bacillus subtilis* bacteriophage φe. *Biochemistry* **13**, 2667-2672.

Dutia, B. M., Frame, M. C., Subak-Sharpe, J. H., Clark, W. N. & Marsden, H. S. (1986). Specific inhibition of herpesvirus ribonucleotide reductase by synthetic peptides. *Nature* **321**, 439-441.

Elder, J. H., Lerner, D. L., Hasselkuslight, C. S., Fontenot, D. J., Hunter, E., Luciw, P. A., Montelaro, R. C. & Phillips, T. R. (1992). Distinct subsets of retroviruses encode dUTPase. *Journal of Virology* **66**, 1791-1794.

El-Hajj, H. H., Wang, L. & Weiss, B. (1992). Multiple mutant of *E. coli* synthesising virtually thymineless DNA during limited growth. *Journal of Bacteriology* **174**, 4450-4456.

El-Hajj, H. H., Zhang, H. & Weiss, B. (1988). Lethality of a *dut* (deoxyuridine triphosphatase) mutation in *Escherichia coli. Journal of Bacteriology* **170**, 1069-1075.

Elias, P., O'Donnell, M. E., Mocarski, E. D. & Lehman, I. R. (1986). A DNA binding protein specific for an origin of herpes simplex virus type 1. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 6322-6326.

Epstein, M. A. (1962). Observations on the mode of release of herpes virus from infected HeLa cells. *Journal of Cell Biology* **12**, 589-597.

Feng, D. -F. & Dolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* **25**, 351-360.

Field, H. J. & Wildy, P. (1978). The pathogenicity of the thymidine kinase-deficient mutants of herpes simplex virus in mice. *J. Hyg.* **81**, 267-277.

Fisher, F. B. & Preston, V. G. (1986). Isolation and characterisation of herpes simplex virus type-1 mutants which fail to induce dUTPase activity. *Virology* **148**, 190-197.

Focher, F., Mazzarello, P., Verri, A., Hubscher, U. & Spadari, S. (1990). Activity profiles of enzymes that control the uracil incorporation into DNA during neuronal development. *Mutation Research* **237**, 65-73.

Frame, M. C., Marsden, H. S. & Dutia, B. M. (1985). The ribonucleotide reductase induced by herpes simplex virus type-1 involves minimally a complex of 2 polypeptides (136K and 38K). *Journal of General Virology* **66**, 1581-1587.

Frenkel, N., Schirmer, E. C., Wyatt, L. S., Katsafanas, G., Roffman, E., Danovich, R. M. & June, C. H. (1990). Isolation of a new herpesvirus from human CD4+ T cells. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 748-752.

Friedberg, E. C., Walker, G. C. & Siede, W. (1995). Base excision repair. In *DNA Repair and Mutagenesis*, pp. 135-190. Washington: American Society for Microbiology.

Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-79

Furlong, D., Swift, H. & Roizman, B. (1972). Arrangement of herpesvirus deoxyribonucleic acid in the core. *Journal of Virology* **10**, 1071-1074.

Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO Journal* **12**, 4425-4431.

Gallant, J. E., Moore, R. D., Richman, D. D., Keruly, J., Chaisson, R. E., Bartlett, J., McAvinue, S., Bryson, Y., Cohen, H., Fischl, M., Bolin, T., Kessler, H., Burrough, Y., Mildvan, D., Fox, A., Richman, D., Freeman, B., Simon, G., Grabowy, K. W., Chernoff, D., Duff, P., Thompson, S., Barrett, K., Awe, R., Chapman, R., Leonard, S., Baines, L., Turner, P., Hawkins, M., Murray, H., Bowers, J., Lane, C., Tilson, H., Andrews, E. & Smiley, L. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. *Journal of Infectious Diseases* **166**, 1223-1227.

Gallo, M. L., Jackwood, D. H., Murphy, M., Marsden, H. S. & Parris, D. S. (1988). Purification of the herpes simplex virus type 1 65-kilodalton DNA-binding protein: properties of the protein and evidence of its association with the virus-encoded DNA polymerase. *Journal of Virology* **62**, 2874-2883.

Gao, M., Robertson, B. J., McCann, P. J., O'Boyle, D. R., Weller, S. K., Newcomb, W. W., Brown J. C. & Weinheimer, S.P. (1998). Functional conservations of the alkaline nuclease of herpes simplex type 1 and human cytomegalovirus. *Virology* **249**, 460-470.

Gelb, L. D. (1990). Varicella zoster virus. In *Virology*, 2nd edn. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

Georgopoulou, U., Kakkanas, A., Miriagou, V., Michaelidou, A. & Mavromara, P. (1995). Characterisation of the US8.5 protein of herpes simplex virus. *Archives of Virology* **140**, 2227-2241.

Gibbs, J. S., Chiou, H. C., Hall, J. D., Mount, D. W., Retondo, M. J., Weller, S. K. & Coen, D. M. (1985). Sequence and mapping analysis of the HSV DNA polymerase gene predict a C-terminal substrate binding domain. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 7969-7973.

Gibson, W. & Roizman, B. (1972). Proteins specified by herpes simplex virus. VIII. Characterisation and composition of multiple capsid forms of subtypes 1 and 2. *Journal of Virology* **10**, 1044-1052.

Giroir, E. L. & Deutsch, W. A. (1987). Drosophila deoxyuridine triphosphatase. Purification and characterization. *Journal of Biological Chemistry* **262**, 130-134.

Golstein, J. N. & Weller, S. K. (1998). The exonuclease activity of HSV-1 UL12 is required for *in vivo* function. *Virology* **244**, 442-457.

Gompels, U. A., Nicholas, J., Lawrence, G., Jones, M., Jones, M., Thomson, B. J., Martin, M. E. D., Efstathiou, S., Craxton, M. & Macaulay, H. A. (1995). The DNA sequence of human herpesvirus 6 - Structure, coding content, and genome evolution. *Virology* **209**, 29-51.

Gottlieb, J., Marcy, A. I., Coen, D. M. & Challberg, M. D. (1990). The herpes simplex virus type 1 UL42 gene product - a subunit of DNA polymerase that functions to increase processivity. *Journal of Virology* **64**, 5976-5987.

