The DNA sequence of the RK strain of human herpesvirus 7

by

A. George Megaw

A Thesis presented for the Degree of Doctor of Philosophy

in

The Faculty of Science at the University of Glasgow

Institute of Virology
Church Street
Glasgow

December 1997

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

ACKNOWLEDGEMENTS

I would like to thank the following people who have contributed to the completion of this thesis.

Prof. D.J. McGeoch, Professor J.B. Clements, Dr. H. Marsden and Professor J.H. Subak-Sharpe, for the provision of the facilities in the Institute of Virology. Dr. A.J. Davison for his support, wisdom and advice (especially in the technical aspects of splicing) throughout the course of this work and also for his thorough criticism of this manuscript; your patience is a virtue. Professor N. Frenkel and Dr. D. Rapaport for provision of HHV-7 strain RK genomic DNA. The washroom and media staff for provision of glassware and general solutions used in this project.

Katherine, I owe a debt of understanding that I shall never be able to repay - I love you. Aoife, who appeared not to notice, your newborn giggles and smiles were like watching the snow fall. Everyone else who helped a little bit every day (Charles, Robin, Wendy, Ali, Moira, Steve and the others) - thank you.

During the course of this study, the author was in receipt of a Medical Research Council studentship and, unless otherwise stated, all results were obtained by the author's own efforts.

SUMMARY

Human herpesvirus 7 (HHV-7) was first isolated by Frenkel and colleagues in 1989 from $CD4^+$ T-cells, during experiments concerning propagation of HHV-6. Uninfected cells underwent spontaneous cytopathic effect following conditions promoting T-cell activation. A new herpesvirus was isolated which was related to but distinct from human herpesvirus 6 (HHV-6) This virus was designated HHV-7 (RK).

HHV-7 is ubiquitous in the human population - probably more than 85% of people are infected. It is likely that HHV-7 is transmitted through saliva, with primary infection occurring before three years of age. The virus has no proven involvement in any disease. HHV-7 is a member of the Betaherpesvirinae and is more closely related to HHV-6 than to human cytomegalovirus (HCMV). The genome is approximately 145 kbp in length and has an overall structure like that of HHV-6, consisting of a long unique region (U) flanked by a direct repeat (DR).

Since 1984, the complete sequences of fourteen herpesvirus genomes have been published and at least four more have been completed. This has allowed comparative studies of individual genes and permitted extensive and detailed understanding of the relationships between the herpesviruses. In addition, recent reports based on molecular sequence analysis, have provided a detailed phylogeny and allowed an evolutionary timescale to be derived.

The aims of this project were to determine the complete DNA sequence of HHV-7 and to analyse the coding potential of the sequence. The work was initiated in collaboration with Professor N. Frenkel (Tel-Aviv University, Israel), who provided genomic DNA from the RK strain. During the course of this work the complete DNA sequences of HHV-6 strain U1102 and HHV-7 strain JI were published by other groups. Therefore, the HHV-7 strain RK sequence was compared with these two sequences in order to appraise evolutionary divergence and to re-evaluate the genetic contents of HHV-6 and HHV-7.

The HHV-7 RK genome was sequenced by direct shotgun cloning of random DNA fragments into an M13 vector. Approximately 7,000 fragments were sequenced and compiled into a

database with the aid of a computer. The genetic content of the sequence was interpreted using a suite of computer programs. This analysis was then extended to reassess the sequence of HHV-6. Special attention was paid to potential splicing.

The HHV-7 RK genome sequence obtained is 153,080 bp in length and comprises U (133,012 bp) flanked on each side by a copy of DR (10,034 bp). Four different reiterated sequences are present in the genome, two in U and two in DR. the latter comprise arrays of a human telomere-like element. Complexity and size resulted in one of these arrays being particularly challenging to resolve, and an unsatisfactory solution in this region is reflected in the likelihood that the sequence obtained is approximately 3 kbp longer than the actual genome. Analysis of the differences between HHV-7 RK and JI indicated that the lineages resulting in the two strains diverged from a common ancestor of the order of 10,000 years ago.

It was concluded on the basis of various analytical criteria that the HHV-7 genome contains 84 genes, only one of which has no direct counterpart in HHV-6. However, the latter is related to an adjacent gene which has an HHV-6 homologue. By the same criteria, HHV-6 contains 85 genes, two of which lack counterparts in HHV-7. One encodes a putative membrane glycoprotein and the other is a homologue of the adeno-associated virus type 2 *rep* gene. Furthermore, sequence comparisons between HHV-7 and HHV-6 also allowed the identification of putative splice sites in eleven genes. One of these genes is predicted to encode a previously unrecognised membrane glycoprotein.

This study lays the foundation for continuing experimental investigation of gene expression in HHV-7, particularly in regard to splicing.

CONTENTS

1.0. INTRODUCTION

FIGURES AND TABLES

FIGURES

TABLES

Abbreviations

°	degrees Celsius
µCi	microcurie
µg	microgram
µM	micromolar
A	adenine
AAV-2	adeno-associated virus 2
AIDS	acquired immune deficiency syndrome
APS	ammonium persulphate
ATP	adenosine triphosphate
BGM	bottom gel mix
bp	base pairs
BP	before present
BPB	bromophenol blue
C	Cytosine
CBMC	cord blood mononuclear cells
cDNA	complementary DNA
CFS	chronic fatigue syndrome
Ci	Curie
CIP	calf intestinal phosphatase
CNS	central nervous system
contig	Contiguous sequence
CPE	Cytopathic effect
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
ddATP	2'3'-dideoxyadenosine-triphosphate
ddCTP	2'3'-dideoxycytidine-triphosphate
ddGTP	2'3'-dideoxyguanosine-triphosphate
ddTTP	2'3'-dideoxythymidine-triphosphate
dGTP	2'-deoxyguanosine-5'-triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleoside triphosphates
DR	Direct repeat
DR	Direct repeat
DR_L	Direct repeat left
DR_R	Direct repeat right
DTT	Dithiothreitol
dTTP	2'-deoxythymidine-5'-triphosphate
dUTPase	Deoxyuridine triphosphate nucleotidohydrolase
E	Early
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetra-acetic acid
EtBr	Ethidium bromide

EtOH	Ethanol
G	Guanine
g	Gram
GCR	G-coupled receptor
gp	glycoprotein
h	hour
HPV	human papilloma virus
IE	immediate early
Ig	immunoglobulin
IPTG	isopropyl-β-D-thiogalactoside
IR_L	internal long repeat
IR_S	internal short repeal
kbp	kilobase pair
L	late
l	litre
M	molar
mg	milligram
MIE	major immediated early
min	minute
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MS	multiple sclerosis
nm	nanometer
OBP	origin-binding protein
ORF	open reading frame
ori	origin of replication
ori_L	origin of replication in U_L
ori_S	origin of replication in U_S
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
PEG	polyethylene glycol
R1	reiterated sequence 1
R2	reiterated sequence 2
R3	reiterated sequence 3
RNA	ribonucleic acid
RNase A	ribonuclease A
rpm	revolutions per minute
RR	ribonucleotide reductase
RT	room temperature
SDS	sodium dodecyl sulphate
T	thymidine
T1	telomeric reiteration 1
T2	telomeric reiteration 2
TEMED	N, N, N', N' -tetramethethylene diamine
TGM	top gel mix
Tris	Tris (hydroxymethyl) aminomethane
TR_L	long terminal repeat

TR_S	short terminal repeat
U	unique
U	unique sequence
U_L	long unique
U_S	short unique
UV	ultraviolet
v/v	volume/volume
w/v	weight/volume
w/w	weight/weight
wt	wild type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Herpesviruses

Virus name	Common abbreviation	ICTV designation
Alcelaphine herpesvirus	AHV-1	AHV-1
Bovine herpesvirus 1	BHV-1	BHV-1
Channel catfish virus	CCV	IHV-1
Epstein-Barr virus	EBV	HHV-4
Equine herpesvirus 1	EHV-1	EHV-1
Equine herpesvirus 2	EHV-2	EHV-2
Equine herpesvirus 4	EHV-4	EHV-4
Herpes simplex virus type 1	HSV-1	HHV-1
Herpes simplex virus type 2	HSV-2	HHV-2
Herpesvirus ateles	HVA	AHV-2
Herpesvirus of turkey	HVT	MHV-1[b]
Herpesvirus saimiri	HVS	SHV-2
Human cytomegalovirus	HCMV	HHV-5
Human herpesvirus 6	HHV-6	HHV-6
Human herpesvirus 7	HHV-7	HHV-7
Kaposi's sarcoma-associated herpesvirus	KSHV	(HHV-8)
Marek's disease virus	MDV	GHV-2
Murid herpesvirus 68	MHV-68	MHV-4
Murine cytomegalovirus	MCMV	MHV-1[a]
Pseudorabies virus	PRV	SHV-1
Salmonid herpesvirus 1	SalHv-1	SalHv-1
Salmonid herpesvirus 2	SalHv-2	SalHv-2
Simian herpesvirus	B virus	CHV-1
Varicella-zoster virus	VZV	HHV-3

[a] Murid herpesvirus 1. [b] Meleagrid herpesvirus 1.

One and three letter abbreviations for amino acid residues

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
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

Chapter 1

Introduction

1.0. FOREWORD

This chapter is divided into four sections. Section 1.1. provides a thumbnail description of several aspects of the Herpesviridae, including clinical manifestations, life cycle, classification, genetic content and evolution. Sections 1.2., 1.3. and 1.4., refer to the members of the human Betaherpesvirinae: human cytomegalovirus (HCMV), human herpesvirus 6 (HHV-6) and human herpesvirus 7 (HHV-7), respectively. These viruses are related and share many common features. Thus, the information in each section is similarly organised, and includes sub-sections concerning general characteristics, epidemiology, tropism, growth properties and diseases of the viruses, with emphasis placed on their genetic makeup. The chapter concludes with a short statement of the thesis aims.

1.1. THE HERPESVIRIDAE

1.1.1. Overview

The Herpesviridae are eukaryotic viruses, and number more than 100. They infect both warm and cold-blooded vertebrates, in a broad range from fish to man (Roizman and Sears, 1993; Fenner *et al.,* 1987) and at least one invertebrate, the oyster (Comps and Cochennec, 1993). The best studied infect humans, cattle, chickens, horses, mice and monkeys. Herpesviruses are well adapted to their hosts. Most display a high degree of host specificity, although a single host may become infected by several herpesviruses; for example, there are at least eight human herpesviruses. However, some, such as pseudorabies virus (PRV) and B virus, are less host specific and may infect a range of species.

Herpesvirus lytic infection typically follows a lifecycle of adsorption and penetration into the cell and migration of the viral DNA to the nucleus where gene expression occurs in a sequentially ordered cascade, with three main phases: immediate early (IE), early (E) and late (L). Transcription, viral DNA replication and capsid assembly take place in the nucleus. DNA is packaged into the capsids which then acquire tegument and envelope. Virions are released from the cell by exocytosis.

A common property of herpesviruses is the ability, following primary infection, to persist in an inapparent (or latent) form for the lifetime of the host, with only a small subset of viral genes

being expressed (Kieff and Liebowitz, 1990; Rock, 1993). Occasionally, latent virus may reactivate under the influence of as yet poorly defined stimuli.

Herpesvirus epidemiology is not linked to geographical or seasonal factors and age does not correlate with susceptibility to infection. It is possible, however, to predict the levels of seroconversion to herpesviruses in a population in relation to age. For example at three years of age 90% of people have antibodies to human herpesvirus 6 (HHV-6). Herpesvirus infection may be transmitted horizontally or vertically by direct contact of mucosal surfaces (through saliva, sexually or in breastmilk), but respiratory droplet infection is also common.

Membership of the Herpesviridae is based on virion morphology. The DNA genome which is in a liquid crystalline state (Booy *et al.,* 1991) is packaged within an icosahedral capsid, 100-110 nm in diameter and composed of 162 capsomeres: 12 pentavalent capsomers at the vertices, 60 hexavalent capsomers at the 20 faces and 90 hexavalent capsomers along the 30 edges (Wildy *et al.,* 1960). The capsid is enclosed within a proteinaceous layer (the tegument); (Roizman and Furlong, 1974), which in turn is surrounded by a host-derived lipid envelope, containing viral glycoproteins (Morgan *et al.,* 1959; Asher *et al.,* 1969; Spear and Roizman, 1972; Stannard *et al.,* 1987). Herpesvirions range in diameter between 160 nm and 230 nm, averaging 180 nm (Szilagyi and Berriman, 1994).

Herpesviruses have large, linear, double-stranded DNA genomes (Furlong *et al.,* 1972) which range in size from 125 kbp (e.g. varicella zoster virus; VZV) to 240 kbp (e.g. murine cytomegalovirus; MCMV) and exhibit an impressively wide range of base compositions from 32-75% G+C (Honess, 1984). A characteristic of herpesvirus genome structure is the presence of terminal or internal repeated sequences, in direct or inverse orientation. Herpesviruses contain between about 70 (e.g. herpes simplex virus type 1; HSV-1) and 200 genes (e.g. human cytomegalovirus; HCMV). The great majority of the genome codes for protein, and the genes are arranged in about equal numbers on each DNA strand.

1.1.2. Clinical manifestations

1.1.2.1. General

Primary infections by herpesviruses are often asymptomatic or slight, and recrudescent infection may occur once (e.g. VZV) or many times (e.g. HSV-1). Herpesvirus infection combined with conditions of immune suppression can prove fatal. For example, bone marrow transplant patients suffering from interstitial pneumonitis due to HCMV infection risk 25% mortality. In addition, infection with certain herpesviruses, such as Epstein-Barr virus (EBV) and Marek's disease virus (MDV), can result in carcinoma or lymphoma (Epstein *et al.,* 1964; Evans and Niederman, 1991; Fenner *et al.,* 1987), and the recently discovered Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8 or HHV-8) probably plays a significant causative role in Kaposi's sarcoma (Moore *et al.,* 1996). More unusually, a few herpesviruses are capable of infecting more than one species under natural or near-natural conditions. B virus causes an HSV-1-like disease in monkeys but brings about fatal ascending paralysis and encephalitis in humans (Fenner *et al.,* 1987). PRV primarily infects swine, but in cattle causes "mad-itch", resulting in death within a few hours; also dogs suffer a rabies-like illness which is also invariably fatal (Fenner *et al.,* 1987).

1.1.2.2. Human herpesviruses

Currently, most research effort is put into the human herpesviruses and herpesviruses of domestically important animals. Humans are host to at least eight herpesviruses, seven of which are etiologic agents of a number of diseases.

HSV-1 is ubiquitous in human populations and is typically acquired by the age of five years via respiratory droplets and through contact with mucosal surfaces or abraded skin. Primary infection is usually subclinical, and lifelong latent infection is established in neurons of the sensory ganglia (Baringer and Swoveland, 1973; Fraser *et al.,* 1981). Subsequent reactivation can be periodically triggered by a number of factors including stress, UV radiation and immune suppression (Hill *et al.,* 1978). During primary and recurrent infection the virus causes vesicular lesions of the skin and mucosa. Infection is generally limited to the oropharynx but may involve the eyes and the genitalia (Whitley, 1985; Cameron, 1993).

The clinical pathologies associated with HSV-1 and herpes simplex virus type 2 (HSV-2) are not strictly separated (Nahmias *et al.,* 1991). However, HSV-2 is transmitted sexually, although infection may occur at birth (Whitley, 1985; Cameron, 1993), and lesions are localised to the mucous membranes and skin of the genitals. HSV-2 reactivation usually occurs more frequently than that of HSV-1 (Timbury, 1991; Whitley, 1990).

VZV can spread by airborne infection or by direct contact (Johnson, 1982) and causes two distinct diseases (Hope-Simpson, 1965; Gelb, 1990). Primary infection causes the childhood disease varicella (chickenpox), a generally benign condition characterised by a fever and a vesicular rash. Complications are rare but symptoms may be more severe in adults, where pneumonia is common. The virus establishes latent infection in the dorsal root or cranial nerve ganglia. VZV reactivation causes the disease zoster (shingles), where painful vesicles appear in the area of skin enervated by the affected ganglia. The incidence and severity of zoster increases with age and in the immunocompromised (Schrimpff *et al.,* 1972; Kennedy, 1987; Timbury, 1991), and can leave the victim with the distressing condition of post-herpetic neuralgia.

EBV initiates infection via the oropharynx (Sixbey *et al.,* 1983, 1984) and establishes latent infection in B-lymphocytes. Primary EBV infection is commonly asymptomatic in children, but during adolescence and adulthood the virus may cause infectious mononucleosis (glandular fever). In addition, EBV is implicated in cancers, such as Burkitt's lymphoma, a highly malignant B-cell lymphoma which is common in African children and virtually confined to areas where malaria is endemic, and nasopharyngeal carcinoma which is common among southern Chinese populations (Neiderman *et al.,* 1976; Epstein and Achong, 1986). The selective geographic and racial distribution of these tumours suggests the influence of genetic and environmental factors, as well as EBV infection, in the progression towards disease (Miller, 1990; Timbury, 1991).

HCMV primary infection is often asymptomatic or self-limiting with the only evidence of infection being an increase in antibody titre (Alford and Britt, 1990). Transmission is thought to occur most commonly by salivary contact, but the virus may also be acquired sexually (Emery and Griffiths, 1990). The virus establishes latency in monocytes (Taylor-Wiedeman *et al.,* 1991). The most

serious aspect of HCMV infection occurs if the host is immunocompromised through immaturity or suppression. For example, HCMV can act as a major pathogen of the foetus and allograft or AIDS patients (Alford and Britt, 1990; Emery and Griffiths, 1990; Black and Pellet, 1993).

HHV-6 is typically acquired by the age of two years (Yoshikawa, 1993) and can persist in T-lymphocytes (Salahuddin *et al.,* 1986), possibly establishing latency there (Frenkel and Wyatt, 1992). The virus has been isolated from the saliva of normal adults (Pietroboni *et al.,* 1988a; Harnett *et al.,* 1990; Levy *et al.,* 1990) and this may act as a vehicle for transmission (Wyatt and Frenkel, 1992). HHV-6 isolates may be segregated into two groups, termed variants A and B (HHV-6A and HHV-6B), based on biological and immunological properties and genomic analysis (Wyatt *et al.,* 1990; Ablashi *et al.,* 1991; Schirmer *et al.,* 1991; Aubin *et al.,* 1993). HHV-6A is not associated with disease but HHV-6B is the etiologic agent of exanthem subitum (roseola infantum), a childhood febrile illness (Yamanishi *et al.,* 1988).

Human herpesvirus 7 (HHV-7) has been shown to replicate preferentially in activated CD4+ T-lymphocytes (Frenkel *et al.,* 1990) and has also been isolated from the saliva of up to 81% of healthy adults (Wyatt and Frenkel, 1992; Black *et al.,* 1993; Yoshikawa *et al.,* 1993; Hidaka *et al.,* 1993). HHV-7 primary infection is usually acquired by three years of age (Wyatt *et al.,* 1991; Clark *et al.,* 1993; Hidaka *et al.,* 1993), but a potential site for HHV-7 latent infection has not been identified, nor has transfer by saliva been proved to be the mode of virus transmission. HHV-7, like HHV-6, has been associated with a proportion of exanthem subitum cases (Tanaka *et al.,* 1994).

Kaposi's sarcoma (KS) is a neoplasm of uncertain histogenesis occurring in both HIV-1-infected and uninfected persons (Moore *et al.,* 1996). Kaposi's sarcoma-associated herpesvirus (KSHV), or HHV-8, was discovered in an AIDS-KS lesion by representational difference analysis and shown to be present in almost all AIDS-KS lesions (Chang *et al.,* 1994). These findings have been confirmed and extended to nearly all KS lesions examined from the various epidemiologic classes of KS (Boshoff *et al.,* 1995; Dupin *et al.,* 1995; Moore *et al.,* 1996; Schalling *et al.,* 1995; Chang *et al.,* 1995).

1.1.3. Life cycle

The life cycle of herpesviruses is divided between lytic and latent phases. This can best be illustrated by briefly considering the life cycle of the best-studied herpesvirus, HSV-1; its general aspects apply to the other herpesviruses. Certain properties of EBV are also considered. Specific details relating to HCMV, HHV-6 and HHV-7 are discussed later.

1.1.3.1. The lytic cycle

Initial association of HSV-1 with the host cell is mediated by viral envelope glycoproteins, which also have important roles in adsorption and penetration of the virus (Campadelli-Fiume, 1994). After the virus envelope fuses with the plasma membrane, the nucleocapsid is released into the cytoplasm and migrates to the nucleus where the DNA enters via nuclear pores. At least some of the tegument proteins also enter the nucleus, although the process by which this occurs is not clear. Transcription and replication of viral DNA and capsid assembly take place in the nucleus.

Gene expression occurs in a co-ordinated, regulated and sequentially ordered cascade with three main phases: immediate early (IE or α), early (E or β) and late (L or γ) (Honess and Roizman, 1975; Roizman and Sears, 1990). Overall, proteins encoded by 74 genes are expressed during the cycle. Some IE proteins are trans-acting regulators of virus genes, and initiate cascade expression. E genes encode several enzymes involved in DNA replication and nucleotide metabolism, as well as a subset of glycoproteins and some uncharacterised proteins. L genes encode many virion structural proteins (Honess and Roizman, 1975).

Replication starts in the nucleus, with circularisation of viral genomic DNA by direct ligation of the termini. Viral DNA synthesis is initiated from the viral origins of replication (ori_S and ori_L) to produce DNA in an endless conformation, probably as head-to-tail concatemers, by a rolling circle mechanism (Roizman, 1979; Jacob *et al.*, 1979). Replicated DNA is cleaved specifically into unit-length molecules and packaged into preformed capsids.

The processes by which the capsid acquires tegument and envelope are not fully understood. The envelope is derived from altered host membranes and contains viral glycoproteins, which

are processed into their mature forms in the Golgi apparatus. Virions are released from the cell by exocytosis (Rixon, 1993; Spear, 1985).

1.1.3.2. The latent cycle

Latent infections are produced by all of the human herpesviruses, but are established at sites that are specific to the virus. For example, HSV-1 and VZV become latent in sensory ganglia, whereas EBV latently infects circulating B lymphocytes (Kieff and Liebowitz, 1990; Bastian *et al.,* 1972).

In its latent form, HSV-1 DNA is present either as a circular molecule or as a concatemer (Rock and Fraser, 1983, 1985; Efstathiou *et al.,* 1986) and is not thought to integrate into the host DNA (Mellerick and Fraser, 1987). EBV latent DNA is also maintained as covalently closed circular episomes (Lindhal *et al.,* 1976), and both HSV-1 and EBV latent genomes are apparently present in multiple copies in the latently infected cell. In addition, EBV infection can result in cell transformation and proliferation that may lead to carcinomas or lymphomas (Zur Hausen *et al.,* 1970).

The products of viral genes expressed during the latent state appear to function in maintenance of and reactivation from latency (Steiner *et al.,* 1989; Dambaugh *et al.,* 1986; Speck and Strominger, 1989), but differ between HSV-1 (latency-associated transcripts or LATs) (Spivack and Fraser, 1987; Deatly *et al.,* 1988) and EBV (Epstein-Barr virus nuclear antigen or EBNA proteins) (Lindahl *et al.,* 1974).

1.1.4. Herpesvirus genes

The functions of many herpesvirus genes have been assigned from experimental data and from comparisons with genes of known function from other organisms. Herpesvirus proteins may be grouped into five functional categories, as illustrated by McGeoch and Schaffer (1993).