Greenspan, J. S., Greenspan, D., Lennette, E. T., Abrams, D. I., Conant, M. A., Petersen, V. & Freese, U. K. (1985). Replication of EBV within the epithelial cells of oral 'hairy' leukoplakia, an AIDS-associated lesion. *New England Journal of Medicine* **313**, 1564-1571.

Gunven, P., Klein, G., Henle, G., Henle, W. & Clifford, P. (1970). Epstein-Barr virus in Burkitt's lymphoma and nasopharyngeal carcinoma. Antibodies to EBV associated membrane and viral capsid antigens in Burkitt lymphoma patients. *Nature* **228**, 1053-1056.

Haffey, M. L., Stevens, J. T., Terry, B. J., Dorsky, D. I., Crumpacker, C. S., Wietstock, S. M., Ruyechan, W. T. & Field, A. K. (1988). Expression of herpes simplex type 1 DNA polymerase in *Saccharomyces cerevisiae* and detection of virus-specific enzyme activity in cell-free lysates. *Journal of Virology* **62**, 4493-4498.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557-580.

Hayashi K., Nakazawa, M., Ishizaki, Y., Hiraoka, N. & Obayashi, A. (1986). Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Research* **14**, 7617-7631.

Hochhauser, S. J. & Weiss, B. (1978). *Escherichia coli* mutants deficient in deoxyuridine triphosphatase. *Journal of Bacteriology* **134**, 157-166.

Hoffmann, I., Widstrom, J., Zeppezauer, M. & Nyman, P. O. (1987). Overproduction and large scale preparation of deoxyuridine triphosphate nucleotidohydrolase from *Escherichia coli*. *European Journal of Biochemistry* **164**, 45-51.

Hoffmann, P. J. (1981). Mechanism of degradation of duplex DNA by the DNase induced by herpes simplex virus. *Journal of Virology* **38**, 1005-1014.

Hokari, S. & Sakagishi, Y. (1987). Purification and characterisation of deoxyuridine triphosphate nucleotidohydrolase from anemic rat spleen - a trimer composition of the enzyme protein. *Archives of Biochemistry and Biophysics* **253**, 350-356.

Holland, L. E., Anderson, K. P., Shipman, C. & Wagner, E. K. (1980). Viral DNA synthesis is required for the expression of specific herpes virus type 1 mRNA species. *Virology* **101**, 50-53.

Honess, R. W., Bodemer, W., Cameron, K. R., Niller, H. H., Fleckenstein, B. & Randall, R. E. (1986). The A+T rich genome of the herpesvirus saimiri contains a highly conserved gene for thymidylate synthase. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3604-3608.

Honess, R. W. & Roizman, B. (1975). Regulation of herpes macro-molecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 1276-1280.

Ingemarsson, R. & Lankinen, H. (1987). The herpes simplex virus type 1 ribonucleotide reductase is a tight complex of the type $\alpha_2\beta_2$ composed of 40K and 140K proteins, of which the latter shows multiple forms due to proteolysis. *Virology* **156**, 417-422.

Ingraham, H. A., Dickey, L. & Goulian, M. (1986). DNA fragmentation and cyto-toxicity from increased cellular deoxyuridylate. *Biochemistry* **25**, 3225-3230.

Jamieson, A. T. & Subak-Sharpe, J. H. (1974). Biochemical studies on the herpes simplex virus-specified deoxypyrimidine kinase activity. *Journal of General Virology* **4**, 481-492.

Janin, J. (1979). Surface and inside volumes in globular proteins. *Nature* **277**, 491-492.

Jones, J. F., Shurin, S., Abramowsky, C., Tubbs, R. R., Sciotto, C. G., Wahl, R., Sands, J., Gottman, D., Katz, B. Z. & Sklar, J. (1988). T cell lymphomas containing Epstein-Barr viral DNA in patients with chronic Epstein-Barr virus infections. *New England Journal of Medicine* **318**, 733-741.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature* **358**, 86-89.

Jones, D. T. (1994). Threader: Protein fold recognition by optimal protein sequence threading – Usage notes. Available by anonymous ftp from ftp.biochem.ucl.ac.uk.

Jones, D. T. (1995). *Threader* computer software program. Available by anonymous ftp from ftp.biochem.ucl.ac.uk.

Jons, A., Granzow, H., Kuchling, R. & Mettenleiter, T. C. (1996). The UL49.5 gene of pseudorabies virus codes for an O-glycosylated structural protein of the viral envelope. *Journal of Virology* **70**, 1237-1241.

Jons, A. & Mettenleiter, T. C. (1996). Identification and characterisation of pseudorabies virus dUTPase. *Journal of Virology* **70**, 1242-1245.

Katz, J. P., Bodin, E. T. & Coen, D. M. (1990). Quantitative polymerase chain reaction analysis of herpes simplex virus DNA in ganglia of mice infected with replication-incompetent mutants. *Journal of Virology* **64**, 4288-4295.

Keir, H. M. & Gold, E. (1963). Deoxyribonucleic acid nucleotidyltransferase and deoxyribonuclease from cultured cell infected with herpes simplex virus. *Biochim Biophys Acta* **72**, 263-276.

Kennard, J., Rixon, F. J., McDougall, I. M., Tatman, J. D. & Preston, V. G. (1995). The 25 amino acid residues at the carboxy-terminus of the herpes simplex virus type 1 UL26.5 protein are required for the formation of the capsid shell around the scaffold. *Journal of General Virology* **76**, 1611-1621.

Khanna, R., Burrows, S. R., Thomson, S. A., Moss, D. J., Cresswell, P., Poulsen, L. M. & Cooper, L. (1997). Class I processing defective Burkitt's lymphoma cells are recognised efficiently by CD4(+) EBV-specific CTLs. *Journal of Immunology* **158**, 3619-3625.

Kinghorn, G. R. (1993). Genital herpes - natural history and treatment of acute episodes. *Journal of Medical Virology*, 33-38.

Knopf, K. W. (1979). Properties of herpes simplex virus DNA polymerase and characterisation of its associate exonuclease activity. *European Journal of Biochemistry* **98**, 231-244.

Knopf, K. W., Kaufman, E. R. & Crumpacker, C. S. (1981). Physical mapping of drug resistance mutation defines an active center of the herpes simplex DNA polymerase enzyme. *Journal of Virology* **39**, 746-757.

Koonin, E. V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Research* **24**, 2411-2415.

Koppe, B., Menendezarias, L. & Oroszlan, S. (1994). Expression and purification of the mouse mammary tumor virus *gag-pro* transframe protein P-30 and characterisation of its dUTPase activity. *Journal of Virology* **68**, 2313-2319.

Kosz-Vnenchaak, M., Coen, D. M. & Knipe, D. M. (1990). Restricted expression of herpes simplex virus lytic genes during establishment of latent infection by thymidine kinase-negative mutant viruses. *Journal of Virology* **64**, 5396-5402.

Kristensson, K., Lycke, E., Roytta, M., Svennerholm, B. & 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]. *Journal of General Virology* **67**, 2023-2028.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 488-492.

Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods in Enzymology* **154**, 367-382.

Kunz, B. A., Kohalmi, S. E., Kunkel, T. A., Mathews, C. K., McIntosh, E. M. & Reidy, J. A. (1994). Deoxyribonucleoside triphosphate levels: a critical factor in the maintenance of genetic stability. *Mutation Research* **318**, 1-64.

Kyte, J. & Doolittle, R. (1982). A simple method for displaying the hydrophathic character of a protein. *Journal of Molecular Biology* **157**, 105-132.

Ladner, R. D., Carr, S. A., Huddleston, M. J., McNulty, D. E. & Caradonna, S. J. (1996a). Identification of a consensus cyclin-dependent kinase phosphorylation site unique to the nuclear form of human deoxyuridine triphosphate nucleotidohydrolase. *Journal of Biological Chemistry* **271**, 7752-7757.

Ladner, R. D., McNulty, D. E., Carr, S. A., Roberts, G. D. & Caradonna, S. J. (1996b). Characterisation of distinct nuclear and mitochondrial forms of human deoxyuridine triphosphate nucleotidohydrolase. *Journal of Biological Chemistry* **271**, 7745-7751.

Lagunoff, M., Randall, G. & Roizman, B. (1996). Phenotypic properties of herpes simplex virus 1 containing a derepressed open reading frame-P gene. *Journal of Virology* **70**, 1810-1817.

Lagunoff, M. & Roizman, B. (1995). The regulation of synthesis and properties of the protein product of open reading frame-P of the herpes simplex virus-1 genome. *Journal of Virology* **69**, 3615-3623.

Larsson, G., Nyman, P. O. & Kvassman, J. O. (1996a). Kinetic characterisation of dUTPase From *Escherichia coli*. *Journal of Biological Chemistry* **271**, 24010-24016.

Larsson, G., Nyman, P. O. & Svensson, L. A. (1996b). Crystallisation and preliminary X-ray analysis of human dUTPase. *Acta Crystallographica Section D-Biological Crystallography* **52**, 1039-1040.

Larsson, G., Svensson, L. A. & Nyman, P. O. (1996c). Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analog (dUDP). *Nature Structural Biology* **3**, 532-538.

Lawrence, G. L., Chee, M., Craxton, M. A., Gompels, U. A., Honess, R. W. & Barrell, B. G. (1990). Human herpesvirus 6 is closely related to human cytomegalovirus. *Journal of Virology* **64**, 287-299.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**, 379-400.

Lerner, D. L., Wagaman, P. C., Phillips, T. R., Prosperogarcia, O., Henriksen, S. J., Fox, H. S., Bloom, F. E. & Elder, J. H. (1995). Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine triphosphatase. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7480-7484.

Levy, J. A., Landay, A. & Lennette, E. T. (1990). Human herpesvirus-6 inhibits human immunodeficiency virus type 1 replication in cell culture. *Journal of Clinical Microbiology* **28**, 2362-2364.

Liang, X., Tang, M., Manns, B., Babiuk, L. A. & Zamb, T. J. (1993). Identification and deletion mutagenesis of the bovine herpesvirus-1 dUTPase gene and a gene homologous to herpes simplex virus UL49.5. *Virology* **195**, 42-50.

Lichtenstein, D. L., Rushlow, K. E., Cook, R. F., Raabe, M. L., Swardson, C. J., Kociba, G. J., Issel, C. J. & Montelaro, R. C. (1995). Replication *in vitro* and *in vivo* of an equine infectious anemia virus mutant deficient in dUTPase activity. *Journal of Virology* **69**, 2881-2888.

Lindahl, T. (1982). DNA Repair Enzymes. *Annual Review of Biochemistry* **51**, 61-87.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

Lirette, R. & Caradonna, S. (1990). Inhibition of phosphorylation of cellular dUTP nucleotidohydrolase as a consequence of herpes simplex virus infection. *Journal of Cellular Biochemistry* **43**, 339-353.

Littler, E., Stuart, A. D. & Chee, M. S. (1992). Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir. *Nature* **358**, 160-162.

Littler, E., Zeuthen, J., McBride, A. A., Sorensen, E. T., Powell, K. L., Walsh-Arrand, J. E. & Arrand, J. R. (1986). Identification of an Epstein-Barr virus-encoded thymidine kinase. *The EMBO Journal* **5**, 1959-1966.

Liu, F. Y. & Roizman, B. (1991). The promoter, transcriptional unit, and coding sequence of herpes simplex virus-1 family 35 proteins are Contained within and in frame with the UL26 open reading frame. *Journal of Virology* **65**, 206-212.

Magrath, I., Jain, V. & Bhatia, K. (1992). Epstein-Barr virus and Burkitt's lymphoma. *Seminars in Cancer Biology* **3**, 285-295.

Mansoor, A., Stevenson, M. S., Li, R. Z., Frekko, K., Weiss, W., Ahmad, M., Khan, A. H., Mushtaq, S., Saleem, M., Raffeld, M., Kingma, D. W. & Jaffe, E. S. (1997). Prevalence of Epstein-Barr viral sequences and EBV LMP1 oncogene deletions in Burkitt's lymphoma from Pakistan: Epidemiological correlations. *Human Pathology* **28**, 283-288.