Firstly, control proteins that influence viral transcription or that modulate the infected cell to facilitate viral replication (Everett, 1987). Secondly, essential components of the DNA replication machinery, including DNA polymerase and associated processivity factor, a protein which

recognises the origins of viral DNA replication in certain herpesviruses, a single-stranded DNA-binding protein and the three constituents of a helicase-primase complex (Challberg, 1991). Thirdly, enzymes engaged in nucleotide metabolism or DNA repair, including thymidine kinase, uracil-DNA glycosylase, dUTPase, ribonucleotide reductase (two subunits), a deoxyribonuclease and thymidylate synthase (in VZV and certain other herpesviruses) (Morrison, 1991). Fourthly, virus structural proteins, including components of the capsid, tegument and envelope (Rixon, 1993). Lastly, proteins involved in pathogenesis or latency, including those that which modulate the immune response of the host, such as G protein coupled receptors in certain herpesviruses (Chee *et al.,* 1990; Gompels *et al.,* 1995; Nicholas, 1996).

Experimental analyses of the phenotypes of HSV-1 mutants indicate that approximately half of its complement of genes is not absolutely essential for viral growth in cell culture on a gene-by-gene basis (Roizman and Sears, 1990; McGeoch and Schaffer, 1993). It is assumed that these genes confer a selective advantage *in vivo*.

1.1.5. Herpesvirus classification

1.1.5.1. Formal classification

A formal system of classification for the herpesviruses has been developed (Roizman *et al.,* 1981; Roizman *et al.,* 1992), in which the family Herpesviridae is divided between three subfamilies, the Alpha-, Beta- and Gammaherpesvirinae. Since little genetic information was available at that time, this division depended on biological criteria alone.

The Alphaherpesvirinae are classified on the basis of a moderately wide host range *in vitro*. They exhibit a relatively short reproductive cycle, efficiently destroy infected cells and spread rapidly in culture. Many members of this subfamily have been shown to establish latent infections in sensory ganglia. Members of the Alphaherpesvirinae include HSV-1, HSV-2, VZV, PRV and equine herpesvirus 1 (EHV-1).

The Betaherpesvirinae typically (though not exclusively) demonstrate a restricted host cell range *in vitro*, and infected cells often become enlarged, forming cytomegalia. The reproductive cycle is relatively long and infection progresses slowly in culture, with cell lysis occurring several days

after infection. Latency among the Betaherpesvirinae is as yet poorly understood but may occur in secretory glands, lymphoreticular cells and kidney cells. HCMV, MCMV, HHV-6 and HHV-7 are now amongst those viruses grouped in the Betaherpesvirinae.

In the Gammaherpesvirinae, host range *in vitro* is typically restricted to cells from animals to which the natural host belongs. The viruses of the Gammaherpesvirinae exhibit reproductive cycles of variable length. They are lymphotropic and infect either B- or T-lymphocytes; some also cause infections in epithelial and fibroblastoid cells. Infection of lymphocytes, even at the lytic stage, often occurs without the production of infectious progeny. Latency is frequently established in lymphoid tissue. Well-studied viruses from this subfamily include EBV, herpesvirus saimiri (HVS) and HHV-8.

Since this initial classification of the herpesviruses, significant effort has been channelled into determining the structures and DNA sequences of herpesvirus genomes, and attempts have been made to use the emerging data in classification. Genome size and nucleotide composition have proved unsuitable criteria in classification, as many members of each subfamily have genomes of similar size and G+C content may vary widely (Honess, 1984). Genome structure is a somewhat more suitable aid in herpesvirus classification. A characteristic of herpesvirus genome structure is the arrangement of reiterated sequences positioned within and at the ends of the genomes, either in direct or inverse orientation. On this basis the herpesviruses may be grouped into six categories (Fig. 1.0). Although there is no exclusive correlation between type of genome structure and the formal herpesvirus classification system, group 5 genomes are characteristic of many of the Alphaherpesvirinae and group 2 genomes of many of the Gammaherpesvirinae (Davison, 1993).

More recently, the classification system was refined (Roizman *et al.,* 1992), to take into account herpesvirus genome structures and more importantly, relationships between DNA or protein sequences and gene arrangements. The three subfamilies were each divided into genera: the Alphaherpesvirinae into the genera Simplexvirus (e.g. HSV-1 and HSV-2) and Varicellovirus (e.g. VZV); the Betaherpesvirinae into the genera Cytomegalovirus (e.g. HCMV), Roseolovirus (HHV-6) and Muromegalovirus (e.g. MCMV); and the Gammaherpesvirinae into the genera Lymphocryptovirus (e.g. EBV) and Rhadinovirus (e.g. HVS).

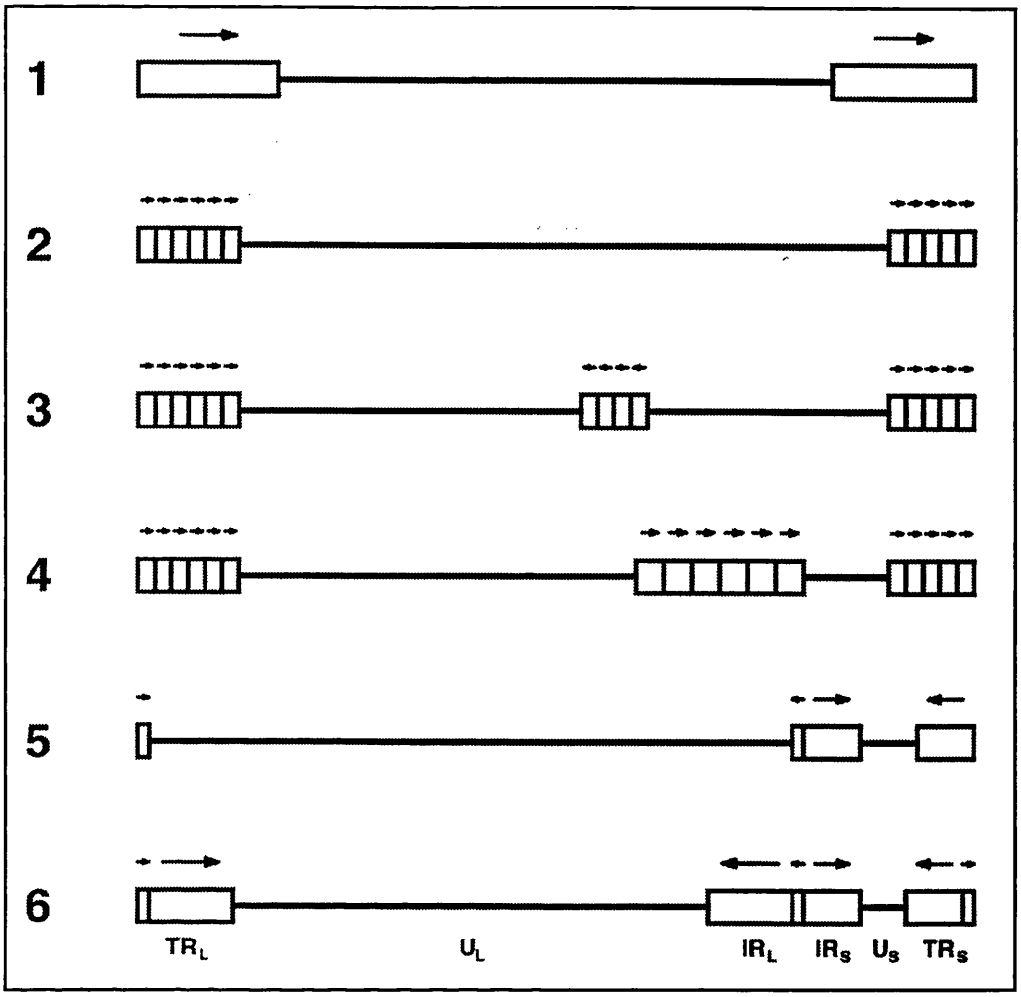


Fig. 1. Types of herpesvirus genome structure

From Davison and McGeoch (1995). Unique and repeated sequences are seen as lines and rectangles, respectively. Arrows indicate locations and orientations of repeated regions. The genomes are not to scale.

1 A long unique sequence flanked by large direct terminal repeats (CCV, EHV-2, HHV-6 and HHV-7) (Chousterman *et al.*, 1979; Davison, 1992; Browning and Studdert, 1989; Telford *et al.*, 1995; Martin *et al.*, 1991; Gompels *et al.*, 1995; Nicholas, 1996).

2 A long unique sequence flanked by multiple smaller tandem direct repeats at the genome termini (HVS and HHV-8) (Bornkamm *et al.*, 1976; Albrecht *et al.*, 1992; Russo *et al.*, 1996).

3 A long unique sequence flanked by multiple smaller tandem direct repeats at the genome termini, and in inverse orientation at an internal location (LHV-1) (Cebrian *et al.*, 1989).

4 Multiple smaller direct tandem repeats at the genome termini plus a different sequence repeated a variable number of times at an internal location (EBV) (Given and Kieff, 1979; Baer *et al.*, 1984).

5 A short unique sequence flanked by large inverted repeats, and a long unique sequence flanked by smaller inverted repeats (EHV-1 and VZV) (Whalley *et al.*, 1981; Telford *et al.*, 1991; Dumas *et al.*, 1981; Davison and Scott, 1986).

6 Short and long unique sequences each flanked by a pair of large inverted repeats; small direct repeat (the a sequence) at the genome termini, also present in inverse orientation at the IRL/IRS junction (HSV-1 and HCMV) (McGeoch *et al.*, 1988; Weststrate *et al.*, 1983; Chee *et al.*, 1990).

1.1.5.2. Genetic classification

Comparative genetic data used as the sole criteria also support a subdivision of mammalian herpesviruses into three subfamilies termed the α-, β-, γ-herpesviruses (reviewed by: Davison, 1993; Davison and McGeoch, 1995; McGeoch *et al.*, 1993, 1995). Perhaps surprisingly, this division for the most part agrees with the formal system. In the genetic scheme the genera Simplexvirus and Varicellovirus correspond to the α_1- and α_2-herpesviruses, the genera Lymphocryptovirus and Rhadinovirus correspond to the γ_1- and γ_2-herpesviruses (Honess, 1984), and the genera Cytomegalovirus/Muromegalovirus and Roseolovirus to the β_1- and β_2-herpesviruses.

Only a few viruses were wrongly classified using the formal system: HHV-6 is genetically a β-herpesvirus and not a member of the Gammaherpesvirinae as suggested by its lymphotropism (Lawrence *et al.*, 1990). Similarly, the lymphotropic MDV and HVT were originally classified as members of the Gammaherpesvirinae (Roizman *et al.*, 1981), their gene content most closely resembles that of the α-herpesviruses (Buckmaster *et al.*, 1988). EHV-2 was suggested as a member of the Betaherpesvirinae, on the basis of its growth properties (Plummer *et al.*, 1969; Roizman *et al.*, 1981), however its DNA sequence confirmed that it is a γ-herpesvirus (Telford *et al.*, 1995).

Genetic data, involving DNA sequences and their interpretations in terms of gene organisation and encoded protein sequences, are now the primary tool in classification. For example, HHV-8 is accepted as a γ_2-herpesvirus solely on the strength of sequence data.

1.1.6. Genetic content

1.1.6.1. Genome sequences

Since 1984, the complete sequences of fourteen herpesvirus genomes have been published: those of EBV, VZV, HSV-1, HCMV, EHV-1, HVS, channel catfish virus (CCV), EHV-2, HHV-6, HHV-7, HHV-8, MCMV, murine herpesvirus 68 (MHV-68), bovine herpesvirus 1 (BHV-1), and alcelaphine herpesvirus (AHV-1) (Baer *et al.,* 1984; Davison and Scott, 1986; McGeoch *et al.,* 1988; Chee *et al.,* 1990; Albrecht *et al.,* 1992; Davison, 1992; Telford *et al.,* 1992; Telford *et al.,* 1995; Gompels *et al.,* 1995; Nicholas, 1996; Russo *et al.,* 1996, Rawlinson *et al.,* 1996; Virgin *et al.,* 1997; Schwyzer *et al.,* 1996; Ensser *et al.,* 1997). At least four more have been completed: those of HSV-2, equine herpesvirus 4 (EHV-4), a second strain of HHV-8, and, in this thesis, a second strain of HHV-7. Together, these sequences total well over two million base pairs.

Analysis of DNA sequences has provided comprehensive insight into genetic content. Herpesviruses contain between about 70 genes in smaller genomes (e.g. HSV-1 and VZV) to perhaps 200 genes in larger genomes (e.g. HCMV). The genes are densely packed and arranged in about equal numbers on each DNA strand. Most are expressed as single exons from their own promoters, although families of genes arranged tandemly on the same strand commonly share a single polyadenylation site (Wagner, 1985). Overlap between genes in different reading frames on the same strand or on opposing strands is rare and usually not extensive. Few genes are spliced in the α-herpesviruses, but splicing is more common in the β- and γ-herpesviruses (McGeoch and Davison, 1995; Davison and Clements, 1996).

The extensive availability of sequence information has also facilitated comparisons between herpesviruses. DNA sequence comparisons are of limited general use in the evaluation of detailed genetic relationships, because of the wide divergence of the family. Relationships are most easily detected at the level of primary amino acid sequence. Mapping and analysis of open reading frames (ORFs: DNA sequences clear of in-frame termination codons that are potentially translatable into proteins) have allowed comparative studies of individual genes and permitted extensive and detailed understanding of the relationships between the herpesviruses. In particular it was apparent at a relatively early stage that the function of a gene in a poorly

understood genome could be inferred from its similarity to a gene from a better characterised genome, such as that of HSV-1 (Davison and Wilkie, 1983; Davison, 1993).

1.1.6.2. Gene conservation

When the genetic contents of the sequenced herpesviruses were compared it became clear that members of the same subfamily share the great majority of genes in a very similar layout. For example, the rather distantly related α-herpesviruses, VZV and HSV-1, share 64 genes arranged colinearly in both genomes (McGeoch and Schaffer, 1993). This has also been demonstrated for the γ herpesviruses (HVS and EBV) and β-herpesviruses (HCMV and HHV-6; Albrecht *et al.,* 1992; Gompels *et al.,* 1995). Differences are much more marked however, when comparisons are made between members of different subfamilies (Davison, 1993), both in that divergence of individual genes is greater and in that the order of conserved genes is disrupted by large scale rearrangement of gene blocks (Davison and Taylor, 1987; Chee *et al.,* 1990; Albrecht *et al.,* 1992). When the genomes of EBV and VZV were compared, conserved genes were observed to be separated into four blocks which are arranged differently in each genome (Davison and Taylor, 1987). Similar comparisons involving HCMV (Chee *et al.,* 1990) showed conserved genes to be present in seven rearranged gene blocks.

About 40 ubiquitous (or "core") genes have been identified which have homologues in all three subfamilies (Davison and Taylor, 1987; Chee *et al.,* 1990; Albrecht *et al.,* 1992; Telford *et al.,* 1992, 1995; Gompels *et al.,* 1995; Nicholas, 1996). The estimate of core genes is approximate, as several genes exhibit lower or undetectable levels of amino acid homology and yet have apparent counterparts in each genome that are similar in size, location, orientation and (in some cases) distribution of hydrophobic residues (Davison, 1993). Relatively few genes are found in representatives of two subfamilies and not the third.

The core genes are located in the long unique region of herpesvirus genomes. They are not restricted to any of the five functional groupings mentioned above (in Section 1.1.4.), but are presumably responsible for aspects common to the lifestyle of all herpesviruses (Davison, 1992; McGeoch and Davison, 1995).

In addition to the core genes, each subfamily contains a subset of unique (or non-core) genes, which are presumably responsible for survival in particular ecological niches. Some genes are unique to genera. For example, EBV encodes latency genes that are not found in HVS (Baer *et al.,* 1984; Albrecht *et al.,* 1992), and HCMV contains several genes lacking counterparts in HHV-6 (Chee *et al.,* 1990; Gompels *et al.,* 1995). Many non-core genes have roles in latency, pathogenesis or immune evasion.

The gene complements of herpesviruses have also diverged by the acquisition of additional cellular genes and development of multigene families (McGeoch *et al.,* 1990). Core genes with cellular homologues (e.g. DNA polymerase), probably derived at very early times from the host cell, and non-core genes with cellular counterparts probably represent more recent acquisitions from the host genome which supply functions of value to a virus in a particular niche (McGeoch and Davison, 1995).

1.1.7. Genome evolution

1.1.7.1. Mechanisms

The herpesviruses are a very divergent family and appear to have used two broad types of molecular mechanisms to arrive at their present state; namely extensive nucleotide mutation and several types of recombination (Davison and McGeoch, 1995; McGeoch, 1989).

Nucleotide mutation involves changes to the base composition of herpesvirus genomes by local substitution/insertion/deletion changes, the end result being the modification of protein functions or the generation of genes *de novo*, (for example the US11 gene of HSV-1; Rixon and McGeoch, 1984). The very large variation in base composition within the Herpesviridae is striking, stretching from 32 to 75% G+C (canine herpesvirus 1 and B-virus, respectively; Honess, 1984), even between herpesviruses in the same genus. PRV and VZV, for example, have G+C compositions of 74% and 46%, respectively. Intragenomic heterogeneity in base composition is also marked, with repeat elements commonly displaying higher G+C compositions than unique sequences (McGeoch, 1989). For example, the HVS genome comprises a unique sequence of 35% G+C flanked by non-coding tandem direct repeats of 71% G+C (Borkmann *et al.,* 1976; Albrecht *et al.,* 1992). The force behind nucleotide compositional bias has not been identified,

although an error-prone DNA polymerase (in HSV-1) (Hall *et al.,* 1985; Honess, 1984) and differing sizes of nucleotide pools caused by the presence or absence of specific enzymes involved in nucleotide anabolism, such as thymidylate synthase (Honess, 1984; Honess *et al.,* 1989), have been suggested as playing a part.

Another interesting facet of nucleotide mutation is "CpG suppression" (shortage of the CG dinucleotide), which is linked to mutations caused by methylation of cytosine residues in CG doublets. CpG suppression is observed commonly throughout γ-herpesvirus genomes, only at certain locations in β-herpesviruses (notably the IE gene regions), and not at all among the α-herpesviruses (Honess *et al.,* 1989; Chee *et al.,* 1990; Gompels *et al.,* 1995; Nicholas, 1996). The site of methylation of herpesviruses is thought to be the latent genome (Honess et al., 1989).

Recombination processes have played roles in genome evolution by gene duplication, gene capture and rearrangement of genes on a small or large scale (McGeoch, 1989). The presence of gene families (or sets of related genes) is ample evidence that gene duplication has occurred in the herpesviruses. Gene families grow by the process of gene duplication, divergence and eventual partitioning of function, and are most common amongst the β-herpesviruses. For example, HCMV contains nine different gene families (Chee *et al*,. 1990). Gene duplication on a smaller scale has also been reported in α-herpesvirus genomes (McGeoch, 1990).

All herpesviruses contain genes with clear cellular homologues which have presumably been obtained by capture of host genes. This has probably occurred by reverse transcription of cellular mRNAs rather than by integration of genomic DNA, since most captured genes are not spliced. Many captured genes encode enzymes involved in nucleotide anabolism and DNA repair, such as DNA polymerase, dUTPase, thymidine kinase, uracil-DNA glycosylase and thymidylate synthase (Harrison *et al.,* 1991; Mullaney *et al.,* 1989; Honess *et al.,* 1986; Thompson *et al.,* 1987). Others may function in modulation of immune or pathogenic processes: for example, the HSV-1 RL1 gene (McGeoch and Barnett, 1991) and possibly the HHV-6 adeno-associated virus type 2 rep gene (Thomson *et al.,* 1991; Gompels *et al.,* 1995). However, it is

notable that no unambiguous cellular homologue has been identified for any herpesvirus structural gene.

Large scale differences in gene layout between the herpesvirus subfamilies show that gene rearrangement occurred in ancient times (Chee *et al.,* 1990; Albrecht *et al.,* 1992; Davison and Wilkie, 1983; Davison and Taylor, 1987). Evidence of gene rearrangement during relatively recent herpesvirus evolution comes from studies of the Us regions of HSV-1 and VZV (Davison and McGeoch, 1986) and from EHV-1 (Telford *et al.,* 1992). Gene rearrangement appears to have depended largely on the expansion and contraction of the flanking inverted repeats.

1.1.7.2. Phylogeny

The conserved set of mammalian herpesvirus core genes has provided strong evidence that the three subfamilies evolved from a common evolutionary origin (Davison, 1992; McGeoch *et al.,* 1995). This points to a common ancestor for mammalian and (from incomplete sequence data) avian herpesviruses which would have contained the set of core genes, plus, perhaps, other genes that are absent from present-day herpesviruses (McGeoch, 1989). Fig. 2a. is an illustrative tree of the Herpesviridae, showing descent of two major groups from a putative ancestor (McGeoch *et al.,* 1995).

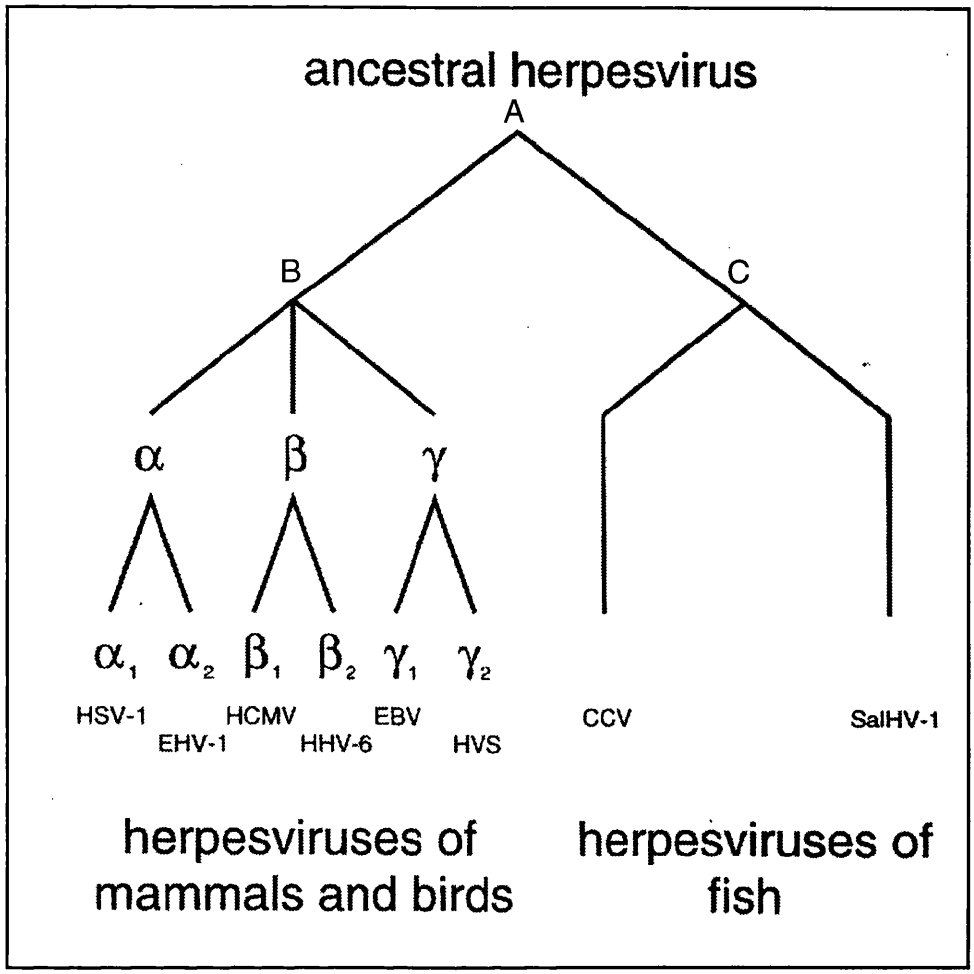
Considering the extremes of divergence displayed by the family, it seems likely that the herpesvirus constitute a very ancient lineage and that the evolutionary space they have explored is extensive (McGeoch, 1992). It is also evident that the herpesvirus evolutionary clock must run faster than that of their hosts (McGeoch, 1989). Many core genes may have been captured from host cells and subsequently adapted to the needs of the virus, as mentioned above, but no convincing examples of cellular homologues of major structural genes have yet been identified. This may mirror the disparity between cellular and viral structure, in contrast to the functions of DNA replication and nucleotide metabolism and repair which are likely to be similar for the virus and host cell. It may be that viral structural genes did evolve from cellular precursors, but that their sequences have been changed radically during evolution to fit novel roles (McGeoch, 1992).