Marsden, H. S. (1992). Disruption of protein-subunit interactions. *Seminars in Virology* **3**, 67-75.

Marsden, H. S., McLean, G. W., Barnard, E. C., Francis, G. J., MacEachran, K., Murphy, M., McVey, G., Cross, A., Abbotts, A. P. & Stow, N. D. (1997). The catalytic subunit of the DNA polymerase of herpes simplex virus type 1 interacts specifically with the C terminus of the UL8 component of the viral helicase-primase complex. *Journal of Virology* **71**, 6390-6397.

Marsden, H. S., Murphy, M., McVey, G. L., Maceachran, K. A., Owsianka, A. M. & Stow, N. D. (1994). Role of the carboxy-terminus of herpes simplex Virus type-1 DNA polymerase in its interaction with UL42. *Journal of General Virology* **75**, 3127-3135.

Mathews, C. K. (1993). Enzyme organisation in DNA precursor biosynthesis. *Nucleic Acids Research* **44**, 167-203.

Mazzarello, P., Focher, F., Verri, A. & Spadari, S. (1990). Misincorporation of uracil into DNA as possible contributor to neuronal ageing and abiotrophy. *International Journal of Neuroscience* **50**, 169-174.

McClements, W., Yamanaka, G., Garsky, V., Perry, H., Bacchetti, S., Colonno, R. & Stein, R. B. (1988). Oligopeptides inhibit the ribonucleotide reductase of HSV by causing subunit separation. *Virology* **162**, 270-273.

McClure, M. A., Johnson, M. S. & Doolittle, R. F. (1987). Relocation of a protease-like gene segment between two retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 2693-2697.

McClure, M. A., Johnson, M. S., Feng, D.-F. & Doolittle, R. F. (1988). Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2469-2473.

McGeoch, D. J. (1987). The genome of herpes simplex virus - structure, replication and evolution. *Journal of Cell Science*, 67-94.

McGeoch, D. J. (1989). The genomes of the human herpesviruses - contents, relationships, and evolution. *Annual Review of Microbiology* **43**, 235-265.

McGeoch, D. J. (1990a). Evolutionary relationships of virion glycoprotein genes in the S regions of alphaherpesvirus genomes. *Journal of General Virology* **71**, 2361-2367.

McGeoch, D. J. (1990b). Protein sequence comparisons show that the pseudoproteases encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Research* **18**, 4105-4110.

McGeoch, D. J., Barnett, B. C. & MacLean, C. A. (1993). Emerging functions of alphaherpesvirus genes. *Seminars in Virology* **4**, 125-134.

McGeoch, D. J. & Cook, S. (1994). Molecular phylogeny of the alphaherpesvirinae subfamily and a proposed evolutionary timescale. *Journal of Molecular Biology* **238**, 9-22.

McGeoch, D. J., Cook, S., Dolan, A., Jamieson, F. E. & Telford, E. A. R. (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *Journal of Molecular Biology* **247**, 443-458.

McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scott, J. E. & Taylor, P. (1988a). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *Journal of General Virology* **69**, 1531-1574.

McGeoch, D. J., Dalrymple, M. A., Dolan, A., McNab, D., Perry, L. J., Taylor, P. & Challberg, M. D. (1988b). Structures of herpes simplex virus type-L genes required for replication of virus DNA. *Journal of Virology* **62**, 444-453.

McGeoch, D. J. & Davison, A. J. (1986). DNA sequence of the herpes simplex virus type-1 gene encoding glycoprotein gH, and identification of homologues in the genomes of varicella-zoster virus and Epstein-Barr virus. *Nucleic Acids Research* **14**, 4281-4292.

McGeoch, D. J., Dolan, A., Donald, S. & Brauer, D. H. K. (1986a). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type-1. *Nucleic Acids Research* **14**, 1727-1745.

McGeoch, D. J., Dolan, A. & Frame, M. C. (1986b). DNA sequence of the region in the genome of herpes simplex virus type-1 containing the exonuclease gene and neighbouring genes. *Nucleic Acids Research* **14**, 3435-3448.

McIntosh, E. M., Ager, D. D., Gadsden, M. H. & Haynes, R. H. (1992). Human dUTP pyrophosphatase - cDNA sequence and potential biological importance of the enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 8020-8024.

McKie, E. A., Hope, R. G., Brown, S. M. & MacLean, A. R. (1994). Characterisation of the herpes simplex virus type-1 strain 17(+) neurovirulence gene RL1 and its expression in a bacterial system. *Journal of General Virology* **75**, 733-741.

McLauchlan, J. & Clements, J. B. (1983). Organisation of the herpes simplex type 1 transcription unit encoding two early proteins with molecular weights of 140000 and 40000. *Journal of General Virology* **64**, 997-1006.

Mead, D.A., Skorupa, E.S. & Kemper, B. (1985) Single stranded DNA SP6 promoter plasmids for engineering mutant RNAs. *Nucleic Acid Res.* **13**,1103-1118.

Mercer, A. A., Fraser, K. M., Stockwell, P. A. & Robinson, A. J. (1989). A homolog of retroviral pseudoproteases in the parapoxvirus, orf virus. *Virology* **172**, 665-668.

Meyers, J. D., Flournoy, N. & Thomas, E. D. (1986). Risk factors for cytomegalovirus infection after human marrow transplantation. *Journal of Infectious Diseases* **153**, 478-488.

Miller, R. T. & Thornton, J. M. (1995). The Threading Analyst: a program to evaluate threading results. Available from ftp.biochem.ucl.ac.uk/pub/px.

Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). Human dUTP pyrophosphatase - uracil recognition by a beta-hairpin and active-sites formed by 3 separate subunits. *Structure* **4**, 1077-1092.

Moller, W. & Amons, R. (1985). Phosphate-binding sequences in nucleotide-binding proteins. *FEBS Letters* **186**, 1-7.

Moore, P. S., Gao, S. J., Dominguez, G., Cesarman, E., Lungu, O., Knowles, D. M., Garber, R., Pellett, P. E., McGeoch, D. J. & Chang, Y. (1996). Primary characterisation of a herpesvirus agent associated with Kaposi's sarcoma. *Journal of Virology* **70**, 9083-9083.

Morrison, J. M. (1991). Enzymes induced by DNA viruses. In *Virus Induced Enzymes*, pp. 215-361. Chichester, England: John Wiley and Sons Ltd.

Moss, B., Gershowitz, A., Stringer, J. R., Holland, L. E. & Wagner, E. K. (1977). 5' terminal and internal methylated nucleosides in HSV type 1 mRNA. *Journal of Virology* **23**, 234-239.

Movra, N. R., Nakamura, K. & Inouye, M. (1980). Amino acid sequence of the signal peptide of *OmpA* protein, a major outer membrane protein of *E.coli*. *Journal of Biological Chemistry* **255**, 27-29.

Mullaney, J., Moss, H. W. M. & McGeoch, D. J. (1989). Gene UL2 of herpes simplex virus type-1 encodes a uracil-DNA glycosylase. *Journal of General Virology* **70**, 449-454.

Nahmias, A. J., Josey, W. E., Naib, Z. M., Luce, C. & Duffey, C. (1970). Antibodies to herpesvirus hominis types 1 and 2 in humans. I. Patients with genital herpetic infections. *American Journal of Epidemiology* **91**, 539-546.

Nash, A. A. & Lohr, J. M. (1992). Pathogenesis and immunology of the herpesvirus infections of the nervous system. In *Neuropathic Viruses and Immunity*, pp. 155-176. Edited by S. Specter. New York: Plenum Press.

Nation, M. D., Guzder, S. N., Giroir, L. E. & Deutsch, W. A. (1989). Control of *drosophila* deoxyuridine triphosphatase - existence of a developmentally expressed protein inhibitor. *Biochemical Journal* **259**, 593-596.

Nazarian, K. (1974). DNA configuration in the core of Marek's disease virus. *Journal of Virology* **13**, 1148-1150.

Nicander, B. & Reichard, P. (1985). Evidence for the involvement of substrate cycles in the regulation of deoxyribonucleoside triphosphate pools in 3T6-cells. *Journal of Biological Chemistry* **260**, 9216-9222.

Nicholas, J. (1996). Determination and analysis of the complete nucleotide sequence of human herpesvirus 7. *Journal of Virology* **70**, 5975-5989.

Nicholas, J., Ruvolo, V., Zong, J., Ciufo, D., Guo, H. G., Reitz, M. S. & Hayward, G. S. (1997). A single 13-kilobase divergent locus in the Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genome contains nine open reading frames that are homologous to or related to cellular proteins. *Journal of Virology* **71**, 1963-1974.

Niederman, J. C., McCollum, R. W., Henle, G. & Henle, W. (1968). Infectious mononucleosis. *Journal of the American Medical Association* **68**, 139-143.

Notredame, C. & Higgins, D. G. (1996). SAGA: sequence alignment by genetic algorithm. *Nucleic Acids Research* **24**, 1515-1524.

Old, L. J., Clifford, P. & Boyse, E. A. (1966). Precipitating antibody in human serum to an antigen present in cultured Burkitt's lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America* **56**, 1699-1704.

Pardo, E. G. & Gutierrez, C. (1990). Cell cycle-dependent and differentiation stage-dependent variation of dUTPase activity in higher plant cells. *Experimental Cell Research* **186**, 90-98.

Pearl, L. H. & Savva, R. (1995). DNA repair in 3 dimensions. *Trends in Biochemical Sciences* **20**, 421-426.

Pearl, L. H. & Savva, R. (1996). The problem with pyrimidines. *Nature Structural Biology* **3**, 485-487.

Peitsch, M. C. (1995). Protein Modelling by Email. *Bio/Technology* **13**, 658-660.

Peitsch, M. C. (1996). ProMod and Swiss-Model: Internet based tools for automated comparative protein modelling. *Biochemical Society Transactions* **24**, 274-279.

Perry, L. J., Rixon, F. J., Everett, R. D., Frame, M. C. & McGeoch, D. J. (1986). Characterisation of the IE110 gene of herpes simplex virus type-1. *Journal of General Virology* **67**, 2365-2380.

Persson, R., Rosengren, A. M., Nyman, P. O., Dauter, Z. & Cedergren-Zeppezauer, E. (1997). Crystallisation and preliminary X-ray crystallographic studies of dUTPase from equine infectious anemia virus. *Protein and Peptide Letters* **4**, 145-148.

Persson, R., Sommer, P., Larsson, G., Zeppezauer, M., Grasser, F. A., Nyman, O. (1999). dUTPase from Epstein-Barr Virus; Recombinant expression, purification and characterisation. (*manuscript*).

Persson, T., Larsson, G. & Nyman, P. O. (1996). Synthesis of 2'-deoxyuridine 5'-(α, β-imido)triphosphate - a substrate analog and potent inhibitor of dUTPase. *Bioorganic & Medicinal Chemistry* **4**, 553-556.

Plummer, G., Bowling, C. P. & Goodheart, C. R. (1969). Comparison of four horse herpesviruses. *Journal of Virology* **4**, 738-741.

Powell, K. L., Littler, E. P. & Purifoy, D. J. M. (1981). Non structural proteins of herpes simplex virus II. Major virus-specific DNA-binding protein. *Journal of Virology* **39**, 894-902.

Prasad, G. S., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H. & Stout, C. D. (1996). Crystal structure of dUTP pyrophosphatase from feline immunodeficiency virus. *Protein Science* **5**, 2429-2437.

Preston, V. G. & Fisher, F. B. (1984). Identification of the herpes simplex virus type-1 gene encoding the dUTPase. *Virology* **138**, 58-68.

Price, A. R. & Fogt, S. M. (1973). Deoxythymidylate phosphohydrolase induced by bacteriophage PBS2 during infection of *Bacillus subtilis*. *Journal of Biological Chemistry* **248**, 1372-1380.

Price, A. R. & Frato, J. (1975). *Bacillus subtilis* deoxyuridine triphosphatase and its bacteriophage PBS2-induced inhibitor. *Journal of Biological Chemistry* **250**, 8804-8811.