Fig. 2. Herpesvirus evolution

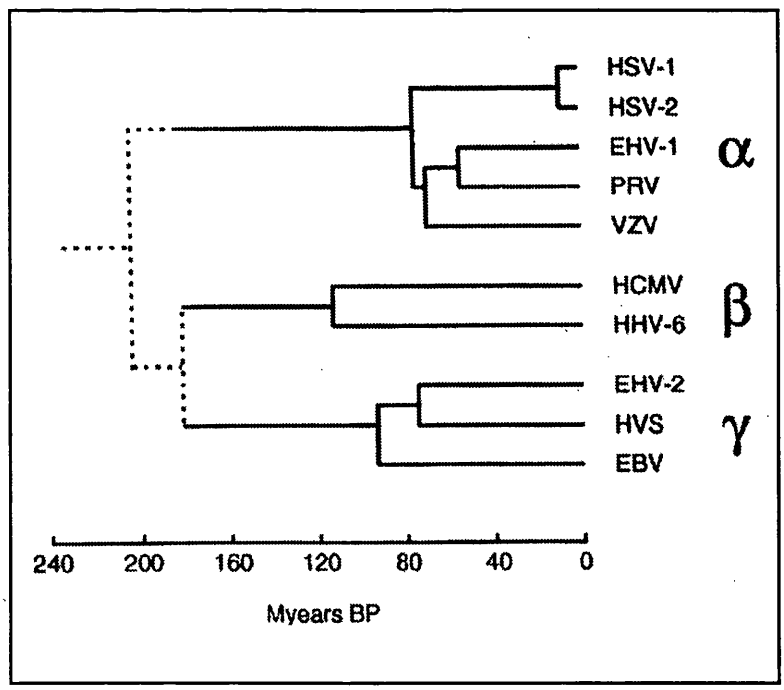
A) A scheme for the evolution of present day herpesviruses from ancestors B and C, which may in turn have originated from ancestor A. It is clear that the mammalian/avian herpesviruses are related and have descended from an ancestral herpesvirus (B) which itself was probably a recognisable herpesvirus with at least 40 genes. Fish herpesviruses have evolved from another progenitor (C) of similar complexity. Nevertheless, the mammalian/avian and fish herpesviruses share evolutionary origins, this places their ancestor (A) much farther in the past. From Davison and McGeoch (1995).

B) Phylogenetic tree distinguishing the α-, β- and γ-herpesvirus subfamilies. Horizontal portions of branches are proportional to assigned divergence values. Subfamily groupings are indicated by Greek letters. The timescale is shown at the foot. The oldest part of the tree is shown as a broken line to indicate the lower level of confidence in data for this region. From McGeoch *et al.* (1995).

A


ancestral herpesvirus
A
B
C
α
β
γ
α1 α2 β1 β2 γ1 γ2
HSV-1
EHV-1
HCMV
HHV-6
EBV
HVS
CCV
SalHV-1
herpesviruses of mammals and birds
herpesviruses of fish

B


HSV-1
HSV-2
EHV-1
PRV
VZV
α
HCMV
HHV-6
β
EHV-2
HVS
EBV
γ
240
200
160
120
80
40
0
Myears BP

The phylogeny of mammalian and avian herpesviruses has resulted in a rather self-contained view of herpesvirus evolution. Our understanding, however, has expanded recently owing to an analysis of the DNA sequence of a fish herpesvirus, CCV, (Davison, 1992). The CCV gene set, with the exception of a few captured genes, lacks any convincing genetic relationship to the mammalian and avian herpesviruses that would indicate a common origin.

Two possible conclusions may be drawn from these findings. CCV may have evolved completely separately from mammalian and avian herpesviruses. Although feasible, this interpretation seems unlikely. CCV particles have the complex virion morphology and capsid structure of herpesviruses (Davison and Davison, 1995; Booy *et al.,* 1991), a group 1 genome similar in size to that of some other herpesviruses (see Fig. 1), and certain aspects of gene arrangement like those of other herpesviruses. Alternatively, CCV and the mammalian and avian herpesviruses may share a common ancestor, but have diverged to an extent that a genetic relationship is no longer detectable. At present, this appears to be the more reasonable view.

Preliminary studies of other fish herpesvirues, salmonid herpesvirus 1 (SalHV-1) (A. J. Davison, unpublished data) and salmonid herpesvirus 2 (SalHV-2) (Bernard and Mercier, 1993) indicate a distant relationship to CCV rather than to mammalian/avian herpesviruses. Thus, it is likely that CCV is a member of a group of relatively little studied fish herpesviruses that is comparable in genetic complexity and evolutionary breadth to the mammalian/avian herpesviruses.

1.1.7.3. Evolutionary timescale

Recent reports, based on molecular sequence analysis (McGeoch and Cook, 1994; McGeoch *et al.,* 1995), have provided a detailed phylogeny and evolutionary timescale for the mammalian/avian herpesviruses.

A robust phylogenetic tree was assembled which clearly distinguished the three recognised herpesvirus subfamilies and the major sublineages (genera), this is illustrated in Fig. 2b. A root position for the tree was established and a timescale was constructed for the tree. The timescale was constructed from the mammalian fossil record, on the assumption that the herpesvirus speciation leading to contemporary herpesviruses within each of the sublineages to a large

extent represents cospeciation with hosts. It was then estimated that the three subfamilies arose approximately 180-220 million years before present, that the major sublineages diverged 80-60 million years before present (probably before the mammalian radiation) and that herpesvirus speciation within the sublineages took place in the last 80 million years.

1.2. HUMAN CYTOMEGALOVIRUS

1.2.1. Basic characteristics

HCMV strains were independently isolated by Smith (1956), Rowe *et al,* (1956) and Weller *et al.* (1957). HCMV is formally designated as human herpesvirus 5 (HHV-5) the prototypic β-herpesvirus. As a member of the Betaherpesvirinae it is most closely related to HHV-6 and HHV-7, which contain homologues to the US22 gene family, in addition to the herpesvirus core genes which are located in the HCMV genome between genes UL45 and UL114 (Lopez and Honess, 1990; Gompels *et al.,* 1995) (see Sections 1.2.6., 1.2.7. and 1.2.9.). HCMV-infected cells contain particles with characteristic herpesvirus morphology (Irmiere and Gibson, 1983 and 1985; Sarov and Abady, 1975; Smith and de Harven, 1974, 1978), in addition to "dense bodies" in almost equal numbers (Craighead *et al.,* 1972; Sarov and Abady, 1975; Irmiere and Gibson, 1983; Klages *et al.,* 1989; Landini *et al.,* 1987) and less abundant "non-infectious enveloped particles" (Sarov and Abady, 1975). HCMV replication *in vitro* is characterised by the appearance of foci containing large refractile cells (or cytomegaly) (Weller *et al.,* 1957).

1.2.2. Epidemiology

HCMV is ubiquitous in the human population, and the prevalence of antibodies in normal adults ranges from 40 to 100% (Krech, 1973). Transmission of HCMV occurs by person-to-person contact and infection can occur at any age. Congenital infection is relatively uncommon, but transmission during birth and from breast milk is very common in developing countries. Infants and young children excrete HCMV in their urine and respiratory tract and may be sources of infection for adults and other children (Stagno and Cloud, 1994). The most important mode of transmission post puberty is probably sexual (Handsfield *et al.,* 1985), especially among sexually active male homosexuals (Drew *et al.,* 1981; Collier *et al.,* 1987), and reinfection is common in sexually active populations (Drew *et al.,* 1984; Spector *et al.,* 1984). HCMV can also be transmitted via blood products and transplanted organs. (Van der Meer *et al.,* 1996)

1.2.3. Cellular tropism

HCMV exhibits tropism *in vitro* most commonly for primary human fibroblast cells. Interestingly, HCMV can be recovered from many different types of cell *in vivo*, including epithelial cells,

peripheral blood mononuclear cells, cells derived from the central nervous system and endothelial cells (Britt, 1996).

1.2.4. Growth properties

Although HCMV is a lytic virus, individuals rarely exhibit clinical evidence of infection. Several *in vitro* properties of the virus, including slow replication, restricted cell tropism, and limited cell-to-cell spread, may contribute to its limited pathogenicity in normal hosts. This important aspect of HCMV biology has been reviewed recently by Stenberg and Kerry (1995) and Mocarski (1996).

As in other herpesviruses, HCMV attachment and cell membrane fusion and penetration are mediated by viral envelope glycoproteins (Rasmussen, 1990; Keay and Baldwin, 1991). Following entry of the virion into the cell the genome is transported to the cell nucleus (Mocarski, 1996).

Expression of HCMV genes is highly regulated and has been divided into three phases: immediate early (IE), early (E) and late (L). Expression of IE genes does not require protein synthesis (DeMarchi, 1981; Stinski *et al.,* 1983, 1991), and the encoded proteins are thought to be involved in the initiation and modulation of virus replication (Stinski *et al.,* 1991). Following IE gene expression, the protein products of the E genes can be detected, many encoding replicative enzymes such as the DNA polymerase. Late gene expression then occurs with the production of virion structural proteins. The virus is notoriously slow growing *in vitro*, and older studies have documented a replicative cycle of nearly 48 hours.

Virus assembly begins with formation of the capsid (Gibson *et al.,* 1993), into which newly replicated concatameric DNA is packaged as unit-length molecules. Tegument components and the envelope are then added and the mature virion is released either by lysis of the infected cell or by a process similar to reverse endocytosis (Britt, 1996).

Like all herpesviruses, HCMV remains with its host for life after primary infection. However, a true latent infection with HCMV has not been demonstrated in the human host, and the site(s) of viral persistence or latency remain unclear, although sensitive polymerase chain reaction (PCR)

methods have identified a monocyte population in peripheral blood of healthy carriers as a site of viral DNA persistence (Taylor-Wiedeman *et al.,* 1991; Mocarski, 1996). Persistence in the presence of an ongoing host immune response remains a characteristic of HCMV, and therapeutic immunosuppression commonly leads to activation of viral replication. Whether the virus enters latency, or continues persistent low level replication, has not been resolved.

1.2.5. Disease and therapy

Primary infection with HCMV in normal adults is often asymptomatic but may be manifested by a self-limiting mononucleosis-like syndrome in less than 1:1000 cases (Horwitz *et al.,* 1986). In contrast, infections in immunocompromised hosts range from similar asymptomatic infection to fulminant life-threatening disease. HCMV is associated with different diseases depending on type and level of immunosuppression.

In AIDS patients, HCMV infection is commonly manifest as retinitis and gastrointestinal dysfunction (Jacobson and Mills, 1988; Jabs *et al.,* 1989; Dietrich and Rahmin, 1991; Gallant *et al.,* 1992; Spector *et al.,* 1993). In allograft transplantations the majority of morbidity for HCMV disease occurs when the recipient is seronegative but the donor is seropositive (Ho *et al.,* 1975; Smiley *et al.,* 1985; Meyers *et al.,* 1986; Appereley and Goldman, 1988; Winston *et al.,* 1990; Wingard *et al.,* 1988). The risk for and the manifestation of HCMV disease are associated with the organ of transplant. For example, pneumonia is a major problem in bone marrow or lung transplant recipients, whereas hepatitis is quite common after liver transplantation.

The risk of fetal infection with HCMV correlates with the lack of pre-existing seroimmunity in women infected with the virus during pregnancy. Infants with congenital infection exhibit many of the findings of both AIDS patients and allograft recipients infected with HCMV (Becroft, 1981; Boppana *et al.,* 1992). Post-natally, HCMV infections are self-limiting, but frequent central nervous system (CNS) sequelae result from damage secondary to fetal infection (Stagno *et al.,* 1983; Demmler, 1991; Boppana *et al.,* 1992; Fowler *et al.,* 1992; Williamson *et al.,* 1992).

The therapeutic agents for invasive HCMV infections include two inhibitors of the viral DNA polymerase, foscarnet and ganciclovir (Faulds and Heel, 1990; Chrisp and Clissold, 1991). However, both agents are toxic (Chrisp and Clissold, 1991) and viral resistance can develop

following treatment (Faulds and Heel, 1990). A third agent, HPMPC, or cidovir, also exhibits potent antiviral activity against HCMV (De Clercq, 1993), but its use is limited by its toxicity (Lalezari *et al.,* 1995; Cherrington *et al.,* 1996).

1.2.6. The HCMV genome

The features of the HCMV genome discussed below have been derived largely from Weston and Barrell (1986), Chee et al. (1990) and Bankier et al. (1991). Table 1 lists the genes present in the HCMV genome, and Fig. 3 shows the genome structure and relative arrangement of the genes

1.2.6.1. Size and structure

The most widely used HCMV strain in the laboratory is a vaccine strain derived from multiple passage in cell culture, AD169. The genome of HCMV strain AD169 is 229,354 bp in length. The nucleotide composition is 57.2% G+C, but varies across the genome (Chee *et al.,* 1990). The AD169 genome structure is in group 1, similar to that of HSV-1 (Fig. 1). It comprises a long unique sequence (U_L: 166,972 bp) flanked by inverted repeats (TR_L: each 11,247 bp), and a short unique sequence (U_S: 35,418 bp) also flanked by inverted repeated sequences (IRS and TRS: both 2,514 bp). Each unique sequence bounded by repeated sequences is free to invert, and hence the genome is found in four equimolar isomers (Fig 3).

Interestingly, recent findings suggest that compared with the published DNA sequence, most isolates of HCMV strain AD169 contain an additional 929 bp after nucleotide 54,612 (Dargan *et al.,* 1997). This sequence is unrelated to the 15 kbp fragment of additional DNA present in clinical isolates and the Toledo strain of HCMV but not in AD169 (Cha *et al.,* 1996).

1.2.6.2. Origin of DNA replication

The HCMV origin of lytic DNA replication (ori_L) is adjacent to the gene encoding the single-stranded DNA binding protein (UL57). Ori_L is complex and maps to a 1.5 kbp region that is extremely rich in repeated sequences and binding sites for various transcription factors (Anders *et al.,* 1992; Anders and Punturieri, 1991; Hamzeh *et al.,* 1990; Masse *et al.,* 1992). This contrasts with the lytic or latent origins of replication in EBV or HSV-1 (Mocarski and Roizman,

TABLE 1. HCMV genes and encoded proteins

Frame	Strand	Start	K-ATG	Stop	Length	MW	Old Name	(ref)	Family	Comments
HCMVJ1L	C	3		929	309	33 176		1		Overlaps J1I & J1S
HCMVTRL1		934	970	1 902	311	34 822	HKLF1	1		= HCMVIRL1
HCMVTRL2		1 893		2 237	115	12 324				= HCMVIRL2
HCMVTRL3		3 141	3 192	3 533	114	13 252				= HCMVIRL3. Glycoprotein?
HCMVTRL4	C	3 785		4 435	217	24 929				= HCMVIRL4. ORF in major early transcript (GREENAWAY and WILKINSON 1987)
HCMVTRL5		4 185	4 266	4 607	114	12 835				= HCMVIRL5
HCMVTRL6	C	5 615	5 947	6 010	111	12 286				= HCMVIRL6
HCMVTRL7	C	6 598	6 843	6 921	82	9 718				= HCMVIRL7
HCMVTRL8		7 227	7 284	7 670	129	14 302				= HCMVIRL8
HCMVTRL9		7 501		7 929	143	15 909				= HCMVIRL9
HCMVTRL10		8 101	8 182	8 694	171	19 035				= HCMVIRL10; D at position 38 is N in IRL10. Glycoprotein
HCMVTRL11		8 648	8 726	9 427	234	26 661			RL11 family	= HCMVIRL11. Glycoprotein
HCMVTRL12		9 431	9 434	10 681	416	47 417			RL11 family	= HCMVIRL12. Glycoprotein
HCMVTRL13		10 778	10 796	11 236	147	15 888				= HCMVIRL13. Glycoprotein exon?
HCMVTRL14		11 140	11 143	11 700	186	21 827			RL11 family	First 35 amino acids identical in IRL14. Glycoprotein exon?
HCMVUL1		11 771	11 810	12 481	224	25 578			RL11 family	Glycoprotein
HCMVUL2	C	12 868	13 047	13 131	60	6 763				Glycoprotein exon?
HCMVUL3	C	13 010	13 324	13 330	105	12 307				
HCMVUL4		13 434	13 464	13 919	152	17 751			RL11 family	Glycoprotein exon?
HCMVUL5		13 986	14 013	14 510	166	18 861			RL11 family	Glycoprotein exon?
HCMVUL6		14 522	14 612	15 463	284	31 447			RL11 family	Glycoprotein exon?
HCMVUL7		15 523	15 526	16 191	222	24 354			RL11 family	Glycoprotein exon?
HCMVUL8		16 198	16 234	16 599	122	13 787			RL11 family	Glycoprotein exon?
HCMVUL9		16 606	16 612	17 295	228	26 889			RL11 family	Glycoprotein
HCMVUL10		17 222		18 199	326	37 366			RL11 family	Glycoprotein exon?
HCMVUL11		18 268	18 295	19 119	275	31 382			RL11 family	Glycoprotein
HCMVUL12	C	19 103	19 321	19 351	73	8 250				Glycoprotein exon?
HCMVUL13		19 143	19 320	20 738	473	54 614				Glycoprotein exon?
HCMVUL14		20 798	20 843	21 871	343	38 567				Glycoprotein
HCMVUL15	C	21 639		22 604	322	35 338				
HCMVUL16		22 342	22 414	23 103	230	26 148				Glycoprotein
HCMVUL17		23 151	23 214	23 525	104	12 672				
HCMVUL18		23 631	23 637	24 740	368	41 736	H3O1	2		Glycoprotein homologous to class 1 HLA (BECK and BARRELL 1988)
HCMVUL19		24 701	24 740	25 033	98	11 281				
HCMVUL20		25 233	25 299	26 318	340	38 703				Glycoprotein. Homologous to TCR-γ?
HCMVUL21	C	26 500	27 024	27 039	175	19 940				
HCMVUL22	C	27 263		27 646	128	14 132				Hydrophobic
HCMVUL23	C	27 866		28 891	342	39 341			US22 family	
HCMVUL24	C	28 936	30 009	30 171	358	40 187			US22 family	
HCMVUL25		30 030	30 057	32 024	656	73 541			UL25 family	
HCMVUL26	C	32 212	32 775	32 994	188	21 156				
HCMVUL27	C	32 834	34 657	34 723	608	69 222				
HCMVUL28	C	34 757		35 893	379	42 739			US22 family	
HCMVUL29	C	35 926	37 005	37 092	360	40 779			US22 family	
HCMVUL30	C	37 138	37 500	37 533	121	14 047				
HCMVUL31		37 682		39 763	694	76 061				
HCMVUL32	C	39 850	42 993	43 050	1048	112 689				Large structural phosphoprotein (pp150) (JAHN et al. 1987)
HCMVUL33		43 128	43 251	44 420	390	43 806			GCR family	Multiply hydrophobic. Homology to G-protein-coupled receptors
HCMVUL34		44 500		46 011	504	56 185				
HCMVUL35		46 042	46 093	48 012	640	72 531			UL25 family	
HCMVUL36EX2	C	48 246		49 751	408.7	47 518	HJLF4	3	US22 family	
HCMVUL36EX1	C	49 354	49 776	49 863	67.3	7 483	HJLF3	3	US22 family	
HCMVUL37EX3	C	49 913		50 842	310	35 476	HJLF2	3		IE glycoprotein exon 3
HCMVUL37EX2	C	50 893		51 015	14.3	1 561	HJLF1	3		IE glycoprotein exon 2
HCMVUL38	C	51 131	52 123	52 138	331	36 738	HZLF3	3		
HCMVUL37EX1	C	52 218	52 706	52 763	162.7	19 116	HZLF2	3		IE glycoprotein exon 1
HCMVUL39		53 024		53 395	124	13 533				
HCMVUL40	C	53 216	53 878	53 893	221	24 368				Glycoprotein
HCMVUL41	C	53 936		54 358	141	16 767				
HCMVUL42	C	54 384		54 854	157	17 066				Glycoprotein exon?
HCMVUL43	C	54 604	55 164	55 245	187	20 993			US22 family	
HCMVUL44	C	55 214	56 512	56 668	433	46 234				Encodes ICP36 protein family (LEACH and MOCARSKI 1989)
HCMVUL45	C	56 656		59 400	915	101 670				Homology to large subunit of ribonucleotide reductase (NIKAS et al. 1986)*

Frame	Strand	Start	K-ATG	Stop	Length	MW	Old Name	(ref)	Family	Comments
HCMVUL46	C	59519	60388	60562	290	33028				Capsid assembly? PERTUISET et al. (1989)*
HCMVUL47		60282	60390	63335	982	109962				
HCMVUL48		62921	63335	70057	2241	253227	HFRF0	4		Virion protein? (BATTERSON et al. 1983*; MCGEOCH et al. 1988a)*
HCMVUL49	C	70403	72112	72334	570	63852	HFLF5	4		
HCMVUL50	C	72072	73262	73283	397	42902	HFLF4	4		Glycoprotein?
HCMVUL51	C	73287	73757	73910	157	16968	HFLF3	4		
HCMVUL52		73748	73796	75799	668	74122	HFRF1	4		
HCMVUL53		75789	75795	76922	376	42314	HFRF2	4		
HCMVUL54	C	76906	80631	80655	1242	137104	HFLF2	4		DNA Polymerase (KOUZARIDES et al. 1987a)
HCMVUL55	C	80775	83492	83654	906	102005	HFLF1	4		gB (CRANAGE et al. 1986)
HCMVUL56	C	83458	86007	86019	850	95870	HFLF0	4		
HCMVUL57	C	86577	90281	90326	1235	133880				Major DNA-binding protein (ANDERS and GIBSON 1988)
HCMVUL58		90864		91235	124	14418				
HCMVUL59	C	91205	91573	91597	123	13945				
HCMVUL60	C	92336		92815	160	18241				
HCMVUL61	C	92847		94139	431	44310				
HCMVUL62	C	94114		94764	217	23686				
HCMVUL63		95331		95717	129	14792				
HCMVUL64	C	95904		96203	100	11245				
HCMVUL65		96315		96620	102	11525				Segments in frame with 67-kDa phosphoprotein sequence of DAVIS and HUANG (1985)
HCMVUL66	C	96475		96816	114	13921				
HCMVUL67	C	97098	97436	97451	113	13218				Glycoprotein exon?
HCMVUL68	C	97750	98079	98100	110	12728				
HCMVUL69	C	98202	100433	100532	744	82679				Transactivator? (MCGEOCH et al. 1988a)*
HCMVUL70	C	100536		103721	1062	120928				DNA replication? (MCGEOCH et al. 1988b)*
HCMVUL71		103239		104471	411	45728				
HCMVUL72	C	104558	105721	105751	388	43576				dUTPase? (PRESTON and FISHER 1984)*
HCMVUL73		105629	105737	106150	138	14868				Glycoprotein
HCMVUL74	C	106128	107525	107585	466	54236				Glycoprotein exon?
HCMVUL75	C	107904	110132	110153	743	84453				gH (CRANAGE et al. 1988)
HCMVUL76		110324	110327	111301	325	36070				
HCMVUL77		110787	110907	112832	642	71188				Virion protein? (ADDISON et al. 1984*; MCGEOCH et al. 1988a)*
HCMVUL78		112864	112924	114216	431	47358				
HCMVUL79	C	114277	115161	115779	295	33846				
HCMVUL80		115084	115198	117321	708	73853				Assembly protein read from internal start (ROBSON and GIBSON 1989)
HCMVUL81	C	117311		117658	116	12796				
HCMVUL82	C	117489	119165	119189	559	61950			UL82 family	pp71 (RUGER et al. 1987)
HCMVUL83	C	119355	121037	121094	561	62900			UL82 family	pp65 (RUGER et al. 1987)
HCMVUL84	C	121312	123069	123306	586	65430				
HCMVUL85	C	123104	124021	124090	306	34596				
HCMVUL86	C	124186	128295	128415	1370	153875	HaLF1	5		Major capsid protein (CHEE et al. 1989b)
HCMVUL87		128265	128355	131177	941	104805				
HCMVUL88		131144	131177	132463	429	47691				
HCMVUL89EX2	C	132466		133629	378	42776				Conserved herpesvirus spliced gene (COSTA et al. 1985)*
HCMVUL90	C	133639	133836	133920	66	7445				
HCMVUL91		133784	133835	134167	111	12028				
HCMVUL92		134020	134140	134742	201	22512				
HCMVUL93		134693	134711	136492	594	68464				
HCMVUL94		136008	136353	137387	345	38382				
HCMVUL89EX1	C	137382	138389	138803	296	34323				Conserved herpesvirus spliced gene (COSTA et al. 1985)*
HCMVUL95		138352	138388	139980	531	57214				
HCMVUL96		139821	140016	140360	115	13108				
HCMVUL97		140373	140484	142604	707	78234	HSRF3	6		Phosphotransferase? (CHEE et al. 1989a)
HCMVUL98		142626	142701	144452	584	65273				DNase (MCGEOCH et al. 1986)*
HCMVUL99		144311	144392	144961	190	20924				Phosphoprotein pp28 (MEYER et al. 1988)
HCMVUL100	C	145229	146344	146413	372	42862				Multiply hydrophobic
HCMVUL101		146353		146697	115	12184				DNA replication? Position only (MCGEOCH et al. 1988b)*
HCMVUL102		146747		149140	798	85615				DNA replication? Position only (MCGEOCH et al. 1988b)*
HCMVUL103	C	149311	150057	150108	249	28637				
HCMVUL104	C	150008	152098	152167	697	78508				Virion protein? (WELLER et al. 1983*; MCGEOCH et al. 1988a)*

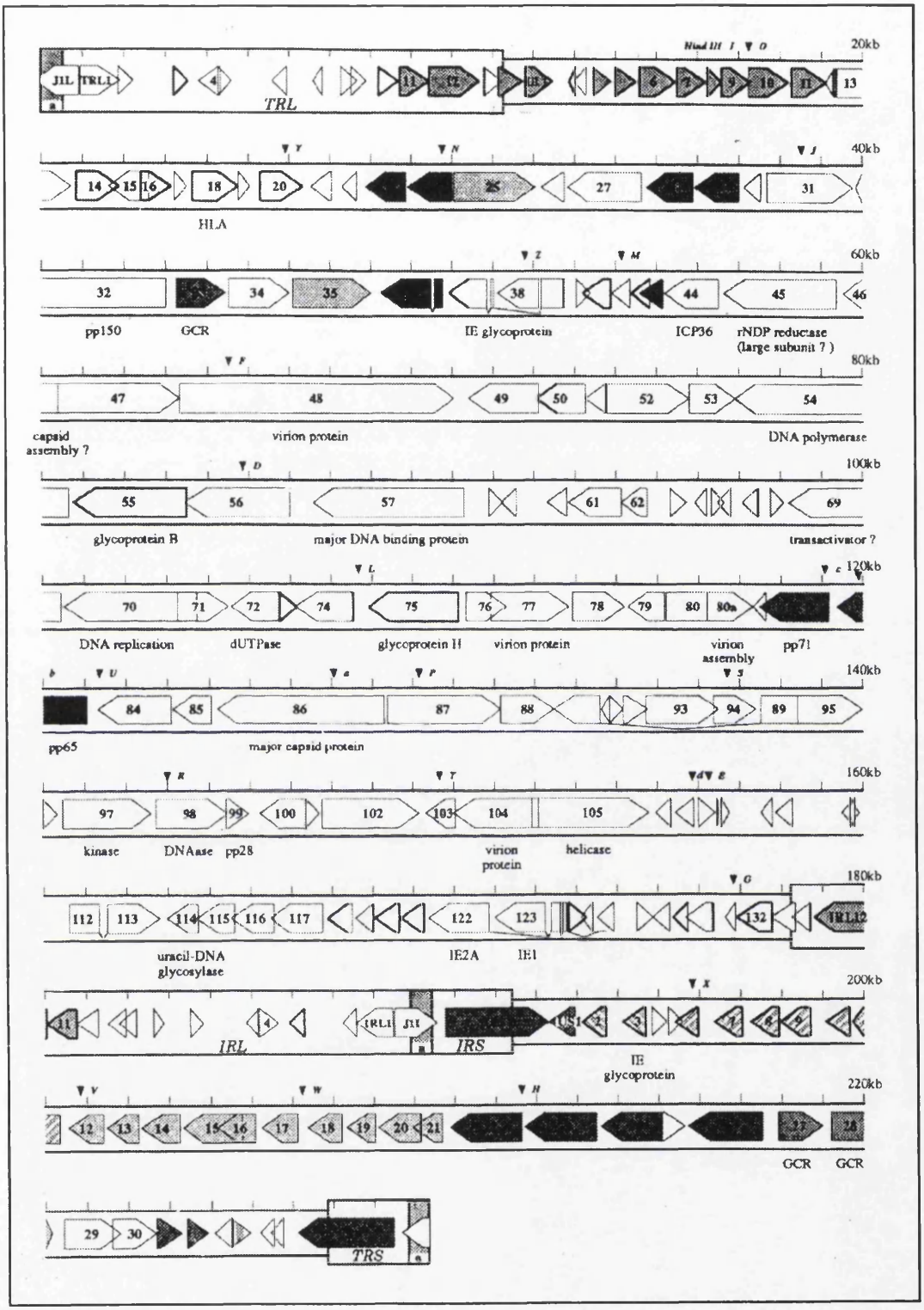
Frame	Strand	Start	K-ATG	Stop	Length	MW	Old Name	(ref)	Family	Comments
HCMVUL105		151 806	151 926	154 793	956	106 501				Helicase (MARTIGNETTI 1987; CRUTE et al. 1989)*
HCMVUL106	C	154 950	155 324	155 330	125	14 500				
HCMVUL107	C	155 420		155 869	150	17 374				
HCMVUL108		156 016		156 384	123	14 501				
HCMVUL109	C	157 517	157 810	157 816	98	11 709				
HCMVUL110	C	157 896		158 276	127	14 224				
HCMVUL111	C	159 479		159 799	107	11 565				
HCMVUL111A		159 615	159 678	159 911	78	8 582				ORF in transforming region (RAZZAQUE et al. 1988)
HCMVUL112		160 484	160 589	161 392	252.3	26 415				Common N-terminus of four phosphoproteins (WRIGHT et al. 1988)
HCMVUL113		161 301		162 797	499	51 105				Probably spliced to UL112; internal splicing? (WRIGHT et al. 1988)
HCMVUL114	C	162 973	163 722	163 758	250	28 354				Uracil-DNA glycosylase (WORRAD and CARADONNA 1988)*
HCMVUL115	C	163 697		164 614	306	34 110				
HCMVUL116	C	164 533		165 564	344	37 519				Glycoprotein exon?
HCMVUL117	C	165 474	166 745	166 757	424	45 464				
HCMVUL118	C	166 861		167 487	209	24 599				Glycoprotein exon?
HCMVUL119	C	167 558	167 983	168 037	142	14 729				Glycoprotein exon?
HCMVUL120	C	168 041	168 643	168 700	201	22 768				Glycoprotein
HCMVUL121	C	168 697	169 236	169 269	180	20 138				Glycoprotein
HCMVUL122	C	169 367		170 878	494.7	51 084				IE2A. Spliced to IE1 EX4. Also KATG at 170599 (STENBERG et al. 1985)
HCMVUL123EX4	C	171 009		172 274	405.7	45 622				IE1 gene exon 4 (STENBERG et al. 1984; AKRIGG et al. 1985)
HCMVUL123EX3	C	172 301		172 654	61.7	6 865				IE1 gene exon 3 (STENBERG et al. 1984; AKRIGG et al. 1985)
HCMVUL123EX2		172 659	172 765	172 873	23.7	2 658				IE1 gene exon 2 (first coding exon) (STENBERG et al. 1984; AKRIGG et al. 1985)
HCMVUL124		172 783	172 798	173 253	152	15 887				Glycoprotein
HCMVUL125	C	173 114		173 419	102	11 000				
HCMVUL126	C	173 508		173 909	134	15 910				
HCMVUL127		174 453	174 495	174 887	131	15 248				
HCMVUL128	C	174 868		175 284	139	16 036				
HCMVUL129	C	175 357		175 704	116	13 288				Glycoprotein exon?
HCMVUL130	C	175 665	176 306	176 438	214	24 653				Glycoprotein exon?
HCMVUL131	C	176 644	176 871	177 042	76	8 243				
HCMVUL132	C	176 934	177 743	177 845	270	29 973				Glycoprotein
HCMVIRL14	C	177 776	178 324	178 327	183	20 750				First 35 amino acids identical in TRL14
HCMVIRL13	C	178 231	178 671	178 689	147	15 888				= HCMVTRL13. Glycoprotein exon?
HCMVIRL12	C	178 786	180 033	180 036	416	47 417			RL11 family	= HCMVTRL12. Glycoprotein
HCMVIRL11	C	180 040	180 741	180 819	234	26 661			RL11 family	= HCMVTRL11. Glycoprotein
HCMVIRL10	C	180 773	181 285	181 366	171	19 034				= HCMVTRL10; N at position 38 is D in TRL10. Glycoprotein
HCMVIRL9	C	181 538		181 966	143	15 909				= HCMVTRL9
HCMVIRL8	C	181 797	182 183	182 240	129	14 302				= HCMVTRL8
HCMVIRL7		182 546	182 624	182 869	82	9718				= HCMVTRL7
HCMVIRL6		183 457	183 520	183 852	111	12 286				= HCMVTRL6. Glycoprotein exon?
HCMVIRL5	C	184 860	185 201	185 282	114	12 835				= HCMVTRL5
HCMVIRL4		185 032		185 682	217	24 929				= HCMVTRL4. ORF in major early transcript (GREENAWAY and WILKINSON 1987)
HCMVIRL3	C	185 934	186 275	186 326	114	13 252				= HCMVTRL3. Glycoprotein?
HCMVIRL2	C	187 230		187 574	115	12 324				= HCMVTRL2
HCMVIRL1	C	187 565	188 497	188 533	311	34 822	HKLF1	1		= HCMVTRL1
HCMVJ11		188 538		189 560	341	36 544				Positions 1 to 309 overlap J1L; 118 to 341 overlap J1S
HCMVIRS1		189 702	189 765	192 302	846	91 050	HQRF1	1	US22 family	V at position 190 is L in TRS1. Sequences diverge after position 549
HCMVUS1	C	192 332		192 967	212	23 481	HQLF3	1	US1 family	
HCMVUS2	C	193 119	193 715	193 850	199	23 112	HQLF2	1	US2 family	Glycoprotein
HCMVUS3	C	194 133	194 690	194 924	186	21 575	HQLF1	1	US2 family	Spliced IE glycoprotein (WESTON 1988)
HCMVUS4		194 832		195 188	119	13 089				
HCMVUS5		195 203	195 230	195 607	126	14 451				
HCMVUS6	C	195 403	195 951	195 975	183	20 640	HXLF6	1	US6 family	Glycoprotein
HCMVUS7	C	196 377	197 051	197 069	225	26 271	HXLF5	1	US6 family	Glycoprotein

Frame	Strand	Start	K-ATG	Stop	Length	MW	Old Name	(ref)	Family	Comments
HCMVUS8	C	197256	197936	197960	227	26634	HXLF4	1	US6 family	Glycoprotein
HCMVUS9	C	197954	198694	198772	247	28054	HXLF3	1	US6 family	Glycoprotein
HCMVUS10	C	199083	199637	199646	185	20772	HXLF2	1	US6 family	Glycoprotein
HCMVUS11	C	199716	200360	200366	215	25265	HXLF1	1	US6 family	Glycoprotein
HCMVUS12	C	200549	201391	201562	281	32470	HVLF6	1	US12 family	Multiply hydrophobic
HCMVUS13	C	201474	202256	202307	261	29461	HVLF5	1	US12 family	Multiply hydrophobic
HCMVUS14	C	202328	203257	203311	310	34198	HVLF4	1	US12 family	Multiply hydrophobic
HCMVUS15	C	203305		204756	484	53049	HVLF3	1	US12 family	Multiply hydrophobic
HCMVUS16	C	204153	205079	205091	309	34718	HVLF2	1	US12 family	Multiply hydrophobic
HCMVUS17	C	205227	206105	206144	293	31910	HVLF1	1	US12 family	Multiply hydrophobic
HCMVUS18	C	206376	207197	207266	274	30195	HWLF5	1	US12 family	Multiply hydrophobic
HCMVUS19	C	207338	208057	208132	240	26424	HWLF4	1	US12 family	Multiply hydrophobic
HCMVUS20	C	208107		209177	357	39890	HWLF3	1	US12 family	Multiply hydrophobic
HCMVUS21	C	208978	209694	209793	239	26586	HWLF2	1	US12 family	Multiply hydrophobic
HCMVUS22	C	209874		211652	593	66971	HWLF1	1	US22 family	Early nuclear protein (MOCARSKI et al. 1988)
HCMVUS23	C	211717	213492	213510	592	68886	HHLF7	1	US22 family	
HCMVUS24	C	213591	215090	215105	500	57928	HHLF6	1	US22 family	
HCMVUS25		215097		215633	179	19655				
HCMVUS26	C	215730	217538	217574	603	70022	HHLF5	1	US22 family	
HCMVUS27		217859	217904	218989	362	41996	HHRF2	1	GCR family	Multiply hydrophobic. Homology to G-protein-coupled receptors
HCMVUS28		219083	219200	220168	323	37189	HHRF3	1	GCR family	Multiply hydrophobic. Homology to G-protein-coupled receptors
HCMVUS29		220420	220426	221811	462	51068	HHRF4	1		
HCMVUS30		221537	221618	222664	349	39115	HHRF5	1		
HCMVUS31		222674		223264	197	22936	HHRF6	1	US1 family	
HCMVUS32		223325	223385	223933	183	22058	HHRF7	1	US1 family	
HCMVUS33	C	224075		224485	137	15775	HHLF3	1		
HCMVUS34		224408	224480	224968	163	17767	HHRF8	1		Glycoprotein exon?
HCMVUS35	C	225212		225538	109	12966	HHLF2	1		
HCMVUS36	C	225429		225758	110	12352				
HCMVTRS1	C	226115	228478	228541	788	83983	HHLF1	1	US22 family	L at position 190 is V in IRS1. Sequences diverge after position 549
HCMVJ1S	C	228683		229354	224	23797				Overlaps J1L & J1I

A compilation of reading frames of HCMV strain AD169, from Chee *et al.* (1990). The orientations, coordinates, and theoretical sizes are tabulated, together with the locations of predicted Kozak consensus ATG codons. For spliced genes exon coordinates represent open reading frame coordinates: donor and acceptor positions are not shown. Lengths are shown in amino acids. References given in the comments section are not referred to in this thesis.

Fig. 3. Genetic layout of the HCMV genome

The genome scale is in kbp, and the thinner and thicker portions of the genome denote the unique regions and repeats, respectively. Abbreviations: TRL/S, long/short terminal repeat; IRL/S, long/short internal repeat; J1, junction frame 1. Protein-coding regions are shown as arrowed boxes pointing in the direction translation, and contain the gene nomenclature. ORFs predicted to be expressed as spliced mRNAs are indicated on the diagram. Abbreviations: HLA, human leukocyte antigen homologue; pp, phosphoprotein; GCR, G-protein-coupled receptor homologue; IE, immediate early; ICP, infected cell protein. From Bankier *et al.* (1991).


20kb
40kb
60kb
80kb
100kb
120kb
140kb
160kb
180kb
200kb
220kb
TRL
HLA
pp150
GCR
IE glycoprotein
ICP36
rNDP reductase (large subunit ?)
capsid assembly ?
virion protein
DNA polymerase
glycoprotein B
major DNA binding protein
transactivator ?
DNA replication
dUTPase
glycoprotein H
virion protein
virion assembly
pp71
pp65
major capsid protein
kinase
DNAase
pp28
virion protein
helicase
uracil-DNA glycosylase
IE2A
IE1
IRL
IRS
IE glycoprotein
GCR
GCR
TRS

1982; Spaete and Frenkel, 1985). An origin of DNA replication operating in the latent phase of HCMV infection has not been identified.

1.2.6.3. Genetic organisation

The HCMV genome contains approximately 200 densely packed genes (J1L, TRL1,-TRL14, UL1-UL132, IRL14-IRL1, J1I, US1-US36, TRS1, J1S) which are arranged about equally on both strands of the genome with little overlapping (Fig. 3). Transcription signals are commonly present within the preceding gene, and some genes share common polyadenylation sites. Splicing appears to be limited to four groups of exons: the IE genes (UL36/UL37 and UL122/UL123), which are spliced in a complex manner (Kouzarides *et al.,* 1988; Weston, 1988; Stenberg *et al.,* 1989), a highly conserved late core gene (UL89) of unknown function (Costa *et al.,* 1985) and UL112 which encodes four phosphoproteins by alternative splicing and may also be spliced to UL113.

1.2.7. Gene complement

A comprehensive list of genes and their functions is given in Table 1, and their positions are shown in Fig. 3. The HCMV IE genes map in two major loci, termed IE (UL36-38) and major IE (or MIE) (UL122-UL123), although four more IE genes (TRL4, TRS1-IRS1, US3 and possibly U69) are located elsewhere (Stasiak and Mocarski, 1992; Stenberg *et al.,* 1984, 1985, 1989). The most abundantly expressed transcripts are IE1 and IE2A (UL122 and UL123, respectively), which are expressed through splicing from a single transcription start site. In contrast to the bulk of the HCMV genome, the MIE region is CpG suppressed (Honess *et al.,* 1989).

HCMV DNA replication uses the set of six herpesvirus-common replication functions that have been best studied in HSV-1: DNA polymerase (UL54) (Hirai *et al.,* 1976; Hirai and Watanabe, 1976; Huang, 1975), single-stranded DNA-binding protein (UL57) (Anders and Gibson, 1988; Kemble *et al.,* 1987), a processivity factor (UL44) which associates with DNA polymerase and prevents its dislocation from the DNA template (Ertl and Powell, 1992), and the predicted three-component helicase-primase complex (UL70, UL101/102 and UL105). However, in addition, several other genes may play a role, including the regulatory genes (mentioned above) required to activate gene expression, as well as two genes that have uncertain roles in viral replication,

UL112/113 which encodes four phosphoproteins by alternative splicing (Wright *et al.,* 1988) and UL84 (He *et al.,* 1992).

HCMV proteins involved in nucleotide metabolism and DNA repair include, uracil-DNA glycosylase (UL114), a putative dUTPase (UL72), deoxyribonuclease and the large, but not the small, subunit of ribonucleotide reductase (UL45). HCMV does not encode a thymidine kinase but does encode another kinase which can modify nucleosides, UL97 (Littler *et al.,* 1992; Sullivan *et al.,* 1992, 1993).

The capsid proteins comprise the major capsid protein (UL86), the two components of the intercapsomeric triplex UL46 and UL85, a small protein associated with the exterior surface of the capsid (UL49) and the protease and the major and minor scaffolding proteins encoded by UL80 and UL80a. The tegument is made up of as many as 20 proteins, many of which are phosphorylated (Gibson, 1981, 1983), with the basic phosphoprotein (UL32) and the lower matrix protein (UL83) being most abundant (Gibson, 1983). Minor tegument proteins include the upper matrix protein (UL82), which functions as a transactivator of gene expression (Liu and Stinski, 1992), the product of UL99 which is highly immunogenic (Irmiere and Gibson, 1985; Martinez *et al.,* 1989; Meyer *et al.,* 1988; Pereira *et al.,* 1984) and UL56 which is likely to be involved in virion maturation (Bogner *et al.,* 1993). However, a full understanding of the functions of many tegument proteins has not yet emerged. At least 17 other genes potentially encode additional tegument proteins (Mocarski, 1996). At least 54 genes have characteristics of glycoproteins, as listed in Table 1. Three envelope glycoproteins are encoded by the core genes: gB (UL55), which is likely to have roles in virion penetration into cells, transmission from cell to cell and fusion of infected cells (Navarro *et al.,* 1993), and gH (UL75) which is thought to be involved in membrane fusion and complexes with the product of UL115 (gL) during transport to the cell surface (Kaye *et al.,* 1992; Spaete *et al.,* 1993).

US27, US28 and UL33 make up the G-coupled receptor (GCR) gene family. Each gene has seven potential hydrophobic domains and contains motifs which show homology to the opsin family of cell surface receptors. This diverse family of receptors mediate their responses via G-proteins and transduce different signals in a variety of systems, including vision, olfaction,

memory and learning, and regulation of the circulatory system, among others (Chee *et al.*, 1990). In addition, UL100 and the 10 genes that go to make up the US12 gene family also contain multiple potential hydrophobic domains, and may act in similar ways to the G-coupled protein receptors.

1.2.8. Captured genes

In addition to genes directly involved in replication, several genes have clear counterparts in the cellular genome and may play roles in the virus-host interaction. HCMV gene UL18 encodes a glycoprotein which has homology to cellular HLA class I proteins (Beck and Barrell, 1988; Chee *et al.*, 1990). The MCMV genome also has a MHC class I homologue (Rawlinson *et al.*, 1996), which is thought to contribute to immune evasion through interference with natural killer cell-mediated clearance. UL20, which is in close proximity to UL18, has low level similarity to regions of the human T-cell receptor γ chain (Beck and Barrell, 1991). The function of UL20 is unknown, but if it is a T-cell receptor homologue, the virus may be able to use the protein to interact with CD3 and infect T-cells (Borst *et al.*, 1987; Brenner *et al.*, 1987).

1.2.9. Gene families

HCMV is strikingly different from other herpesviruses in the evident expansion of the genome by gene duplication and divergence (Weston and Barrell, 1986; Chee *et al.*, 1990; Mocarski, 1993). Fifty-two of the predicted open reading frames listed in Table 1 can be grouped Into nine families of related genes: three pairs (UL25 and UL35; UL82 and UL83; and US2 and US3) and six larger groups: US1 gene family, US6 gene family, RL11 gene family, US12 gene family, US22 gene family and GCR gene family. Unfortunately, little is known about the function of many members of these families, although three families (R11, US2 and US6) are predicted to encode glycoproteins, and possibly compose a gene superfamily. Some US22 genes are spliced and expressed as IE proteins in HCMV or MCMV, but it is not known whether this extends to all of the family (Kouzarides *et al.*, 1988; Messerle *et al.*, 1991). There is evidence for a role in transcriptional activation for UL36 (Colberg-Poley *et al.*, 1992) and IRS1/TRS1 (Stasiak and Mocarski, 1992) but similarly it is not known if all US22 gene products are involved in transactivation. The gene families are clustered in U_S and at one end of U_L, parts of the genome

that appear to be unique to the β-herpesviruses. Most are conserved in MCMV (Rawlinson *et al.,* 1996), and some in HHV-6 and HHV-7 (Gompels *et al.,* 1995; Nicholas, 1996).