Price, A. R. & Warner, H. R. (1969). Bacteriophage T4-induced deoxycytidine triphosphate-deoxyuridine triphosphate nucleotidohydrolase: its properties and its role during phage infection of *Escherichia coli*. *Virology* **39**, 882-892.

Pri-Hadash, A., Hareven, D. & Lifschitz, E. (1992). A meristem-related gene from tomato encodes a dUTPase - analysis of expression in vegetative and floral meristems. *Plant Cell* **4**, 149-159.

Pyles, R. B., Sawtell, N. M. & Thompson, R. L. (1992). Herpes simplex virus type-1 dUTPase mutants are attenuated for neurovirulence, neuroinvasiveness, and reactivation from latency. *Journal of Virology* **66**, 6706-6713.

Quinn, J. P. & McGeoch, D. J. (1985). DNA sequence of the region in the genome of herpes simplex virus type-1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Research* **13**, 8143-8163.

Rawlinson, W. D., Farrell, H. E. & Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *Journal of Virology* **70**, 8833-8849.

Read, G. S. & Frenkel, N. (1983). Herpes simplex virus mutants defective in the virion associated shut-off host polypeptide synthesis and exhibiting abnormal synthesis of alpha (immediate early) viral polypeptides. *Journal of Virology* **46**, 498-512.

Reichard, P. (1988). Interactions between deoxyribonucleotide and DNA synthesis? *Annual Review of Biochemistry* **57**, 349-374.

Reichard, P. (1993). From RNA to DNA, why so many ribonucleotide reductases. *Science* **260**, 1773-1777.

Reichard, P. & Nicander, B. (1985). On the regulation of deoxyribonucleotide synthesis. *Current Topics in Cellular Regulation* **26**, 403-409.

Richter, J., Puchtler, I. & Fleckenstein, B. (1988). Thymidylate synthase gene of herpesvirus ateles. *Journal of Virology* **62**, 3530-3535.

Rixon, F. J. (1993). Structure and assembly of herpesviruses. *Seminars in Virology* **4**, 135-144.

Rixon, F. J. & Clements, J. B. (1982). Detailed structural analysis of two spliced HSV-1 immediate early mRNAs. *Nucleic Acids Research* **10**, 2241-2256.

Rodriguez-Boulan, E. J. & Pendergast, M. (1980). Polarised distribution of viral envelope glycoproteins in the plasma membrane of infected epithelial cells. *Cell* **20**, 45-54.

Rogers, G. A., Larsson, G. & Nyman, P. O. (1997). The activity and mechanism of the enzyme dUTPase from *Escherichia coli. (manuscript)*.

Roizman, B. (1982). The family Herpesviridae: general description, taxonomy, and classification. In *The herpesviruses*, pp. 1-23. Edited by B. Roizman. New York: Plenum Press.

Roizman, B. & Furlong, D. (1974). The replication of herpesviruses. In *Comprehensive Virology*, pp. 229-403. Edited by H. Fraenkel-Conrat & R. R. Wagner. New York: Plenum Press.

Roizman, B. & Roane, P. R. (1964). Multiplication of herpes simplex virus. II. The relation between protein synthesis and the duplication of viral DNA in infected HEp-2 cells. *Virology* **22**, 262-269.

Roizman, B. & Sears, A. E. (1990). Herpes simplex viruses and their replication. In *Virology*, 2nd edn. Edited by B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath & B. Roizman. New York: Raven Press.

Roizman, B., Desrosiers, R. C., Fleckenstein, B., Lopez, C., Minson, A. C. & Studdert, M. J. (1992). The Family Herpesviridae - an update. *Archives of Virology* **123**, 425-449.

Rose, G., Geselowitz, A., Lesser, G., Lee, R. & Zehfus, M. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838.

Ross, J., Williams, M. & Cohen, J. I. (1997). Disruption of the varicella-zoster virus dUTPase and the adjacent ORF9A gene results in impaired growth and reduced syncytia formation *in vitro*. *Virology* **234**, 186-195.

Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *Journal of Molecular Biology* **232**, 585-599.

Rost, B. & Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55-77.

Rost, B. (1996) PHD: Predicting one dimensional protein structure by profile based neural networks. *Methods in Enzymology* **266**, 525-539.

Russo, J. J., Bohenzky, R. A., Chien, M. C., Chen, J., Yan, M., Maddalena, D., Parry, J. P., Peruzzi, D., Edelman, I. S., Chang, Y. & Moore, P. S. (1996). Nucleotide sequence of the Kaposi's sarcoma-associated herpesvirus (HHV-8). *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14862-14867.

Sander, C. & Schneider, R. (1991). Database of homology-derived structures and the structural meaning of sequence alignment. *Proteins* **9**, 56-68.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467.

Savva, R. & Pearl, L. H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase inhibitor protein complex. *Nature Structural Biology* **2**, 752-757.

Sayle, R. A. (1994). Rasmol molecular viewing program. Available by anonymous ftp from ftp.dcs.ed.ac.uk:/pub/rasmol.

Schnipper, L. E. & Crumpacker, C. S. (1980). Resistance of herpes simplex virus to acycloguanosine: role of viral thymidine kinase and DNA polymerase loci. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 2270-2273.

Shao, L., Rapp, L. M. & Weller, S. K. (1993) Herpes simplex virus 1 alkaline nuclease is required for efficient egress of capsids from the nucleus. *Virology* **196**, 146-162.

Shieh, M. T. & Spear, P. G. (1994). Herpesvirus induced cell fusion that is dependent on cell surface heparan sulfate or soluble heparan. *Journal of Virology* **68**, 1224-1228.

Shieh, M. T., Wudunn, D., Montgomery, R. I., Esko, J. D. & Spear, P. G. (1992). Cell surface receptors for herpes simplex virus are heparan sulfate proteoglycans. *Journal of Cell Biology* **116**, 1273-1281.

Shlomai, J. & Kornberg, A. (1978). Deoxyuridine triphosphatase of *Escherichia coli*. *Journal of Biological Chemistry* **253**, 3305-3312.

Silverstein, S., Bachenheimer, S. L., Frenkel, N. & Roizman, B. (1973). Relationship between post-transcriptional adenylation of herpes virus RNA and mRNA abundance. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 2101-2105.