1.3. HUMAN HERPESVIRUS 6

1.3.1. Basic characteristics

HHV-6 was first isolated in 1986 from the peripheral blood lymphocytes of patients with lymphoproliferative disorders, some of whom were also infected with HIV-1 (Salahuddin *et al.,* 1986). The virus was subsequently isolated from AIDS patients (Becker *et al.,* 1988; Downing *et al.,* 1987; Lopez *et al.,* 1988; Tedder *et al.,* 1987), children with exanthem subitum (ES), patients with a variety of other disorders (Agut *et al.,* 1988; Becker *et al.,* 1988) and healthy adults (Lopez *et al.,* 1988). HHV-6 has a typical herpesvirus morphology (Biberfeld *et al.,* 1987; Yoshida *et al.,* 1989; Kramarsky and Sander, 1992), with a very distinct tegument layer (Roffman *et al.,* 1990). Like HCMV, the most noticeable effect of HHV-6 infection *in vitro* is the enlargement (cytomegaly) of host cells and the occurrence of multinucleated cells (syncytia) 3 to 5 days after infection (Pellett and Black, 1996).

Two major subgroups or variants (A and B) of HHV-6 have been identified by restriction endonuclease analysis, sequencing, monoclonal antibody reactivity and cell culture properties (Ablashi *et al.,* 1993; Schrimer *et al.,* 1991; Aubin *et al.,* 1991). The A strains, including GS and U1102, are less frequently observed and have no definite association with human disease, but are often isolated from immunocompromised patients, suggesting a possible link with AIDS. The prototype of B strains is Z29, which is the etiologic agent of ES (Yamanishi *et al.,* 1988).

1.3.2. Epidemiology

Serological surveys indicate that maternal antibodies to HHV-6 can be detected in more than 90% of infants' serum at birth, but that these decrease during the first six months to as low as 5%. Antibody levels to HHV-6 then increase between six months and five years, with 80% of children developing an antibody response by 13 months (Balachandra *et al.,* 1989; Briggs *et al.,* 1988; Brown *et al.,* 1988; Huang *et al.,* 1992; Yoshikawa *et al.,* 1989; Okuno *et al.,* 1989). In most geographical areas, about 90% of people over the age of two years are HHV-6 seropositive (Linde *et al.,* 1988; Linde *et al.,* 1990; Parker and Weber, 1993; Saxinger *et al.,* 1988; Yoshikawa *et al.,* 1989; Okuno *et al.,* 1989; Levy *et al.,* 1990).

Evidence of nearly identical restriction endonuclease profiles from cases during an ES outbreak in an orphanage nursery (Okuno *et al.,* 1991), and from siblings and mothers of children with ES (Mukai *et al.,* 1994), suggests that HHV-6 is transmitted horizontally to infants, although the mode of transmission remains uncertain. It has been widely reported that HHV-6 is present in the saliva and throat swabs of nearly all adults (Levy *et al.,* 1990; Pietroboni *et al.,* 1989; Pietroboni *et al.,* 1988; Brown *et al.,* 1988; Asano *et al.,* 1992), that HHV-6 antigens can be detected in saliva and bronchial gland epithelium (Fox *et al.,* 1990; Krueger *et al.,* 1990), and that DNA has been detected in saliva by polymerase chain reaction (PCR) (Cone *et al.,* 1993; Jarrett *et al.,* 1990; Kido *et al.,* 1990). Questions have been raised over the specificity of some of the analytical reagents used to detect HHV-6 in these tests, with respect to HHV-7 (Pellett and Black, 1996), but it seems reasonable to suggest that HHV-6 infection can be transmitted through saliva.

Possible vertical transmission of HHV-6 has also been suggested. HHV-6 is present in vaginal swabs, including specimens from pregnant women (Leach *et al.,* 1994; Okuno *et al.,* 1995), suggesting the possibility of perinatal and sexual infection. Also, Hall et al (1994) reported that nearly 30% of new-borns had HHV-6 DNA in their mononuclear cells. In contrast, immunoglobulin (Ig) M with activity against HHV-6 is not found in peripheral blood from new-borns (Farr *et al.,* 1990) and rarely in specimens of cord blood cells (Dunne *et al.,* 1992). In addition, only one of 52 aborted fetuses of HIV-1 positive women was HHV-6 positive by PCR (Ando *et al.,* 1992). HHV-6 DNA sequences have not been found in breast milk (Dunne and Jevon, 1993; Takahashi *et al.,* 1988), suggesting that breast feeding is not a route of transmission.

1.3.3. Cellular tropism

HHV-6 replicates optimally in $CD4^+$ T-cells (Lusso *et al.,* 1988; Takahashi *et al.,* 1989) and displays limited replication in $CD8^+$ T-cells, natural killer cells, monocytes, epithelial cells, brain-derived cells (Levy *et al.,* 1990; He *et al.,* 1996) and EBV infected B lymphocytes (Ablashi *et al.,* 1991). CD4 is not believed to be the membrane receptor used by HHV-6, as infection is not inhibited by anti-CD4 antibodies (Lusso *et al.,* 1989). The cellular receptor of HHV-6 is not known. Interestingly, HHV-6 A strain induces CD4 in $CD3^+$ $CD4^-$ $CD8^+$ lymphocytes, rendering them susceptible to HIV-1 infection (Lusso *et al.,* 1991).

1.3.4. Growth properties

The growth cycle of HHV-6 requires 4 to 5 days for the production of extranuclear enveloped virions (Black *et al.,* 1989; Kramarsky and Sander, 1992). However, a regulatory cascade for the virus has not been described, and little is known of the regulation of specific viral genes, largely as a result of the difficulty involved in growing high titre virus (Pellett and Black, 1996).

Of interest are unusual cytoplasmic invaginations (tegusomes) in the infected cell, in which capsids appear to acquire a tegument layer (Roffman *et al.,* 1990). Tegusomes have been observed in cells infected with HHV-7, but have not been noted in cells infected with other herpesviruses.

At any given time, 5% of the population may be seropositive for HHV-6 IgM (Suga *et al.,* 1992); HHV-6 DNA can be detected in the lymphocytes of up to 90% of healthy individuals (Cone *et al.,* 1993; Cuende *et al.,* 1994; Jarrett *et al.,* 1990), and HHV-6 antigens can be found in salivary glands (Fox *et al.,* 1990; Krueger *et al.,* 1990), lymph node tissue (Levine *et al.,* 1992) and neurons and glial cells in the brain (Challoner *et al.,* 1995). Such a high level of HHV-6 in the human population is unlikely to be due solely to reinfection. It is thought that following primary infection, HHV-6 develops a persistent infection that persists for the lifetime of the host. However, the site(s) of HHV-6 persistent infection is not known, although one report has been proposed that the virus may reside in monocyte/macrophage cells (Kondo *et al.,* 1991).

1.3.5. Disease and therapy

HHV-6 has been cited as possible cause of certain malignancies, such as Hodgkin's disease (Clark *et al.,* 1990; Torelli *et al.,* 1992; Di Luca *et al.,* 1994), non-Hodgkin's lymphoma (Jarrett *et al.,* 1988; Josephs *et al.,* 1988), angioimmunoblastic lymphadendopathy with dysproteinemia (Luppi *et al.,* 1993a), Langerhans cell-hystiocytosis (Leahy *et al.,* 1993), Kaposi's sarcoma (Bovenzi *et al.,* 1993), cervical carcinoma (Chen *et al.,* 1994a) and oral carcinoma (Yadav, *et al.,* 1994). However, a direct involvement of HHV-6 in human malignancies has not been proven, and it is difficult to differentiate virus that is a passenger in tumour tissue from that which has a role in tumour etiology. Nonetheless, HHV-6 DNA has been reported to integrate into host cell chromosomes (Luppi *et al.,* 1993b), and has the ability to transform cultured murine fibroblasts

and human epidermal keratinocytes *in vitro* (Razzaque, 1990; Razzaque *et al.,* 1993; Thompson *et al.,* 1994). In addition, HHV-6 DNA can accelerate the tumourigenesis induced by human papillomavirus-immortalised (HPV) human cervical cells transplanted into nude mice (Chen *et al.,* 1994b).

The role of HHV-6B as the causative agent of ES, also known as roseola infantum and sixth disease, was proposed in 1988 (Yamanishi *et al.,* 1988). In most cases, ES is characterised by a high fever which typically lasts for three days and a rash (roseola) that appears as the fever subsides, lasting one to three more days. However, primary HHV-6 infection may lack the typical rash (Suga *et al.,* 1989; Pruksananonda *et al.,* 1992). The disease normally resolves without complications (Asano *et al.,* 1991 and 1994; Okada *et al.,* 1993), but high fever can in some cases lead to convulsions, respiratory tract and tympanic inflammation, intestinal symptoms, meningitis and meningoencephalitis (Pellett and Black, 1996; Suga *et al.,* 1989; Pruksananonda *et al.,* 1992). In a manner similar to HCMV infection, individuals who acquire primary HHV-6 infection in adulthood can develop a self-limiting febrile illness which resembles infectious mononucleosis (Steeper *et al.,* 1988) or hepatitis (Irving *et al.,* 1990; Sobue *et al.,* 1991).

HHV-6 has been associated with chronic fatigue syndrome (CFS), but serological studies reveal conflicting reports. CFS patients have been reported to have higher than average antibody titres for HHV-6 (Ablashi *et al.,* 1988; Balachandran *et al.,* 1991; Buchwald *et al.,* 1992; Dale *et al.,* 1989; Levine *et al.,* 1992; Read *et al.,* 1990; Reeves *et al.,* 1992) or antibody titres for HHV-6 that are similar to control patients (Gold *et al.,* 1990; Marshall *et al.,* 1991; Wakefield *et al.,* 1988). These results must be interpreted in light of the fact that higher than average antibody titres to other viruses (e.g. HCMV, EBV, measles virus) are also seen in CFS patients compared to controls (Holmes *et al.,* 1987).

Current evidence points to a possible role for HHV-6 in the pathology of multiple sclerosis (MS). HHV-6 has an association with acute nervous system disease (Asano *et al.,* 1992; Caserta *et al.,* 1994; Drobyski *et al.,* 1994; Hall *et al.,* 1994; Huang *et al.,* 1991; Kondo *et al.,* 1993; Suga *et al.,* 1993; Ward and Grey, 1994; Yamanishi *et al.,* 1988). The virus can grow in cells derived from the nervous system (Ablashi *et al.,* 1987), persists in the CNS of children (Caserta *et al.,* 1994), and is present in nearly all adult brains (Challoner *et al.,* 1995; Luppi *et al.,* 1994). In

addition, MS sufferers have higher than normal antibody titres to HHV-6, the virus has been detected in the cerebrospinal fluid of MS sufferers, and the distribution of HHV-6 antigen in their brains appears significantly different from normal (Sola *et al.,* 1993; Wilborn *et al.,* 1994; Challoner *et al.,* 1995).

As with HCMV, there is some evidence that HHV-6 may act as an opportunistic agent in immunosuppressed patients suffering from AIDS or undergoing organ or bone marrow transplants (Okuno *et al.,* 1990). The conditions that are associated with infection in the immunocompromised host include interstitial pneumonia in AIDS, bone marrow and organ transplant patients (Carrigan *et al.,* 1991; Knox and Carrigan, 1994; Cone *et al.,* 1993) and encephalitis in AIDS and bone marrow transplant patients (Drobyski *et al.,* 1993; Carrigan and Knox, 1994). The evidence linking HHV-6 with these conditions is based on the temporal association between the signs of disease, antibody levels to HHV-6 and detection of viral antigens and DNA.

No systematic trials have been performed to test the efficacy of drugs for the treatment of HHV-6 infection *in vivo*. However, several compounds have been tested for their ability to inhibit HHV-6 growth in cell culture. Of the compounds tested, gancyclovir and foscarnet inhibit HHV-6 infectivity more effectively than acyclovir, a pattern shared with HCMV (Agut *et al.,* 1991 and 1989; Burns and Sandyford, 1990; Russler *et al.,* 1989; Streicher *et al.,* 1988). In addition, the cytokines IFN-α and IL-2 exert inhibitory effects on HHV-6 infection in cell culture (Kikuta *et al.,* 1990; Roffman and Frenkel, 1990).

1.3.6. Co-infection with other viruses

HHV-6 has been reported to interact with several other viruses. HHV-6 superinfects B cells latently infected with EBV and may induce EBV lytic replication (Ablashi *et al.,* 1988; Flamand *et al.,* 1993). The observed increase in HHV-6 specific antibodies during HCMV primary infection suggests that HHV-6 may be activated by HCMV. Similarly, HHV-7 primary infection can activate (presumably latent) HHV-6, which can then outgrow the HHV-7 infection (Frenkel and Wyatt, 1992). HHV-6 has been shown to transactivate a HPV promoter and enhance expression of HPV RNA (Chen *et al.,* 1994a).

Perhaps most interesting is that HHV-6 has been proposed as a cofactor with HIV-1, capable of speeding the progression of AIDS. Several lines of evidence support this claim. The HIV-1 transactivator, tat, enhances HHV-6 replication (Sieczkowske *et al.,* 1995); HHV-6 (strains A and B) is a potent transactivator of HIV-1 (Lusso *et al.,* 1989; Ensoli *et al.,* 1989; Horvat *et al.,* 1989); and HHV-6 induces two cytokines, TNF-α and IL-1β (Flamand *et al.,* 1991), which are known to enhance replication of HIV-1 *in vivo.* These findings are made more relevant because HHV-6 and HIV-1 share tropism for $CD4^+$ T-cells (Lusso *et al.,* 1988).

In contrast, studies of the prevalence of antibodies to HHV-6 and HIV-1 are confusing and have variously shown no difference in HHV-6 prevalence between HIV positive groups and control groups (Brown *et al.,* 1988; Essers *et al.,* 1991; Fox *et al.,* 1988), higher (Ablashi *et al.,* 1988) or lower prevalence to HHV-6 in HIV-1 positive people, or no positive correlation between seroprevalence to HHV-6 and progression to AIDS (Spira *et al.,* 1990; Chen *et al.,* 1992).

1.3.7. The HHV-6 genome

The features of the HHV-6 genome discussed below have been derived largely from Lawrence *et al.* (1990), Lindquister and Pellett (1991), Martin *et al.* (1991), Nicholas and Martin (1994), Nicholas (1994) and Gompels *et al.* (1995). Table 2 gives a list of the genes present in the HHV-6 genome, Fig. 4 shows the genome structure and relative arrangement of the genes, and Fig. 5 compares the gene arrangements of the human β-herpesviruses.

1.3.7.1. Size and structure

The published sequence of HHV-6 strain U1102 (Gompels *et al.,* 1995) is 159,322 bp in length (Nicholas, 1996) and consists of a unique region (U) (143,147 bp), flanked by single copies of a directly repeated sequence (formally DR_L and DR_R; referred to here as DR) (8,087 bp), as shown in Fig. 4.0. The overall nucleotide composition is 43% G + C, but it is lower in U (41%) and higher in DR (58%) (Gompels *et al.,* 1995; Lindquister and Pellett, 1991). Genes located in the direct repeat are prefixed by the term DR and genes in the unique sequence by U.

There are three major reiterations designated R1, R2 and R3 positioned in or adjacent to the IE-A region near the right end of U (Gompels *et al.,* 1995). R1 is located within U86, which is

TABLE 2. HHV-6 genes and encoded proteins

name	strand	start	stop	aa:mw	old name	closest homologue	gene family	gene block	- comment
LT1	C	338	3	112:12816					
DR1	-	501	791	97:11677			DR1/DR6		- CXC motif
DR2	-	791	2650	620:67151		HCMVUS26*	HCMVUS22		
DR3	C	2979	2404	192:19404					
DR4	-	2746	3045	100:10562					
DR5	C	4171	3737	145:15637					
DR6	-	4725	5033	103:12129			DR1/DR6		- CXC motif
DR7	-	5629	6717	363:40831		HCMVUS22*	HCMVUS22		- transformation, transactivator
DR8	-	7237	7566	110:12766	SJRF1				- SR domain
LJ1	C	8432	7470	321:36736	SJLF1				- across DRL junctional telomeric repeats
U1	-	8245	8613	123:14338	SJRF2				- SR domain
U2	C	9816	8719	366:39815	SHL1	HCMVUL23*	HCMVUS22		
U3	C	11276	10158	373:43700	SHL2	HCMVUL24*	HCMVUS22		
U4	C	13092	11488	535:61963	SHL3	HCMVUL27*	U4/U5		
U5	C	14548	13217	444:51587	SSL1	HCMVUL27*	U4/U5		
U6	-	14619	14864	82:8878					
U7	C	15936	14911	342:39357	SSL2	HCMVUL28*	HCMVUS22		
U8	C	17091	16024	356:41103	SFL1	HCMVUL29*	HCMVUS22		
U9	C	17552	17241	104:11819	SFL2				
U10	-	17604	18911	436:50123	SFR1,P1RF0	HCMVUL31*			
U11	C	21578	18969	870:97073	P1LF1	HCMVUL32*			- pp100 major antigenic structural protein, basic phosphoprotein, BPP
U12EX	-	21680	21710	347:39747+					- U12 exon,spliced donor/acceptor 21710,21800+
U12	-	21856	22809	318:36611	P1RF1	HCMVUL33*	GCR		- G-protein coupled receptor homology;EBV induced EBI1
U13	-	22898	23215	106:12003	EFRF1				
U14	-	23316	25142	609:69561	EFRF2	HCMVUL25*			- homology to HCMV'UL25/35' family
U15	C	25992	25663	110:13416	EFLF3				
U16EX	C	27349	26262	312:36092+		HCMVUL36EX2*			- IE-B;U16exon,spliced acceptor/donor 27034,27187
U16	C	-27116	26262	285:32908	EFLF2,IE-B	HCMVUL36*	HCMVUS22		- IE-B;transactivator,
U17	C	27349	26951	133:15538	EFLF1	HCMVUL36EX1*	HCMVUS22		- IE-B;U16exon17,acceptor/donor 27034,27187
U18	C	29389	28511	293:33219	EJLF6	HCMVUL37EX3*			- IE-B;homology to HCMV IE glycoprotein
U19	C	30818	29652	389:43317	EJLF4	HCMVUL38*			- IE-B
U20	C	32337	31072	422:48689	EJLF3		IG		- glycoprotein;Ig chain C domain
U21	C	33641	32343	433:49396	EJLF2				- glycoprotein
U22	C	34347	33742	202:23409	EJLF1				- glycoprotein
U23	C	35085	34378	236:26524	EqLF1				- glycoprotein
U24	C	35655	35395	87:10092	EoLF1				- glycoprotein exon? glyco site,TM
U25	C	36814	35867	316:37097	EPLF3	HCMVUL43*	HCMVUS22		
U26	C	37809	36925	295:33336	EPLF2				
U27	C	38978	37800	393:44811	EPLF1	HCMVUL44***		I	- pp41,pol processivity,transactivator,HCMV ICP36
U28	C	41434	39023	804:93354	P2LF2	HCMVUL45***		I	- large sub-unit ribonucleotide reductase;RR1
U29	C	42356	41460	299:34205	P2LF1	HCMVUL46***		I	- capsid assembly and DNA maturation; minor capsid protein, mCP
U30	-	41884	45129	1082:124060	P2RF1	HCMVUL47***		I	- HCMV capsid assembly,myosin
U31	-	45150	51380	2077:239950	HHRF1	HCMVUL48***		I	- large tegument protein,high molecular weight protein, HMWP
U32	C	51721	51458	88:9742					
U33	C	53135	51726	470:54701	XILF3	HCMVUL49***		I	- capsid protein
U34	C	53916	53089	276:31672	XILF2	HCMVUL50***		I	- possible virion protein,TM
U35	C	54253	53936	106:12441	XILF1	HCMVUL51* *			
U36	-	54252	55703	484:56409	XIRF1	HCMVUL52***		I	- probable virion protein
U37	-	55710	56501	264:30846	XIRF2	HCMVUL53***		I	- nuclear phosphoprotein?
U38	C	59588	56553	1012:115822	XILF0,Pol	HCMVUL54***		II	- DNA polymerase
U39	C	62080	59591	830:93240	gpB	HCMVUL55***		II	- glycoprotein B
U40	C	64214	62037	726:82879	tp	HCMVUL56***		II	- transport protein
U41	C	67620	64225	1132:127765	mdbp	HCMVUL57***		II	- major DNA binding protein
U42	C	70598	69057	514:59771		HCMVUL69***		III	- conserved herpesvirus transactivator-
U43	C	73405	70826	860:100082		HCMVUL70***		III	- helicase/primase complex; HSV primase
U44	-	73446	74084	213:24297		HCMVUL71***		III	
U45	C	75218	74091	376:43399	BHLF3	HCMVUL72***		III	- putative dUTPase
U46	-	75291	75542	84:9779	BHRF1	HCMVUL73***		III	- membrane/secreted protein
U47	C	77867	75915	651:73251	variable gp	HCMVUL74*			- membrane/secreted glycoprotein
U48	C	80118	78037	694:79544	gH,BHLF1	HCMVUL75***		IV	- glycoprotein H
U49	-	80277	81032	252:29293	BHRF2	HCMVUL76***		IV	- fusion protein
U50	-	80812	82476	555:63596	BHRF3	HCMVUL77***		IV	- virion protein
U51	-	82574	83476	301:34715	XKRF1	opioid[a]/HCMVUL78*	GCR		- G-protein coupled receptor homology; HVS GCR
U52	C	84274	83501	258:30031	XKLF1	HCMVUL79*			
U53	-	84281	85864	528:58637	0R,XKRF1	HCMVUL80***		IV	- protease; U53a in-frame assembly protein, AP
U54	C	87427	86054	458:51499	1L	HCMVUL82/83*			- tegument pp65/72K, possible transactivator IE genes
U55	C	88803	87508	432:50172	2L	HCMVUL84*			
U56	C	89873	88986	296:33465	3L	HCMVUL85***		V	- probable capsid protein
U57	C	93912	89878	1345:151953	4L,mcp	HCMVUL86***		V	- major capsid protein,MCP

name	strand	start	stop	aa:mw	old name	closest homologue	gene family	gene block	- comment
U58	-	93924	96239	772:88748	5R	HCMVUL87**			
U59	-	96239	97288	350:39884	6R	HCMVUL88*			
U60	C	98256	97291	322:36084	7L	HCMVUL89EX2***		VI	- late spliced gene(U60/66),possible DNA packaging protein
U61	C	98578	98234	115:13578					
U62	-	98427	98681	85:9579	8R	HCMVUL91*			
U63	-	98632	99279	216:24783	9R	HCMVUL92**			
U64	-	99260	100585	442:51392	10R	HCMVUL93***		VI	
U65	-	100545	101549	335:37878	11R	HCMVUL94***		VI	
U66	C	102486	101572	305:35930	12L	HCMVUL89EX1***		VI	- late spliced gene(U60/66),possible DNA packaging protein
U67	-	102458	103516	353:39533	13R	HCMVUL95***		VI	
U68	-	103519	103860	114:13075	14R	HCMVUL96**			
U69	-	103866	105551	562:63718	15R	HCMVUL97***		VI	- ganciclovir kinase; conserved phosphotransferase
U70	-	105562	107025	488:56646	16R	HCMVUL98***		VI	- alkaline exonuclease
U71	-	106965	107195	77:8469	17R				- position HCMV pp28k,HSV myristilated virion protein
U72	C	108312	107281	344:38993	18L	HCMVUL100***		VI	- integral membrane protein, gM
U73	-	108325	110664	780:89719	19R,HDRF0	HSVUL9+			- origin binding protein
U74	-	110636	112621	662:76318	HDRF1	HCMVUL102***		VI	- helicase/primase complex
U75	C	113408	112662	249:28762	HDLF2	HCMVUL103***		VI	
U76	C	115305	113320	662:77236	HDLF1	HCMVUL104***		VI	- possible virion protein
U77	-	115100	117571	824:93288	HDRF2	HCMVUL105***		VI	- helicase/primase complex; helicase
U78	C	119038	118712	109:12725	EDLF5				
U79	-	120164	121195	344:39273	EDRF1	HCMVUL112*			- HCMV in vitro replication;spliced
U80	-	-121170	121763	198:22256	EDRF2	HCMVUL113*			- HCMV in vitro replication; spliced
U81	C	122577	121813	255:29039	EDLF4	HCMVUL114***		VII	- uracil-DNA glycosylase
U82	C	123405	122656	250:28962	gL,EDLF3	HCMVUL115***		VII	- glycoprotein gL; gH accessory protein
U83	-	123528	123818	97:10411	EDRF3				- CC chemokine?
U84	C	124953	123928	342:39557	EDLF2	HCMVUL117*			- spliced in HCMV
U85	C	125853	124984	290:32901	EDLF1		IG		- OX-2 homology; glycoprotein
U86	C	-128136	125992	715:80040	BCLF1	HCMVUL122*			- IE-A;HCMVIE2 homology, SR domain repeats
U87	C	130043	127554	830:91391	BCLF0				- IE-A;glycoprotein?, highly charged,pro repeats
U88	-	131034	132272	413:44136					- IE-A;open all frames, cys repeats
U89	C	135610	133094	839:93712	pRF3/4;RF2				- IE-A;HCMV IE1 position, transactivator
U90	C	135948	135667	94:10651	pRF2;RF1				- IE-A; spliced U89
U91	-	-136485	136826	114:12897					- IE-A;spliced antisense IE1?
U92	C	-138492	138052	147:15878					- kpn repeats, part duplicate U93
U93	C	-139124	138534	197:21371					- kpn repeats, part duplicate U92
U94	C	142866	141397	490:55849	HCLF2	AAV2 Rep 68/78			- parvovirus replication, transactivation
U95	-	142941	146303	1121:124069	HCRF2	MCMVIE2*	HCMVUS22		- positional/sequence homolgue MCMVIE2
U96	C	146940	146644	99:12128	HCLF1				
U97	C	148077	147811	89:10384					
U98	C	-149391	148744	296:24698					
U99	C	149766	149488	93:10559					- signal sequence
U100	C	150437	149871	189:21609	gp82/105				- spliced glycoprotein gp82/105
RJ1	C	151571	151143	143:16221					- across DRR junctional telomeric repeats
DR1	-	151734	152024	97:11677			DR1/DR6		- CXC motif
DR2	-	152024	153883	620:67151		HCMVUS26*	HCMVUS22		
DR3	C	154212	153637	192:19404					
DR4	-	153979	154278	100:10562					
DR5	C	155404	154970	145:15637					
DR6	-	155958	156266	103:12129			DR1/DR6		- CXC motif
DR7	-	156862	157950	363:40831		HCMVUS22*	HCMVUS22		- transformation, transactivator
DR8	-	158470	158799	110:12766					- SR domain

A compilation of reading frames of HHV-6 strain U1102, from Gompels *et al.* (1995). [a] Start is the first base ATG (or its complement), except as indicated by a dash, where first base in ORF or exon is listed. Stop is third base in stop codon. + indicates spliced gene encoded molecular weight.