Slabaugh, M. B. & Roseman, N. A. (1989). Retroviral protease-like gene in the vaccinia virus genome. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 4152-4155.

Slabaugh, M. B., Davis, R.E., Roseman, N. A. & Mathews, C.K. (1993). Vaccinia virus ribonucleotide reductase expression and isolation of the recombinant large subunit. *Journal of Biological Chemistry* **268**, 17803-17810.

Sommer, P., Kremmer, E., Bier, S., Konig, S., Zalud, P., Zeppezauer, M., Jones, J. F., Muellerlantzsch, N. & Grasser, F. A. (1996). Cloning and expression of the Epstein-Barr virus-encoded dUTPase - Patients with acute, reactivated or chronic virus infection develop antibodies against the enzyme. *Journal of General Virology* **77**, 2795-2805.

Spear, P. G. & Roizman, B. (1972). Purification and structural proteins of the herpes virion. *Journal of Virology* **9**, 143-159.

Steagall, W. K., Robek, M. D., Perry, S. T., Fuller, F. J. & Payne, S. L. (1995). Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virology* **210**, 302-313.

Stow, N. D. (1982). Localization of an origin of DNA replication within the TRS/IRS repeated regijon of the herpes simplex virus type 1 genome. *EMBO Journal* **1**, 863-867.

Stow, N. D. & McMonagle, E. C. (1983). Characterization of the TRS/IRS origin of DNA replication of herpes simplex virus type 1. *Virology* **130**, 427-438.

Strahler, J. R., Zhu, X. X., Hora, N., Wang, Y. K., Andrews, P. C., Roseman, N. A., Neel, J. V., Turka, L. & Hanash, S. M. (1993). Maturation stage and proliferation-dependent expression of dUTPase in human T cells. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4991-4995.

Straus, S. E. (1989). Clinical and biological differences between recurrent herpes simplex virus and varicella-zoster virus infections. *JAMA-Journal of the American Medical Association* **262**, 3455-3458.

Stryer, L. (1988). Viruses and Oncogenes. In *Biochemistry*, 3rd edn. New York: W.H. Freeman and Company.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 60-89.

Su, I. J., Hsieh, H. C., Lin, K. H., Uen, W. C., Kao, C. L., Chen, C. J., Cheng, A. L., Kadin, M. E. & Chen, J. Y. (1991). Aggressive peripheral T cell lymphomas containing Epstein-Barr viral DNA - a clinicopathological and molecular analysis. *Blood* **77**, 799-808.

Sullivan-Bolyai, J., Hull, H. F., Wilson, C. & Corey, L. (1983). Neonatal herpes simplex virus infection in King County, Washington - increasing incidence and epidemiologic correlates. *JAMA-Journal of the American Medical Association* **250**, 3059-3062.

Takahashi, K., Sonoda, S., Higashi, K., Kondo, T., Takahashi, H., Takahashi, M. & Yamanishi, K. (1989). Predominant CD4 lymphocyte-T tropism of human herpesvirus-6-related virus. *Journal of Virology* **63**, 3161-3163.

Telford, E. A. R., Lankinen, H. & Marsden, H. (1990). Inhibition of equine herpesvirus type 1-induced ribonucleotide reductase by the nonapeptide YAGAVVNDL. *Journal of General Virology* **71**, 1373-1378.

Telford, E. A. R., Studdert, M. J., Agius, C. T., Watson, M. S., Aird, H. C. & Davison, A. J. (1993). Equine herpesviruses-2 and herpesviruses-5 are gammaherpesviruses. *Virology* **195**, 492-499.

Telford, E. A. R., Watson, M. S., Aird, H. C., Perry, J. & Davison, A. J. (1995). The DNA sequence of equine herpesvirus-2. *Journal of Molecular Biology* **249**, 520-528.

Telford, E. A. R., Watson, M. S., McBride, K. & Davison, A. J. (1992). The DNA sequence of equine herpesvirus-1. *Virology* **189**, 304-316.

Thompson, R., Honess, R. W., Taylor, L., Morran, J. & Davison, A. J. (1987). Varicella-zoster virus specifies a thymidylate synthetase. *Journal of General Virology* **68**, 1449-1455.

Threadgill, D. S., Steagall, W. K., Flaherty, M. T., Fuller, F. J., Perry, S. T., Rushlow, K. E., Legrice, S. F. J. & Payne, S. L. (1993). Characterisation of equine infectious anemia virus dUTPase - Growth properties of a dUTPase-deficient mutant. *Journal of Virology* **67**, 2592-2600.

Trimble, J. J., Murthy, S. C., Bakker, A., Grassmann, R. & Desrosiers, R. C. (1988). A gene for dihydrofolate reductase in a herpesvirus. *Science* **239**, 1145-1147.

Turelli, P., Guiguen, F., Mornex, J. F., Vigne, R. & Querat, G. (1997). dUTPase-minus caprine arthritis encephalitis virus is attenuated for pathogenesis and accumulates G to A substitutions. *Journal of Virology* **71**, 4522-4530.

Turelli, P., Petursson, G., Guiguen, F., Mornex, J. F., Vigne, R. & Querat, G. (1996). Replication properties of dUTPase-deficient mutants of caprine and ovine lentiviruses. *Journal of Virology* **70**, 1213-1217.

Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S. & Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 154-157.

Verri, A., Mazzarello, P., Biamonit, G., Spadari, S. & Focher, F. (1990). The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic Acids Research* **18**, 5775-5780.

Vertessy, B.G., Larsson, G., Persson, T., Bergman, A. C., Persson, R. & Nyman, P. O. (1998). The complete triphosphate moiety of non-hydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in *E. coli* dUTP pyrophosphatase. *FEBS Letters* **421**, 83-88.

Vertessy, B. G., Persson, R., Rosengren, A. M., Zeppezauer, M. & Nyman, P. O. (1996). Specific derivatization of the active site tyrosine in dUTPase perturbs ligand binding to the active site. *Biochemical and Biophysical Research Communications* **219**, 294-300.

Vertessy, B. G., Zalud, P., Nyman, P. O. & Zeppezauer, M. (1994). Identification of tyrosine as a functional residue in the active site of *Escherichia coli* dUTPase. *Biochemica et Biophysica Acta-Protein Structure and Molecular Enzymology* **1205**, 146-150.