[B] ***conserved in α-, β- and γ-herpesviruses; **conserved in β- and γ-herpesviruses; *conserved in β-herpesviruses; **conserved in β- and α-herpesviruses.

[C] Properties for proteins are derived mostly from homologues described in HSV-1, VZV, EHV-1, or HCMV (Davison and Scott, 1986; McGeoch et al., 1988; Chee et al., 1990; Telford et al., 1992).

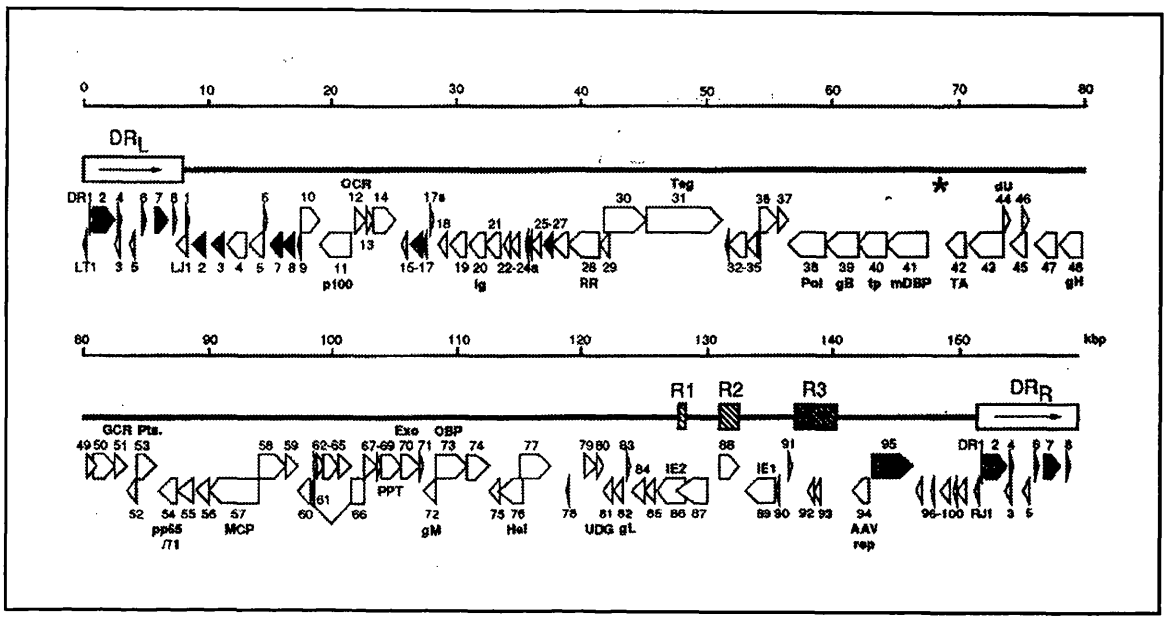


Fig. 4. Genetic layout of the HHV-6 genome.

Predicted HHV-6 gene organisation, from Gompels *et al*. (1995). Repeat regions are boxed, DRL, DRR, R1, R2, R3, and the long unique (UL) region is indicated by a solid line. Protein coding regions are indicated as open arrows and are numbered DR1-DR8 in the direct repeats and U1-U100 in UL. The ori-lyt is indicated by a star on UL. Abbreviations are, GCR (G-protein coupled receptor), RR1 (large subunit of ribonucleotide reductase), TEG (tegument protein), POL (DNA polymerase), tp (transport protein), mDBP (major single-stranded DNA binding protein), TA (conserved herpesvirus transactivator), Pts (protease/assembly protein), exo (alkaline exonuclease), OBP (origin binding protein), UDG (uracil DNA glycosylase), Hel (helicase), MCP (major capsid protein), AAV rep (Adeno-associated virus replication protein homologue), IG (immunoglobulin superfamily). The US22 gene family are shaded.

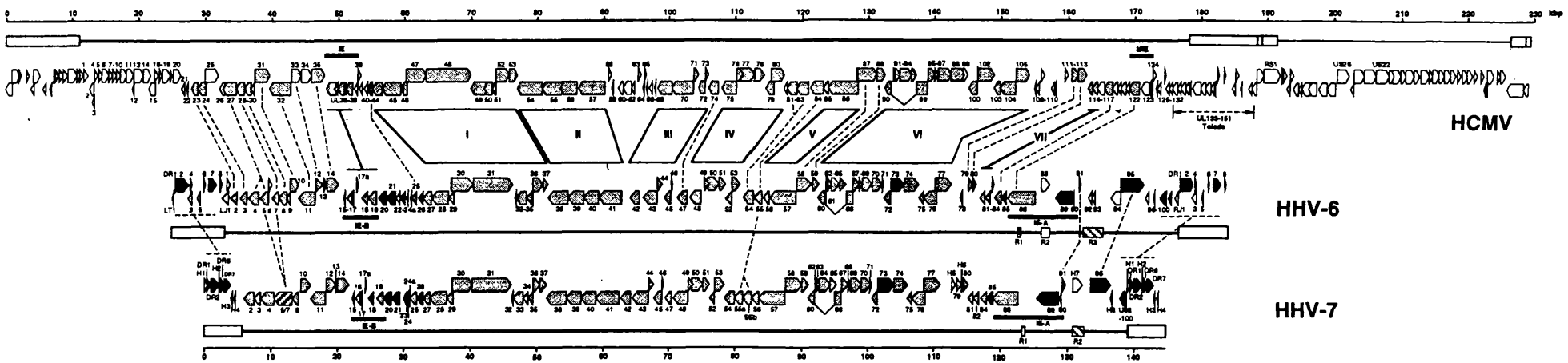


Fig. 5. Comparison of the genetic content of HCMV, HHV-6 and HHV-7.

Alignment of the genomes of HCMV (AD169), HHV-6 (U1102), and HHV-7 (JI), showing homologous (shaded) and nonconserved (open) genes in these β-herpesviruses; from Nicholas (1996). Solid areas indicate genes between HHV-6 and HHV-7 but not present in HCMV. The herpesvirus-conserved gene blocks (I to VII) are indicated. Major divergence loci between HCMV and HHV-6/HHV-7 correspond to terminal sequences, regions between HCMV UL38-UL43 and UL57-UL69, and sequences within and around the MIE locus (UL122/123). HHV-6/HHV-7 US22 genes U95, DR2, and DR2, and DR7 have approximately positional counterparts in HCMV, but these corresponding genes are not necessarily the nearest homologs. The locus at which additional genes (UL133 and UL151) have recently been identified in HCMV strain Toledo (Cha et al., 1996) is included.

homologous to HCMV IE-2 (Birney *et al.,* 1993; Nicholas, 1994; Gompels *et al.,* 1995). The R2 region contains approximately 250 bp of simple TG dinucleotide repeats and creates a large ORF in all 6 reading frames (Gompels *et al.,* 1995), although it is doubtful that any are actually expressed as protein. R3 is an array of more than 25 copies of a 110 bp tandem reiteration.

Both ends of each copy of DR contain arrays of a simple repeat $(GGGTTA)_n$, similar to human telomeric sequences (Kishi *et al.,* 1988; Martin *et al.,* 1991a; Gompels *et al.,* 1995). The telomeric repeats form an imperfect array near the left end of DR, and approximately 60 perfect tandem copies of the sequence near the right end of DR (Gompels and Macaulay, 1995; Thomson *et al.,* 1994), although the number of copies in each region may vary (Martin *et al.,* 1991a; Gompels and Macaulay, 1995; Thomson *et al.,* 1994). The significance of the telomeric repeats has not been established. Cleavage/packaging sites, pac-1 and pac-2, are positioned between the telomeric repeats and the ends of DR (Gompels *et al.,* 1995; Thomson *et al.,* 1994).

1.3.7.2. Relationships to other herpesviruses

The core genes are compactly arranged across 86 kb in the centre of U. Overall, the HHV-6 genome is arranged colinearly with that of HCMV, but diverges near the genome ends (Efstathiou *et al.,* 1992; Neipel *et al.,* 1991). As a result, genes contained at the left end and in U_S of the HCMV genome, including the HLA class 1 homologue (UL18) and most of the glycoprotein gene families (RL11, US6 and US12) are absent from the HHV-6 genome. In a reciprocal fashion those genes unique to HHV-6 are found largely outside the region of colinearity with HCMV (Gompels *et al.,* 1995).

1.3.7.3. Gene complement

The HHV-6 genome contains 121 closely packed ORFs (LT1, DR1-DR8, LJ1, U1-U100, RJ1, DR1-DR8), encoded on both DNA strands, with little overlapping (Gompels *et al.,* 1995). Splicing is thought to operate for several pairs of ORFs: U60 and U66 (Lawrence et al., 1990), U17 and U16 (Nicholas and Martin, 1994), U90 and U91 (Schiewe et al., 1994) and U96 to U100 (Pfeiffer et al., 1995).

The HHV-6 genome contains two IE loci, IE-A (U86-U91) and IE-B (U16-U19), which are roughly analogous to HCMV MIE and IE-2 respectively, and which are also CpG suppressed (Kouzarides *et al.,* 1988; Nicholas and Martin, 1994). HHV-6 genes with homologues in HCMV are listed in Table 2.

Several HHV-6 genes lack detectable counterparts in HCMV. These are also listed in Table 2 and include: the telomeric repeat ORFs LT1, LJ1 and RJ1; the R2 ORF U88; glycoprotein ORFs U20, U21, U22, U23, U24 (although there are glycoprotein genes in the corresponding region of the HCMV genome they are unrelated); the spliced ORFs U96-U100 encoding gp105; U85, which is a homologue of OX-2 and a member of the Ig family; the origin binding protein encoded by U73; the AAV-2 rep gene homologue, U94; the IE proteins encoded by U89/U90 and the U91 protein; and ORFs DR1, DR3, DR4, DR5, DR8, U1, U6, U9, U13, U15, U26, U32 ,U61, U78, U83, U92 and U93. Many of the genes unique to HHV-6 are located towards the genome ends and include regions which may be expressed by splicing.

1.3.7.4. Origin of lytic DNA replication

HHV-6 contains an origin of DNA replication (ori) upstream of the major DNA-binding protein gene (U41) in a position corresponding to ori_L in HCMV (Hamzeh *et al.,* 1990; Anders *et al.,* 1992; Gompels *et al.,* 1992; Jones and Teo, 1992; Masse *et al.,* 1992; Dewhurst *et al.,* 1993). Ori_L is complex and comprises an AT-rich sequence and two 137 bp imperfect direct repeats. It contains two origin-binding protein binding sites (OBP-1 and OBP-2), and is adjacent to a GC-rich motif. It is noteworthy that the origin binding protein of HHV-6 (U73) is a counterpart of the α-herpesvirus origin binding protein, and lacks a homologue in HCMV (Chee *et al.,* 1990). Thus, HHV-6 ori_L has features of both the β- and α-herpesviruses, in containing a complex set of repeat motifs and sequences which bind the origin binding protein.

1.3.7.5. Gene families and captured genes

The HHV-6 genome contains several gene families, DR1/6 (DR1 and DR6), U4/5, (U4 and U5) US22 (DR2, DR7, U2, U3, U7, U8, U17, U25, and U95) GCR (U12 and U51) and Ig (U20 and U85), some of which are related HCMV genes (Chee *et al.,* 1990) (see Table 2).

The DR1/DR6 proteins share a CxC motif (X = unspecified residue) and appear to be unique to HHV-6. U4/5 are tandemly duplicated positional homologues of HCMV UL27. Members of the US22 gene family are the most numerous related genes in the HHV-6 genome, having nine members (two repeated in DR). They are related to varying extents to the 12 members in HCMV (see Table 2) (Kouzarides *et al.,* 1988; Chee *et al.,* 1990); U95 is also related to the MCMV IE2 gene (Messerle *et al.,* 1991). There is evidence for a role in transcriptional activation for HHV-6 U16 (Geng *et al.,* 1992; Nicholas and Martin, 1994) and DR7 (Thompson *et al.,* 1994), DR7 also appears to have a role in cellular transformation *in vitro* (Thompson *et al.,* 1994).

The U12 and U51 proteins are cellular GCR homologues, and are distantly related to HCMV GCR gene family (Gompels *et al.,* 1995; Chee *et al.,* 1990). GCRs are similar to chemokine receptors, which interact with a family of proinflammatory cytokines that function in the migration and activation of leukocytes and thus mediate the inflammatory response (Oppenheim *et al.,* 1991; Gerard and Gerard, 1994). Interestingly, HHV-6 U83 may also act as a CC chemokine.

U20 and U85 encode glycoproteins with domains similar to those of the Ig superfamily (Gompels *et al.,* 1995), although the significance of the former is doubtful. These proteins may play a role in cell-to-cell contact via protein-protein interactions (Williams and Barclay, 1988).

Notably, HHV-6 encodes a homologue of the parvovirus, adeno-associated virus type 2 rep gene. The function of the gene product in HHV-6 is unknown, but in AAV this protein functions as a transcriptional regulatory protein and is an essential component of the DNA replication machinery (Thomson *et al.,* 1991). On the basis of close (24% identity) amino acid sequence similarities between the AAV-2 and HHV-6 rep genes, HHV-6 U94 probably represents a gene captured during coinfection of a single cell by AAV-2 and an HHV-6 progenitor (Thomson *et al.,* 1991).

1.4. HUMAN HERPESVIRUS 7

1.4.1. Basic characteristics

HHV-7 (strain RK) was first isolated in 1989 from the $CD4^+$ T-cells of a 26 year old healthy individual (Frenkel *et al.,* 1990), during experiments concerning propagation of HHV-6. Uninfected cells underwent spontaneous cytopathic effect (CPE) following conditions promoting T-cell activation, and subsequent electron microscopic analyses revealed particles with the characteristic herpesvirus morphology including a clearly visible tegument layer. However, the virus was distinct from other herpesviruses, including HHV-6 A and B variants, as shown by restriction endonuclease blot hybridisation analyses, nucleotide sequencing, and interactions with monoclonal antibodies. The novel virus was designated HHV-7 (Frenkel *et al.,* 1990).

1.4.2. Epidemiology

HHV-7 is ubiquitous in the human population - probably more than 85% are infected (Ablashi *et al.,* 1995; Wyatt *et al.,* 1991; Berneman *et al.,* 1992b). The prevalence rates both for adults and for children appear lower in Japan than in the USA or Europe (Yoshikawa *et al.,* 1993; Wyatt *et al.,* 1991).

It is likely that HHV-7 is transmitted through saliva, many workers have isolated HHV-7 from saliva and noted that the salivary system is a source of persistent production of infectious HHV-7 (Wyatt and Frenkel, 1992; Sada *et al.,* 1996; Hidaka *et al.,* 1993; Black *et al.,* 1993). HHV-7 has also been consistently isolated from the peripheral blood mononuclear cells (PBMC) of young children (Ablashi *et al.,* 1995). In addition, HHV-7 has been found at the cervix (Okuno *et al.,* 1995), suggesting possible sexual transmission of the virus.

HHV-7 infection occurs early in childhood, perhaps later than HHV-6 infection at three years of age (Wyatt *et al.,* 1991; Yoshikawa *et al.,* 1993), but one study reported little difference in the prevalence of antibodies to both viruses in relation to age (Clark *et al.,* 1993). Although the great majority of human sera contain antibodies to both HHV-6 and HHV-7, as tested by immunofluorescent antibody (Berneman *et al.,* 1992b; Wyatt *et al.,* 1991), very few sera contain antibodies to HHV-6 or HHV-7 alone.

1.4.3. Cellular tropism

HHV-7 has a more limited host range then HHV-6, but grows well in CD4+ lymphocytes (PBMC) and adequately in the immature T-cell line SUP T1 (Frenkel *et al.,* 1990; Berneman *et al.,* 1992b). The CD4 molecule is a critical part of the cell receptor used by HHV-7 (Lusso *et al.,* 1994), although it has been proposed that HHV-7 infection selectively down-modulates surface expression of cellular CD4 (Lusso *et al.,* 1994; Furukawa *et al.,* 1994). This is in direct contrast to HHV-6, for which CD4 is not the cellular receptor, and which is thought to enhance CD4 presentation during infection (Lusso *et al.,* 1991; Furukawa *et al.,* 1994).

As HIV-1 is also associated with the CD4 receptor and infects CD4+ T-cells, it is interesting that preadsorption of CD4+ T-cells with HHV-7 results in inhibition of HIV-1 replication (Lusso *et al.,* 1994).

1.4.4. Growth properties

Little can be said about the HHV-7 lytic cycle except that it is likely to be similar to that of HHV-6. A regulatory gene cascade has not been described for HHV-7, and little is known of the regulation of specific viral genes, largely as a result of the difficulty involved in growing high titre virus. As with HHV-6, tegusomes are evident in HHV-7 infected cells.

The high prevalence of HHV-7 and its presence in the saliva of healthy individuals (Wyatt and Frenkel, 1992; Wyatt *et al.,* 1991; Berneman *et al.,* 1992b; Black *et al.,* 1993; Ablashi *et al.,* 1995; Frenkel and Roffman, 1996) are unlikely to be maintained by lytic reinfection alone, but probably involves latent virus. HHV-7 was first isolated from CD4+ T-cells, most likely following reactivation from latency, and PBMCs have been proposed as a site of HHV-7 latency (Frenkel *et al.,* 1990; Katsadanas *et al.,* 1996). HHV-7 infection has been reported to mediate reactivation of HHV-6 in cells infected with both viruses (Frenkel and Wyatt, 1992; Katsadanas *et al.,* 1996).

1.4.5. Disease

Unlike HHV-6, which is widely accepted as a causative agent of ES (Yamanishi *et al.,* 1988), HHV-7 has no proven involvement in any disease. Recent reports of several seroconversions to HHV-7 during convalescence from ES have suggested that HHV-7 may cause a proportion of

ES cases (Tanaka *et al.,* 1994; Hidaka *et al.,* 1994; Ueda *et al.,* 1994; Asano *et al.,* 1995; Portolani *et al.,* 1995). However, in all but one of these cases the patients had previously seroconverted to HHV-6. Also, it has been suggested that primary HHV-7 infection may appear to cause ES as a result of its ability to reactivate HHV-6 (Asano *et al.,* 1995; Frenkel *et al.,* 1992; Frenkel and Roffman, 1996; Katsafanos *et al.,* 1996).

HHV-7 was recovered from one patient with CFS (the JI strain; Berneman *et al.,* 1992), but serological studies have found no association between HHV-7 infection and the disease. One report (Kawa Ha *et al.,* 1993) described a case of chronic EBV-like infection where no viruses other than HHV-7 could be isolated, but again no pathogenic link was established. In addition, HHV-7 is known not to be the etiologic agent of Kawasaki disease (Burns, 1994).

1.4.6. The HHV-7 genome

The features of the HHV-7 genome discussed below have been derived largely from Secchiero *et al.,* (1995), Ruvolo *et al.,* (1996) and Nicholas (1996). Table 3 gives a list of the genes represented in the HHV-7 genome, Fig. 5 compares the gene arrangements of the human β herpesviruses and Fig. 6 shows the genome structure and relative arrangement of the genes.

1.4.6.1. Size and structure

The complete nucleotide sequence of HHV-7 strain JI was derived using plasmid and lambda clones by Nicholas (1996). The following description of the genome content is derived from this analysis except where stated otherwise.

Genome structure is identical to that of HHV-6, with a single long unique component (U) flanked by large terminal direct repeats (DR_L and DR_R), approximately 133 kbp and 6 kbp in length, respectively (Nicholas, 1996; Ruvolo *et al.,* 1996; Secchiero *et al.,* 1995). The genome of HHV-7 stain JI is thus approximately 14 kbp shorter than that of HHV-6 strain U1102. The nucleotide composition of HHV-7 is 43% G+C overall, and is lower (38%) in U and higher (54%) in DR (Secchiero *et al.,* 1995).

TABLE 3. HHV-7 genes and encoded proteins

ORF[a]	Sense[b]	Position[b]				Length (aa)[c]	Homology with HHV-6 (%Sim/%Id)[d]	Homolog(s)[e]								Comments
		Start	Stop	N-Term Met?	Poly(A)			HHV-6	HCMV	HVS	EBV	EHV-2	EHV-1	VZV	HSV-1	
H1	+	33	542	N	853*	169										
DR1	+	368	826	N	853*	152	57.1/44.0	DR1								US22 gene family, DR1/DR6 homology
DR2	+	898	2100	N	3917*	400	58.1/37.0	DR2								US22 gene family
H2	+	2267	2506	Y	3917*	79										
DR6	+	2562	3050	Y	3917*	161	71.6/58.8	DR6								US22 gene family, DR1/DR6 homology
DR7	+	3122	3910	Y	3917*	262	77.8/60.5	DR7								US22 gene family, transactivator
H3	−	4224	3976	Y	3040	82										
H4	−	4745	4449	Y	3040	98										
U2	−	7417	6338	Y	6248*	359	69.6/51.1	U2	UL23							US22 gene family
U3	−	8732	7578	Y	7514	384	65.4/49.3	U3	UL24							US22 gene family
U4	−	10382	8754	Y	7514	542	75.5/59.2	U4	UL27							
U5/7	−	13004	10407	Y	10411+	865	71.2/49.7	U5	UL27							US22 gene family (HHV-6 U7)
							80.2/60.7	U7	UL28							
U8	−	14262	13174	Y	11850	362	69.0/51.0	U8	UL29							US22 gene family
U10	+	14608	15963	Y	15885*	451	71.6/51.5	U10	UL31							
U11	−	18249	15982	Y	15971	755	55.6/30.7	U11	UL32							Structural phosphoprotein
U12	+	18396	19436	Y	19481	346	67.7/45.7	U12	UL33							GCR homolog, chemokine receptor
U13	+	19521	19817	Y	20424*	98	54.6/37.1	U13								
U14	+	19885	21831	Y	21925*	648	67.9/53.5	U14	UL35							HCMV UL25/35 gene family
U15	−	22244	22564	Y	22075	106	84.9/64.1	U15								
U16	−	23547(A)	22772	N	22754	264	74.8/56.5	U16	UL36x2							IE-B, transactivator (spliced to U17)
U17Ex	−	23836	23620(D)	Y		72	55.5/41.7	U17Ex	UL36x1							IE-B, transactivator (spliced to U16)
U17	−	23836	23570	Y	22754	88	54.5/39.8	U17								
U17a	+	24318	24587	Y	24707	89	46.6/29.3	U17a								
U18	−	25600	24713	Y	24578	295	71.4/43.9	U18	UL37x3							IE-B, homologous to HCMV IE glycoprotein
U19	−	26922	25945	Y	25504	325	59.9/40.1	U19	UL38							IE-B
U20	−	28211	27036	Y	27027	391	53.2/23.4	U20								Ig gene family?
U21	−	29494	28202	Y	28127	430	55.8/32.2	U21								Glycoprotein
U23	−	30418	29903	Y	29904+	171	55.6/29.4	U23								Glycoprotein, EHV-1 gJ homology
U24	−	30772	30524	Y	30455*	82	63.2/36.8	U24								Glycoprotein
U24a	−	31129	30776	Y	30671	117	57.1/33.9	U24a								
U25	−	31898	30936	Y	30671	320	70.7/48.9	U25	UL43							US22 gene family, transactivator
U26	−	32869	31988	Y	31959	293	60.4/30.4	U26								
U27	−	33951	32857	Y	31959	364	83.8/67.6	U27	UL44	(59)	(BMRF1)	(59)	(18)	(16)	(UL42)	DNA polymerase processivity factor
U28	−	36484	34064	Y	34059	806	70.2/47.1	U28	UL45	61	BORF2	61	21	19	UL39	Ribonucleotide reductase (large subunit)
U29	−	37347	36487	Y	36457	286	77.5/53.3	U29	UL46	62	BORF1	62	22	20	UL38	Minor capsid protein (mCP)
U30	+	37362	40178	Y	37548	938	66.0/45.9	U30	UL47	63	BOLF1	63	(23)	(21)	(UL37)	Capsid assembly, myosin
U31	+	40179	46358	Y	46994	2,059	65.3/46.2	U31	UL48	64	BPLF1	64	24	22	UL36	Large tegument protein
U32	−	46627	46355	Y	46385*	90	80.7/65.9	U32		(66)	(BFRF2)	(66)	(25)	(23)	(UL35)	
U33	−	48041	46608	Y	46385*	477	78.5/59.4	U33	UL49							Virion protein
U34	−	48768	47992	Y	47906	258	75.0/59.1	U34	UL50	67	BFRF1	67	(26)	(24)	(UL34)	Virion protein?
U35	−	49119	48805	Y	48775	104	78.6/58.3	U35	UL51				27	25	UL33	
U36	+	49118	50575	Y	51089*	485	77.3/58.2	U36	UL52	68	BFLF1	68	28	26	UL32	Probable virion protein
U37	+	50577	51356	Y	51613*	259	80.5/63.0	U37	UL53	69	BFLF2	69	29	27	UL31	

U38	–	54401	51363	Y	51353	1,012	81.6/66.7	U38	UL54	9	BALF5	9	30	28	UL30	DNA polymerase
U39	–	56869	54401	Y	53919*	822	72.1/56.4	U39	UL55	8	BALF4	8	33	31	UL27	Glycoprotein B (gB)
U40	–	58997	56832	Y	56686	721	75.4/56.0	U40	UL56	7	BALF3	7	32	30	UL28	Transport protein (tp)
U41	–	62395	59000	Y	58773*	1,131	84.1/63.3	U41	UL57	6	BALF2	6	31	29	UL29	Major DNA-binding protein
U42	–	64352	62772	Y	62714*	526	75.7/57.0	U42	UL69	57	BMLF1	57	5	4	UL54	Transactivator
U43	–	67086	64501	Y	64470	861	76.9/61.2	U43	UL70	56	BSLF1	56	7	6	UL52	Primase
U44	+	67143	67754	Y	67750	203	77.0/58.5	U44	UL71	55	BSRF1	55	(8)	(7)	(UL51)	
U45	–	68898	67759	Y	67763+	379	69.3/51.2	U45	UL72	54	BLLF2	54	9	8	UL50	dUTPase
U46	+	68930	69190	Y	69186+	86	72.8/53.1	U46	UL73	53	BLRF1	53	(10)	(9A)	(UL49a)	
U47	–	70579	69638	Y	69235*	313	46.6/25.4	U47	UL74							
U48	–	72889	70817	Y	70714	690	63.7/39.0	U48	UL75	22	BXLF2	22	39	37	UL22	Glycoprotein H
U49	+	73003	73722	Y	73765*	239	68.6/51.9	U49	UL76	20	BXRF1	20	37	35	UL24	Fusion protein
U50	+	73538	75202	Y	76287	554	73.4/54.9	U50	UL77	19	BVRF1	19	36	34	UL25	Virion protein
U51	+	75304	76188	Y	76287	294	64.6/36.1	U51	UL78	74						GCR, opioid^{R} homolog
U52	–	76949	76185	Y	75932	254	79.9/55.5	U52	UL79	18	BVRF1.5a/b	18				
U53	+	76957	78495	Y	79010	512	67.1/52.4	U53	UL80	17	BVRF2	17	35	33	UL26	Protease/assembly protein
U54	–	79870	78503	Y	78443	455	66.3/41.4	U54	UL82/83							Tegument protein transactivator
U55A	–	81201	79918	Y	79891	427	56.8/32.6	U55	UL84							Replication function?
U55B	–	82577	81285	Y	79891	430	46.4/20.7	U55	UL84							Replication function?
U56	–	83511	82630	Y	82431	293	82.6/65.2	U56	UL85	26	BDLF1	26	43	41	UL18	Capsid protein
U57	–	87551	83514	Y	83098	1,345	82.6/68.4	U57	UL86	25	BcLF1	25	42	40	UL19	Major capsid protein (MCP)
U58	+	87563	89890	Y	90111*	775	75.7/61.3	U58	UL87	24	BcRF1	24				
U59	+	89838	90881	Y	90942*	347	58.9/38.7	U59	UL88							
U60	–	92005(A)	90878	N	90793	394	86.9/75.7	U60	UL89x2	29b	BDRF1	29b	44	42	UL15x2	Late spliced gene (U60/U66) DNA packaging
U62	+	92017	92244	Y	92564	75	81.7/50.7	U62	UL91	30	BDLF3.5	30				
U63	+	92216	92851	Y	93008	211	84.5/71.8	U63	UL92	31	BDLF4	31				
U64	+	92829	94148	Y	94363*	439	64.5/40.7	U64	UL93	32	BGLF1	32	45	43	UL17	
U65	+	94111	95103	Y	95318	330	77.1/60.1	U65	UL94	33	BGLF2	33	46	44	UL16	
U66	–	95985	95122(D)			309	76.2/62.3	U66	UL89x1	29a	BGRF1	29a	47	45	UL15x1	Late spliced (U60/U66) DNA packaging
U67	+	95984	97024	Y	97041	346	70.3/51.7	U67	UL95	34	BGLF3	34	(48)	(46)	(UL14)	
U68	+	97024	97868	Y	97971*	114	71.1/48.9	U68	UL96	35	BGLF3.5					
U69	+	97371	99011	Y	99096*	546	71.8/54.0	U69	UL97	36	BGLF4	36	49	47	UL13	Phosphotransferase
U70	+	99013	100455	Y	100554	480	72.5/52.0	U70	UL98	37	BGLF5	37	50	48	UL12	Alkaline exonuclease
U71	+	100392	100613	Y	100609+	73	64.4/53.4	U71								
U72	–	101676	100636	Y	100640+	346	81.1/59.0	U72	UL100	39	BBRF3	39	52	50	UL10	Integral membrane protein (gM)
U73	+	101693	104456	Y	104067	787	75.4/58.2	U73						51	UL9	Origin-binding protein (OBP)
U74	+	104007	105986	Y	106055	659	64.4/40.6	U74	UL102	(41)	(BBLF3)		(54)	(52)	(UL8)	Helicase/primase complex
U75	–	106743	105973	Y	105977+	256	64.9/45.6	U75	UL103	42	BBRF2	42	55	53	UL7	
U76	–	108589	106667	Y	106633*	640	78.5/59.7	U76	UL104	43	BBRF1	43	56	54	UL6	Virion protein?
U77	+	108435	110897	Y	111009	820	85.7/74.9	U77	UL105	44	BBLF4	44	57	55	UL5	Helicase
H5	+	112811	113311	N	114614	166										
U79	+	113502	114203	Y	114614	233	70.6/44.7	U79	UL112							HCMV replication, spliced (UL112/UL113)
H6	+	114257	114505	N	114614	82	64.1/48.7	U79(C)								HHV-6 U79 homology (C terminus)
U80	+	114557	115189	N	115347*	210	57.0/40.4	U80	UL113							HCMV replication, spliced (UL112/113)
U81	–	115948	115184	Y	115161	254	79.1/58.5	U81	UL114	46	BKRF3		61	59	UL2	Uracil-DNA glycosylase
U82	–	116778	116038	Y	115986	246	62.1/40.0	U82	UL115	(47)	(BKRF2)		(62)	(60)	(UL1)	Glycoprotein L
U84	–	118043	117111	Y	116463*	310	65.7/46.5	U84	UL117							Spliced in HCMV
U85	–	118913	118071	Y	117885	280	60.8/37.7	U85								OX-2 homology, glycoprotein
U86	–	122708	119091	Y	119082	1205	57.0/41.4	U86	UL122							IE-A, HCMV IE2 homology
U89	–	128668	125420	Y	125424+	1,082	60.7/36.4	U89								IE-A, transactivator

U90	−	129051	128776	N		91	69.6/39.1	U90	IE-A, exon in HHV-6
U91	+	129122	129625	Y	130111	167	47.7/27.0	U91	
H7	+	130829	132112	Y	132157	427			DraI repeats
U95	+	133382	136204	Y	136293	940	52.8/29.7	U95	MCMV IE2 homolog, US22 gene
H8	−	136579	136307	Y	136257	90			
U98	−	138451	137945	N		168	52.2/31.2	U98	Homology to HHV-6 gp82/105
U99	−	138692	138375	N		105	61.3/37.6	U99	Homology to HHV-6 gp82/105
U100	−	138999	138751	Y		82	44.4/23.5	U100	Homology to HHV-6 gp82/105
H1′	+	139080	139589	N	139900*	169			
DR1′	+	139415	139873	N	139900*	152	57.1/44.0	DR1	US22 gene family, DR1/DR6 homology
DR2′	+	139945	141147	N	142964*	400	58.1/37.0	DR2	US22 gene family
H2′	+	141314	141553	Y	142964*	79			
DR6′	+	141609	142097	Y	142964*	161	71.6/58.8	DR6	US22 gene family, DR1/DR6 homology
DR7′	+	142169	142957	Y	142964*	262	77.8/60.5	DR7	US22 gene family, transactivator
H3′	−	143271	143023	Y	142087	82			
H4′	−	143792	143496	Y	142087	98			

A compilation of reading frames of HHV-7 strain JI, from Nicholas (1996). [a] HHV-7 ORFs are named after their HHV-6 homologs (Gompels, et al., 1995); ORFs unique to HHV-7 are prefixed with "H" and numbered 1 to 8.

[b] The positions and orientations of the ORFs are indicated, together with the positions of the first downstream polyadenlyation signals (AATAAA, ATTAAA*). Polyadenlyation signals overlapping ORF C-terminal sequences are indicated (+). Determined (U16/U17Ex) and predicted (U60/U66) splice donor (D) and acceptor (A) sites are indicated.

[c] The sizes of ORF translation products (starting at the N-terminal methionines where these occur) in amino acids (aa) are shown.

[d] The values for percent similarity (% Sim) and identity (% Id) between HHV-7 and HHV-6 homologs are based on BESTFIT alignments (Devereux *et al.*, 1984), with gap and length weights set at 3.0 and 0.1, respectively.

[e] Homologous genes were identified by database searches and pairwise alignments. Listings of homologous genes were based on these analyses and on data from comparisons of other herpesvirus proteins (Albrecht, *et al.*, 1992; Chee *et al.*, 1990; Gompels *et al.*, 1995; McGeoch, 1989). Genes that show only limited sequence similarities to HHV-7/HHV-6/HCMV genes but are colinear (and in some cases functionally analogous) are given in parentheses.

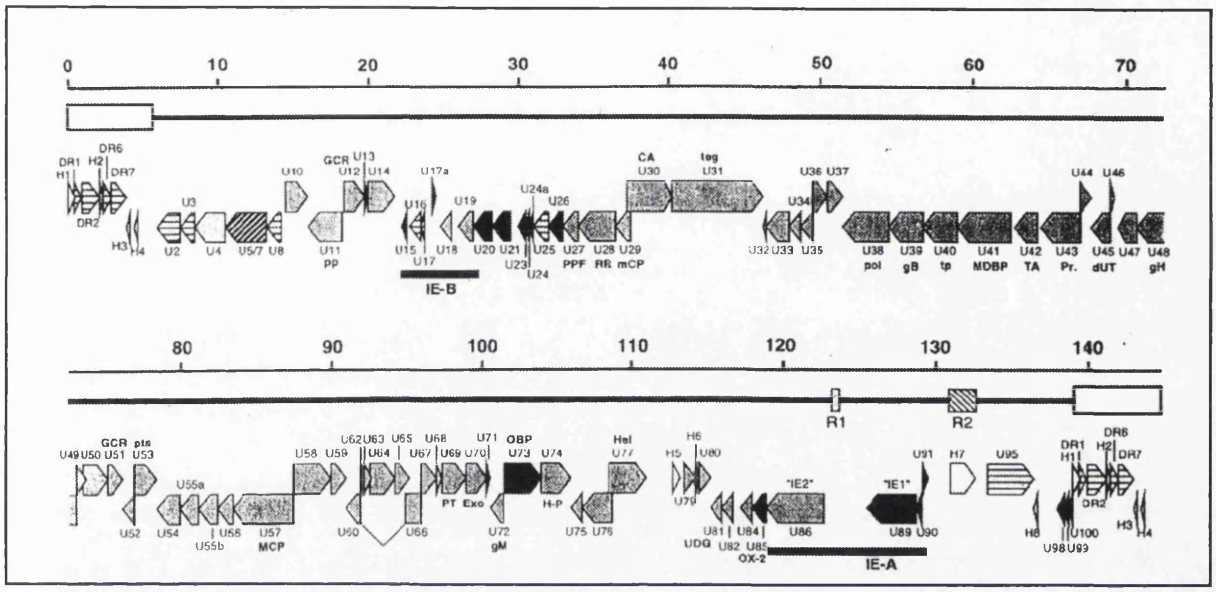


Fig. 6. Genetic layout of the HHV-7 genome.

Predicted HHV-7 gene organisation, from Nicholas (1996). Repeat regions are boxed and the long unique (UL) region is indicated by a solid line. Protein coding regions are indicated as open arrows. Several are homologous to HHV-6 genes (solid areas) and HHV-6/HCMV genes (shaded areas). Genes apparently unique to HHV-7 are unshaded. Lightly shaded ORFs (horizontal lines) correspond to members of the US22 gene family (DR1, DR2, DR6, DR7 and U95 are not conserved in HCMV).

HHV-7 ORFs with homologs in HHV-6 are named after their HHV-6 conterparts; HHV-7 unique ORFs are prefixed with "H" and sucessively numbered (from left to right on the genome). ORF U5/7 encodes a translation product with N-terminal homology to HHV-6 U7 (and HCMV U28) and C-terminal homology to HHV-6 U5 (and HCMV U27). Abbreviations are: pp, pp100 structral phosphoprotein; GCR, G-protein-coupled receptor; PPF, DNA polymerase processivity factor; RR, ribonucleotide reductase; mCP, minor capsid protein; CA, capsid assembly protein; teg, large tegument protein; pol, DNA polymerase; gB, glycoprotein B; tp, transport protein; MDBP, major DNA-binding protein; TA, herpesvirus conserved transactivator; Pr., primase; dUT, dUTPase; gH, glycoprotein H; pts, protease/assembly protein; MCP, major capsid protein; PT, phosphotransferase; Exo, alkaline exonuclease; gM, glycoprotein M; OBP, origin binding protein; H-P, helciase-primase complex component; Hel, helicase; UDG, uracil-DNA glycosylase; OX-2, OX-2 membrane antigen.

In common with HHV-6, HHV-7 has repetitive sequences upstream of the IE-A locus, designated R1 and R2. HHV-7 R1 is located between U86 and U89 in an analogous position similar to the R2 repeat in HHV-6. HHV-6 R2 is an array of TG repeats but HHV-7 R1 is complex, described as comprising two complete 84 bp repeats and two 67 bp partial repeats, with multiple copies of a simple repeat (TAAAT) scattered within and around the larger repeats. HHV-7 R2 is homologous to HHV-6 R3 (Gompels *et al.,* 1995; Martin *et al.,* 1991a, b) and is located upstream from the IE-A locus. It is described as comprising 16 well conserved 105 bp repetitive motifs, flanked by two copies of a partially conserved repeat. Size heterogeneity in this region of the HHV-7 genome has been reported, and may correspond to different numbers of 105 bp repeats in different viral isolates (Ruvolo *et al.,* 1996).

HHV-7 does not contain an equivalent of the HHV-6 R1 repeat present in U86 (Nicholas, 1994), although the HHV-7 U86 gene product does contain a region which is similarly rich in serine and basic residues at a corresponding location.

As in HHV-6, the HHV-7 genome contains arrays of human telomere-like tandem repeats ($[GGGTTA]_n$ and related sequences) close to both ends of DR (Secchiero *et al.,* 1995; Ruvolo *et al.,* 1996). The HHV-7 telomeric repeat region at the left end of DR is longer and more complex than that in HHV-6, and that at the right end is shorter and simpler. In addition, the HHV-7 telomeric repeat regions at the right end of DR varies in length between 1.2 and 1.7 kbp in plasmid clones (Nicholas, 1996).

The function of the telomeric repeats is not known, but their presence in the genomes of several lymphotropic herpesviruses, including α-herpesviruses (MDV) and γ-herpesviruses (EHV-2) (Kishi *et al.,* 1991, 1988; Telford *et al.,* 1995), has provoked a range of suggestions. Possible functions for the telomeric arrays are mentioned in the discussion.

The origin of lytic replication (ori_L) is positioned upstream of the major DNA-binding protein gene (U41), as in HHV-6, and is closely related to, but considerably shorter than, its HHV-6 counterpart (van Loon *et al.,* 1997). HHV-7 also encodes a homologue of the α-herpesvirus origin-binding protein (U73).

1.4.6.2. Genetic organisation

The HHV-7 genome contains 101 closely packed ORFs encoded on both DNA strands, with little overlapping (Fig. 6). Splicing is thought to occur in several genes, but splice sites have been predicted only for two pairs of ORFs (U60 and U66; U17 and U16).

1.4.6.3. Gene complement

HHV-7 gene functions have been inferred from previous experimental findings with other better characterised herpesviruses (notably HSV-1, HCMV and HHV-6) or from the functional characteristics of cellular homologues (see Table 3). The arrangement of the genes within the HHV-7 genome is shown in Fig. 6, and the gene arrangements of HCMV, HHV-6 and HHV-7 are compared in Fig. 5. HHV-6 and HHV-7 are very similar in gene layout, and predicted protein sequences are generally well conserved (as listed in Table 3).

Eight HHV-7 ORFs (H1-H8) lack detectable counterparts in HHV-6, and 22 HHV-6 ORFs lack HHV-7 counterparts (LT1, DR3, DR4, DR5, DR8, LJ1, U1, U6, U9, U12EX, U22, U61, U78, U83, U87, U88,U92, U93, U94, U96, U97,RJ1). Thus, HHV-7 lacks an equivalent of the AAV-2 *rep* gene. Gene fusion and duplication events also appear to have occurred in HHV-7. Two HHV-6 ORFs (U5 and U7) are fused in HHV-7 to give U5/7. Also, U55 is duplicated in HHV-7, giving U55A and U55B.

1.5. AIMS OF THE THESIS

The project was initiated in October 1993, as a collaboration with Professor N. Frenkel (Tel-Aviv University, Israel) who provided genomic DNA from HHV-7 strain RK. The aims were to determine the complete DNA sequence of HHV-7 by random shotgun cloning of genomic DNA and to interpret the coding potential of the sequence. During the course of this work the complete DNA sequences of HHV-6 strain U1102 and HHV-7 strain JI were published (Gompels *et al.,* 1995; Nicholas, 1996). Therefore, the HHV-7 strain RK sequence was compared with these two sequences in order to approve evolutionary divergence and to re-evaluate the genetic content of HHV-6 and HHV-7.

Chapter 2

Materials and Methods

2.1. MATERIALS

2.1.1. Chemicals

The chemicals used were of analytical grade and most of these were supplied by BDH Chemicals or Sigma Chemical Co. Exceptions were ammonium persulphate (APS) and N,N,N',N'-tetramethylethylenediamine (TEMED) (Bio-Rad Laboratories Ltd), dGTP, dATP, dTTP, dCTP, ATP (each at 100 mM), deaza-dGTP, ddGTP, ddATP, ddTTP, ddCTP (each at 5 mM) (Pharmacia) and N,N-dimethyl formamide (Cruachem Chemicals).