Vertessy, B. G. & Zeppezauer, M. (1994). Identification of tyrosine as an active site residue involved in the catalytic mechanism of *Escherichia coli* dUTPase. *Biochemical Society Transactions* **22**, S 233-S 233.

Virgin, H. W., Latreille, P., Wamsley, P., Hallsworth, K., Weck, K. E., Dal-Canto, A. J. & Speck, S. H. (1997). Complete sequence and genomic analysis of murine gammaherpesvirus 68. *Journal of Virology* **71**, 5894-5904.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 615-619.

Wadsworth, S., Jacob, R. J. & Roizman, B. (1975). Anatomy of the herpes simplex virus DNA II. Size composition and arrangement of inverted terminal repeats. *Journal of Virology* **5**, 1487-1497.

Wagaman, P. C., Hasselkuslight, C. S., Henson, M., Lerner, D. L., Phillips, T. R. & Elder, J. H. (1993). Molecular cloning and characterisation of deoxyuridine triphosphatase from feline immunodeficiency virus (FIV). *Virology* **196**, 451-457.

Wagner, E. K. (1985). Individual HSV transcripts: Characterisation of specific genes. In *The Herpesviruses*, pp. 45-104. Edited by B. Roizman: Plenum Press.

Wagner, E. K. & Summers, W. C. (1978). Structure of the joint regions and the termini of the DNA of herpes simplex virus type 1. *Journal of Virology* **27**, 374-387.

Wang, L. H. & Weiss, B. (1992). Dcd (dCTP Deaminase) Gene of *Escherichia coli* - mapping, cloning, sequencing, and identification as a locus of suppressors of lethal *dut* (dUTPase) mutations. *Journal of Bacteriology* **174**, 5647-5653.

Wang, Z. & Mosbaugh, D. W. (1989). Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS-2 encodes a binding protein specific for uracil-DNA glycosylase. *Journal of Biological Chemistry* **264**, 1163-1171.

Ward, P. L., Barker, D. E. & Roizman, B. (1996). A novel herpes simplex virus 1 gene UL43.5, maps antisense to the UL43 gene and encodes a protein which colocalises in nuclear structures with capsid proteins. *Journal of Virology* **70**, 2684-2690.

Warner, H. R., Duncan, B. K., Garret, C. & Neuhard, J. (1981). Synthesis and metabolism of uracil-containing deoxyribonucleic acid in *Escherichia coli*. *Journal of Bacteriology* **145**, 687-695.

Weiss, L. M., Strickler, J. G., Warnke, R. A., Purtilo, D. T. & Sklar, J. (1987). Epstein-Barr viral DNA in tissues of Hodgkin's disease. *American Journal of Pathology* **129**, 86-91.

Wheeler, L., Wang, Y. & Mathews, C. K. (1992). Specific associations of T4 bacteriophage proteins with immobilised deoxycytidylate hydroxymethylase. *Journal of Biological Chemistry* **267**, 7664-7670.

Wheeler, L. J., Ray, N. B., Ungermann, C., Hendricks, S. P., Bernard, M. A., Hanson, E. S. & Mathews, C. K. (1996). T4-phage gene-32 protein as a candidate organising factor for the deoxyribonucleoside triphosphate synthetase complex. *Journal of Biological Chemistry* **271**, 11156-11162.

Wiedbrauk, D. C. & Johnston, S. G. (1993). Herpes simplex virus. In *Manual of Clinical Virology*, pp. 111. New York: Raven Press.

Wildy, P., Russell, W. C. & Horne, R. W. (1960). The morphology of herpes virus. *Virology* **15**, 486-500.

Williams, M. V. & Cheng, Y. (1979). Human deoxyuridine triphosphate nucleotidohydrolase. Purification and characterization of the deoxyuridine triphosphate nucleotidohydrolase from acute lymphocytic leukemia. *Journal of Biological Chemistry* **254**, 2897-2901.

Williams, M. V. (1988). Herpes simplex virus-induced dUTPase - target site for antiviral chemotherapy. *Virology* **166**, 262-264.

Williams, M. V., Holliday, J. & Glaser, R. (1985). Induction of a deoxyuridine triphosphate nucleotidohydrolase activity in Epstein-Barr virus-infected cells. *Virology* **142**, 326-333.

Williams, M. V. & Parris, D. S. (1987). Characterisation of a herpes simplex virus type-2 deoxyuridine triphosphate nucleotidohydrolase and mapping of a gene conferring type specificity for the enzyme. *Virology* **156**, 282-292.

Wohlrab, F. & Francke, B. (1980). Deoxyribopyrimidine triphosphatase activity specific for cells infected with herpes simplex virus type 1. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1872-6.

Wohlrab, F., Garrett, B. K. & Francke, B. (1982). Control of expression of the herpes simplex virus-induced deoxypyrimidine triphosphatase in cells infected with mutants of herpes simplex virus types 1 and 2 intertypic recombinants. *Journal of Virology* **43**, 935-942.

Wolfenden, R., Andersson, L., Cullis, P. & Southgate, C. (1981). Affinities of amino acid side chains for solvent water. *Biochemistry* **20**, 849-855.

Worrad, D. M. & Caradonna, S. (1988). Identification of the coding sequence for herpes simplex virus uracil-DNA glycosylase. *Journal of Virology* **62**, 4774-4777.

Wu, C. A., Nelson, N. J., McGeoch, D. J. & Challberg, M. D. (1988). Identification of herpes simplex virus type-L genes required for origin-dependent DNA synthesis. *Journal of Virology* **62**, 435-443.

Wyatt, L. S., Rodriguez, W. J., Balachandran, N. & Frenkel, N. (1991). Human herpesvirus-7 - Antigenic properties and prevalence in children and adults. *Journal of Virology* **65**, 6260-6265.

Yamanishi, K., Kondo, T., Kondo, K., Hayakawa, Y., Kido, S., Takahashi, K. & Takahashi, M. (1990). Exanthem subitum and human herpesvirus 6 (HHV-6) infection. *Advances in Experimental Medical Biology* **278**, 29-37.