2.1.2. Solutions and buffers

TE:	10 mM Tris-HCl pH 8 1 mM EDTA
10 x TBE:	109 g/l Tris 55 g/l boric acid 9.3 g/l EDTA
50 x TAE buffer:	242 g/l Tris 57.1 g/l glacial acetic acid 11.6 g/l EDTA
DF dyes:	37.2 g/l EDTA 100 g/l Ficoll 400 5 x TBE 1% (w/v) bromophenol blue
Buffer A:	330 mM Tris-acetate (pH 7.9) 100 mM magnesium acetate 660 mM potassium acetate 5 mM dithiothreitol (DTT) (supplied by Boehringer Mannheim as buffer A)

5 x ligase buffer:	250 mM Tris-HCl (pH 7.5) 50 mM $MgCl_2$ 5 mM DTT 25% (w/v) PEG 6000
X-gal:	40 mg/ml 5-bromo-4-chloro-3-indoyl β-D-galactopyranoside in dimethyl formamide
IPTG:	30 mg/ml isopropylthio-β-D-galactoside
PEG/NaCl:	20% (w/v) PEG 6000 2.5 M NaCl
Phenol/TE:	phenol equilibrated with TE
Sodium acetate/ethanol:	10 ml 3 M sodium acetate (pH 5.5) 240 ml ethanol
TM buffer:	100 mM Tris-HCl (pH 8.0) 100 mM $MgCl_2$
Chase solution:	0.25 mM each of dGTP, dATP, dTTP and dCTP

dNTP/ddNTP solutions: volumes are in µl

	G	A	T	C
water	2000	1000	2000	2000
0.5 mM 7-deaza-dGTP	50	1000	1000	1000
0.5 mM dTTP	1000	1000	50	1000
0.5 mM dCTP	1000	1000	1000	50
5 mM ddGTP	15			
5 mM ddATP		2		
5 mM ddTTP			50	
5 mM ddCTP				7

40% Acrylamide solution: 40% (w/v) Acrylogel 5 premix; deionised

Top gel mixture:
460 g/l urea
0.5 x TBE
acrylamide solution to 6% (v/v)

Bottom gel mixture:
460 g/l urea
50 g/l sucrose
50 mg/l bromophenol blue
2.5 x TBE
acrylamide solution to 6% (v/v)

Formamide dyes:
1 g/l xylene cyanol FF
1 g/l bromophenol blue
10 mM EDTA
in deionised formamide

50 x TAE buffer/formamide solution:
960 µl deionised formamide
40 µl 50 x TAE

2.1.3. Enzymes

All restriction endonucleases and buffers were obtained from Bethesda Research Laboratories (BRL) or Boehringer Mannheim. Other enzymes included:

Calf intestinal phosphotase (1 unit/μl):	Boehringer Mannheim
T4 DNA polymerase (3 U/μl):	Boehringer Mannheim
T4 DNA ligase (4 U/μl):	Stratagene
Klenow fragment of *E. coli* DNA polymerase I (20-50 U/μl)	provided by Dr. E.A.R. Telford (Joyce and Grindley, 1983)

2.1.4. Radiochemicals

Deoxyadenosine (^{35}S) thiophosphate (dATP^{35}S) (NEG0345; Du Pont)	Specific activity: 500 Ci/mmol (12.5 μCi/μl)

2.1.5. Bacterial growth media

2YT broth:	85 mM NaCl 1% (w/v) bactopeptone 1% (w/v) yeast extract
L-Broth:	177 mM NaCl 1% (w/v) bactopeptone
L-Broth Agar:	1.5% (w/v) agar in L-broth
Top agar:	0.6% (w/v) bacto-agar in water

2.1.6. Bacterial strains

Escherichia coli MAX Efficiency DH5αF'IQ competent and lawn cells (phenotype: F'ϕ80d*lac*ZΔM15 Δ(*lac*ZYA-argF)U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_K^-, m_K^+) *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1/F' *pro*AB$^+$ *lacI*^{q}ZΔ M15 *zzf*::Tn5[Km^1]) were used (Gibco-BRL).

2.1.7. DNA

HHV-7 (strains RK) genomic DNA was provided by Prof. N. Frenkel (Tel-Aviv University). Bacteriophage M13mp19 RF I DNA (0.01μg/μl) and DNA markers (123 bp DNA ladder (1 μg/μl) and 100 bp DNA ladder (1 μg/μl)) were obtained from Gibco-BRL. Oligonucleotide primers were synthesised by Dr J McLauchlan, Mr J McGeehan and Mr R Reid using a Cruachem PS250 DNA synthesiser.

2.1.8. Miscellaneous

The following items were used in addition to standard laboratory materials and equipment:

Whatman 3 MM chromatography paper:	Whatman International Ltd
Kodak X-OMAT XS-1 film:	Kodak Ltd
BioMax MR autoradiographic film:	Kodak Ltd
Electroporator II and electroporation cuvettes 0.1 cm:	Invitrogen
EDP-Plus and EDP-Plus M-8 pipettes:	Rainin Instrument Co. Inc.
Geneclean II Kit:	BIO 101, Inc.
S/P Multi-Tube Vortexer:	Baxter
ECPS 3000/150 Electrophoresis constant power supply:	Pharmacia

2.2. METHODS

2.2.1. PREPARATION OF DNA

2.2.1.1. Production of viral DNA

The production of viral DNA was carried out in the laboratory of Professor N Frenkel. Four samples of HHV-7 RK genomic DNA, totalling several micrograms, were provided.

Human cord blood mononuclear cells (CBMCs) were treated with phytohemagglutinin for 2 days in RPMI 1640 medium containing 10% (v/v) fetal calf serum, 50 mg/ml gentamycin (Frenkel *et al.,* 1990; Berneman *et al.,* 1992a; Black and Pellett, 1993). The cells were then incubated for 2 h at 37° with concentrated aliquots of HHV-7 (RK). Following adsorption, the cells were diluted with medium to 10^6 cells/ml (approximately 300 ml total), and incubated further at 37° (Frenkel and Rapaport, 1995). The typical cytopathic effect (CPE) of HHV-7 infection, characterised by ballooning of cells and the appearance of limited syncytia, became evident at about 14 days after infection. The cells were harvested by centrifugation, rinsed in phosphate-buffered saline and pelleted by once more. The cells were resuspended in 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.6% (v/v) Nonidet P40, Dounce homogenised and separated into nuclear and cytoplasmic fractions by centrifugation. The fractions were treated with 0.5% (w/v) sodium deoxycholate and incubated with 50 mg/ml DNase I and 10 mg/ml RNase A, and capsids were prepared by density centrifugation on sucrose gradients as described by Gibson and Roizman (1972) and Vlazny *et al.* (1982). DNA was extracted from capsids by SDS-proteinase K lysis as described by Di Luca *et al.* (1990).

2.2.1.2. Restriction enzyme digestion

Digestion of DNA using restriction enzymes was carried out in accordance with the manufacturer's instructions. A typical digest would involve 1 U of enzyme per 0.5 μg DNA in a volume of 20μl, incubated for 1 h at 37°.

2.2.1.3. Agarose gel electrophoresis

100 ml of 0.6-1.5% (w/v) agarose in 1 x TBE (or TAE) was dissolved by boiling and allowed to cool to 65°. The agarose was poured into a horizontal slab gel template (15 cm x 10 cm) that had been sealed with tape and contained a Teflon well-forming comb. When the agarose was set the tape was removed, and the gel was placed into an electrophoresis tank and filled with 1 x TBE buffer. DNA samples (along with appropriate size markers) were mixed with 0.2 volumes of DF dyes prior to loading. Following electrophoresis at 100 V for 1-2 h, the DNA was stained for 30 min with 0.5 μg/ml ethidium bromide (EtBr) in 100 ml 1 x TBE. The DNA was visualised by short wave UV transillumination (302 nm) for analytical gels or by long wave UV transillumination (365 nm) for preparative gels. The gels were photographed using a photo-imager system (Appligene).

2.2.1.4. Precipitation of viral DNA

50% PEG was mixed with HHV-7 genomic DNA containing 1 M NaCl to a final concentration of 8% and incubated overnight at 4°. The DNA was pelleted by centrifugation for 10 min at 12,000 rpm, washed twice with 95% ethanol, dried in a lyophilizer, and resuspended in 50 μl of water. 5 μl was electrophoresed on a 1.5% agarose gel (containing 1 x TBE) for 2 h at 50 V. The DNA was stained with EtBr and photographed.

2.2.1.5. Sonication of viral DNA

50 μl of HHV-7 genomic DNA prepared by PEG precipitation was placed in a 1.5 ml microfuge tube and sonicated in a Pulsatron sonicating waterbath for 60 sec. 5 μl was electrophoresed at 50 V for 2 h on a 1.5% agarose gel (containing 1 x TAE). The gel was stained with EtBr and photographed.

2.2.1.6. Size selection of sonicated DNA fragments

50 μl of sonicated DNA fragments was mixed with 10 μl of DF dyes and loaded on a 1.5% (w/v) agarose gel (containing 1 x TAE). A 100 bp ladder was included as a

DNA marker in a separate lane. The samples were electrophoresed at 50 V for 2 h. The agarose gel was divided in two halves, and the half containing the DNA ladder was stained with EtBr and photographed. Reassembly of the gel over a UV light source made it possible to excise blocks of agarose containing DNA fragments of a specific size without the need to contaminate the DNA with EtBr stain. Thus, 400-700 bp fragments of DNA were separated from the bulk of the randomly sonicated DNA.

2.2.1.7. DNA purification using Geneclean

Sonicated DNA was recovered from the agarose blocks using a commercial kit (Geneclean II). The blocks were mixed with three volumes (1 g/ml) of sodium iodide solution, and incubated at 55° until the agarose had dissolved. 6 µl of 'Glassmilk' silica matrix (which binds DNA in the presence of high concentrations of sodium iodide) was added to each sample and mixed by rotation for 30 min. The silica was pelleted by brief centrifugation and the supernatant was removed. The silica was resuspended by vortexing in 200µl of wash solution and pelleted. This was repeated three times and the silica allowed to air dry for 12 min. The pellet was resuspended in 25 µl of TE and the DNA was eluted by incubation at 55° for 5 min. The silica was pelleted by brief centrifugation and the DNA solution was transferred to a fresh microfuge tube and stored at -20°.

2.2.2. PREPARATION OF TEMPLATES

2.2.2.1. T4 DNA polymerase end repair

45 µl of sonicated DNA fragments was mixed with 5.0 µl of 10 x Buffer A, 5.0 µl of 2 mM dNTPs and 3 µl (3U) of T4 DNA polymerase, and incubated for 1 h at 37°. The end-repaired DNA was purified by phenol/chloroform extraction and ethanol precipitation. The purified DNA was dissolved in 20 µl of TE.

2.2.2.2. Preparation of M13mp19 *Sma*I

20 μg of M13 mp19 RF I DNA (1 μg/μl) was added to 191 μl of water and mixed with 10 μl (100 U) of *Sma*I, 4 μl (4 U) of calf intestinal phosphatase and 25 μl of 10 x Buffer A, and incubated for 5 h at 25°. The linearized, dephosphorylated M13 DNA was purified by phenol/chloroform extraction and ethanol precipitation and resuspended in water at 0.1 μg/μl.

2.2.2.3. Ligation

2 μl of sonicated, end-repaired HHV-7 DNA fragments was added to 9.5 μl of water and mixed with 2 μl (0.2 μg) of *Sma*I-cleaved, dephosphorylated M13 mp19 RF DNA, 4 μl of 5 x ligase buffer, 2 μl of 10 mM ATP and 0.5 μl (2 U) of T4 DNA ligase and incubated overnight at 16°. Two 20 μl control ligations were also employed, both lacking HHV-7 DNA fragments and one lacking ligase. The ligations were diluted to a final volume of 100 μl with water and the DNA was purified by careful phenol/chloroform extraction and ethanol precipitation. The ligated DNA was resuspended in 50 μl of water and stored at -20°.

2.2.2.4. Transfection and electroporation

300 ml of L-broth agar was melted in a microwave oven, cooled to approximately 50° and poured into ten 90 mm Petri dishes. The plates were allowed to set and were then dried open and inverted for 1 h at 37°.

100 μl of competent *E. coli* cells (DH5αF') was added to a precooled 15 ml Falcon 2006 tube, followed by 5 μl of ligated DNA. The cells were mixed gently with the DNA and incubated on ice for 30 min. The cells were transformed by heat shock at 43° for 2 min and placed on ice for 2 min. The transformed cells were then transferred to 15 ml of top agar containing 50 μl of IPTG, 100 μl of X-gal and 1 ml of a fresh overnight culture of *E. coli* DH5αF' (grown in 2YT broth). The top agar

was spread onto the L-broth agar plates (3 ml/plate) and allowed to set for 30 min. The plates were inverted and incubated overnight at 37° and stored at 4° for up to 20 days. A typical transfection would yield a mixture of blue plaques (phage without insert) and clear plaques (phage containing insert DNA).

Electroporation was employed as an alternative method of transfecting bacterial cells. *E. coli* (DH5αF') cells were grown to mid-log phase in 400 ml of L-broth, chilled, centrifuged at 3,000 rpm in a Sorvall RC-5B superspeed centrifuge for 30 min, and then washed extensively with water to reduce the ionic strength of the cell suspension. The cells were resuspended in 10% glycerol at a concentration of 3 x 10^{10} cells/ml, frozen in dry ice, and stored at -70°.

Electroporation was carried out according to the manufacturer's instructions (Invitrogen). 2 µl of ligated DNA was mixed with 50 µl of electrocompetent cells, deposited into an 0.1 cm electroporation cuvette, which was electropulsed at 1500 V, with a capacitance of 50 µF and a resistance of 150 Ω. 200 µl of a fresh overnight culture of *E. coli* (DH5αF') was added to facilitate transfer of the samples to tubes containing top agar, IPTG and X-gal, as described above. The top agar mixtures were spread onto the L-broth agar plates, which were then inverted, incubated overnight at 37° and stored at 4°.

2.2.2.5. Preparation of DNA templates

200 ml of 2YT broth was inoculated with 2 ml of a fresh overnight culture of *E. coli* (DH5αF') grown in 2YT broth. 1.2 ml aliquots were dispensed into six 24 well plates. Clear recombinant M13 plaques were visualised on a light box and transferred to the wells using sterile cocktail sticks. The plates were incubated at 37° for 6 h in a humidified benchtop shaking incubator. 1 ml aliquots were transferred to 1.5 ml microfuge tubes and centrifuged for 5 min at 12,000 rpm to pellet the bacteria. The supernatants were tipped into fresh tubes, and 120 µl of

20% PEG/2.5M NaCl was added. The samples were mixed vigorously and incubated overnight at 4°.

Phage was pelleted by centrifugation for 5 min at 12,000 rpm and the supernatants were aspirated using a water pump. Each pellet was resuspended in 100 µl of TE and mixed vigorously with 50 µl of phenol/TE for 1 min using a multitube vortexer. The samples were vortexed twice more with time intervals of 1 h and 10 min, respectively, between vortexing. Following centrifugation at 12,000 rpm for 2 min, the upper aqueous layer were transferred to fresh tubes. 250 µl of sodium acetate/ethanol was added to each tube, and the samples were mixed several times by inversion and incubated overnight at -20°.

Phage DNA was pelleted by centrifugation for 5 min at 12,000 rpm. The supernatants were aspirated using a water pump, and the pellets were washed with 500 µl of 95% ethanol. The ethanol was aspirated using a water pump and the pellets were air dried at 37° for 30 min, dissolved in 30 µl of TE, and transferred to 96 well microtitre plates. The plates were sealed and stored at -20°.

2.2.3. DNA SEQUENCING

2.2.3.1. Preparation and deprotection of primers.

Oligonucleotides were eluted from the column matrix in 1.5 ml of ammonia and deprotected by incubation at 55° for 5 h. The oligonucleotides were lyophilised overnight, resuspended in 200 µl of water and stored at -20°.

2.2.3.2. Polyacrylamide gel electrophoresis

The purity of freshly prepared oligonucleotides was checked by electrophoresis on a 16% polyacrylamide gel.

Two well washed glass plates separated by 1.5 mm spacers were sealed with 38 mm Scotch electrical tape and secured by two large foldback clips positioned over each spacer. 125 µl of 25% (w/v) APS and 125 µl of TEMED were added to 50 ml of acrylamide solution in 1 x TAE. The mixture was poured into the gel sandwich, and a 20 tooth Teflon well-forming comb was inserted. The gel was allowed to set for 10 min before the clips and tape were removed from the bottom of the gel. The gel sandwich was secured to the electrophoresis apparatus and 1 l of 1 x TAE buffer was added to the electrophoresis tanks.

5 µl of each oligonucleotide solution was mixed with 5 µl of TAE/formamide solution, denatured for 5 min at 100°, and cooled on ice for 5 min. The 20 tooth Teflon comb was removed from the gel and the samples were loaded onto the wells alongside a 10 µl marker sample of formamide dyes. The samples were electrophoresed at 150 V for 3 h until the bromophenol blue dye had migrated half-way through the gel. The gel sandwich was removed from the electrophoresis apparatus and the plates were separated. The gel was covered with clingfilm and placed on a fluorescent screen, visualised by shading under UV irradiation and photographed.

2.2.3.3. Annealing primer

Annealing reactions were carried out in 96 well microtitre plates. DNA templates were thawed and vortexed briefly, and 4.5 µl from each template was transferred to a fresh microtitre plate using an EDP-Plus M-8 pipette. 15 µl of annealing mixture (225 µl M13 specific primer (diluted 1:1000), 180 µl TM buffer and 1095 µl water) was dispensed onto the wall of each well using an EDP-Plus pipette. The microtitre plate was sealed, vortexed and briefly centrifuged in a Beckman GPR tabletop centrifuge. The plate was incubated at 37° for 30 min and stored at -20°.

2.2.3.4. Sequencing reactions

Template DNA was sequenced by the dideoxynucleotide chain termination method based on that of Sanger *et al.* (1977). Sequencing reactions were performed in 96 well microtitre plates, each plate providing material sufficient for eight sequencing gels. Annealed DNA templates were thawed and briefly vortexed. 2.5 µl of annealed template DNA was transferred into each of four wells (for G, A, T, or C reactions) in four fresh 96 well microtitre plates using an EDP-Plus M-8 pipette.

Four sequencing mixtures sufficient for 96 sequencing reactions were made up: 10 µl 0.1M dithiothreitol (DTT), 250 µl of the appropriate dNTP mixture (dGTP, dATP, dTTP or dCTP), 8 µl dATP^{35}S and 2 µl Klenow fragment (diluted 1:15). 2.5 µl aliquots from these mixtures were dispensed to the sides of appropriate wells using an EDP-Plus pipette. The microtitre plates were sealed, vortexed and briefly centrifuged in a Beckman GPR tabletop centrifuge, and incubated at 37° for 10 min. Using an EDP-Plus pipette, 2.5 µl of chase solution (containing Klenow fragment) was aliquoted to the side of every well. The microtitre plates were sealed, vortexed, centrifuged and incubated at 37° for 10 min. The sequenced templates were stored at -20°.

2.2.3.5. Sequencing gel electrophoresis

Buffer gradient gels were prepared in pairs or groups of four. For each gel, two clean siliconised glass plates were polished with 95% ethanol, and the sequencing gel sandwich assembled around a pair of vertically positioned (0.4 mm) spacers. The sandwich was sealed on the sides and bottom using 38 mm Scotch electrical tape and secured with two large foldback clips positioned over each spacer.

60 ml of top gel mixture (TGM) and 12 ml of bottom gel mixture (BGM) were cooled on ice. 25 µl or 95 µl respectively of 25% (w/v) ammonium persulphate (APS) was dispensed into the BGM and TGM solutions. The same volumes of TEMED were also dispensed and the two solutions were mixed.

12 ml of polymerising TGM were drawn into a 25 ml pipette, followed by 12 ml of polymerising BGM. The interface was disturbed with a few air bubbles and the mixture was pipetted slowly into the gel sandwich. Bubbles were dislodged by striking the glass plates vigorously. The remaining 35 ml of polymerising TGM was dispensed evenly into the sandwich and bubbles were removed. The gel sandwich was placed almost horizontally and two narrow 24 lane sharks-tooth combs were inserted adjacently, 0.5 cm into the gel (teeth uppermost). Three large foldback clips were positioned to secure the combs, and the gel was allowed to set for 30 min. The clips and the tape from the bottom of the gel sandwich were removed. The gel was clamped into a BRL model S2 sequencing kit and the reservoirs were filled with 0.5 x TBE. The combs were removed from the gel and the well was washed out three times with 0.5 x TBE using a 60 ml syringe fitted with a needle.

Several non-buffer gradient gels were also made. These were assembled identically to the buffer gradient gels mentioned above but lacked BGM and were comprised wholely of TGM.

The microtitre plate containing sequenced templates was thawed, and 2.5µl of formamide dyes were added to each well using an EDP-Plus pipette. The covered plate was vortexed briefly, placed in a boiling waterbath for 1 min to denature the samples and chilled on ice for 5 min. The combs were repositioned with their teeth touching the surface of the gel, and 2µl from each well were loaded onto the gel from left to right, in the order GATC.

Samples were electrophoresed on buffer gradient gels at 60W until the lower dye front (bromophenol blue) reached the bottom of the gel (approximately 2 h). The gel sandwich was removed from the sequencing apparatus, and the glass plates were separated using the point of a pair of scissors. The gel was soaked three times with 10% acetic acid for 5 min in a fume hood, transferred onto two sheets of

3 MM paper, and covered with clingfilm. The gel was dried on a gel drier for 1 hour at 80° under vacuum and exposed to X-ray film overnight.

2.2.4.0. GENOME ASSEMBLY AND ANALYSIS

2.2.4.1. Reading and assembly of sequences

DNA sequence data were read from the autoradiographs using a computer-linked gel reading device (Summagraphics digitizer) and were labelled according to template number on the microtitre plates. The database was compiled by a computer-driven process of overlapping individual sequences, using version 1.0 of Staden's sequence assembly package (SAP) (Staden, 1987). Software was operated in a DEC Alpha 2100 computer.

2.2.4.2. Analysis of the completed sequence

The completed DNA sequence of HHV-7 RK was manipulated and analysed using the Wisconsin package, Genetics Computer Group (GCG) Madison, Wisconsin (Devereux *et al.*, 1984), running on Open VMS AXP, version 7.1. Several of the programs used are listed below. Assemble was used to construct the genome sequence from the database consensus sequence. Composition was used to determined the nucleotide composition of the sequence. Gap was used to align the DNA and protein sequences. Candidate open reading frames were identified using Frames. Findpatterns was used to identify candidate polyadenylation sites. Codon usage was examined using Codonfrequency and Codonpreference. Hydrophobicity profiles were prepared using PepPlot, using the parameters described by Kyte and Doolittle (1982). DNA sequences were translated using Ptrans (Taylor, 1986).

Chapter 3

Results

3.1. RANDOM SHOTGUN SEQUENCING

Of the complete herpesvirus sequences obtained in this Institute, four (CCV, EHV-1, EHV-2 and EHV-4) were determined by direct random shotgun sequencing of genomic DNA (Davison, 1992; Telford *et al.*, 1992, 1995, unpublished data). The procedure has been described by Davison (1991) and involves sequencing by the dideoxynucleotide chain termination technology (Sanger *et al.*, 1977) large numbers of random genomic DNA fragments generated by sonication of genomic DNA and cloned directly into bacteriophage M13 vector. The DNA sequences are visualised by autoradiography following electrophoresis on high resolution denaturing gels, and "read" using a computer linked gel reading device (usually a digitiser). A database is compiled using Staden's sequence assembly program (SAP) (Staden, 1987) and every nucleotide in the completed sequence is manually checked on both strands and edited by reference to the autoradiographs. Lastly, the sequence is analysed using programs from Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin. For this project, given the limited availability of HHV-7 DNA, the whole procedure was tested using HSV-1 DNA until good quality sequence data was reliably produced.

3.2. PREPARATION OF VIRAL DNA

Viral DNA was produced in the laboratory of Professor Niza Frenkel at the Tel-Aviv university, Israel, and the procedure is described in Section 2.2.1.1. Four samples of HHV-7 RK genomic DNA (HHV-7 a to d), totalling several micrograms, were provided by Professor Frenkel.

3.3. PREPARATION OF SEQUENCES

3.3.1. Identity and purity of the DNA

The identity of one representative sample was checked by restriction endonuclease analysis with *Hin*dIII and *Sal*I, followed by agarose gel electrophoresis (Fig. 7). Although the restriction patterns were somewhat blurred, they appeared to match published profiles (Wyatt and Frenkel, 1992), thus confirming that this sample was HHV-7 RK DNA.

Aliquots of the four samples were assessed for contamination with smaller DNA fragments by electrophoresis. A few micrograms of full length genomic DNA was detected in each sample

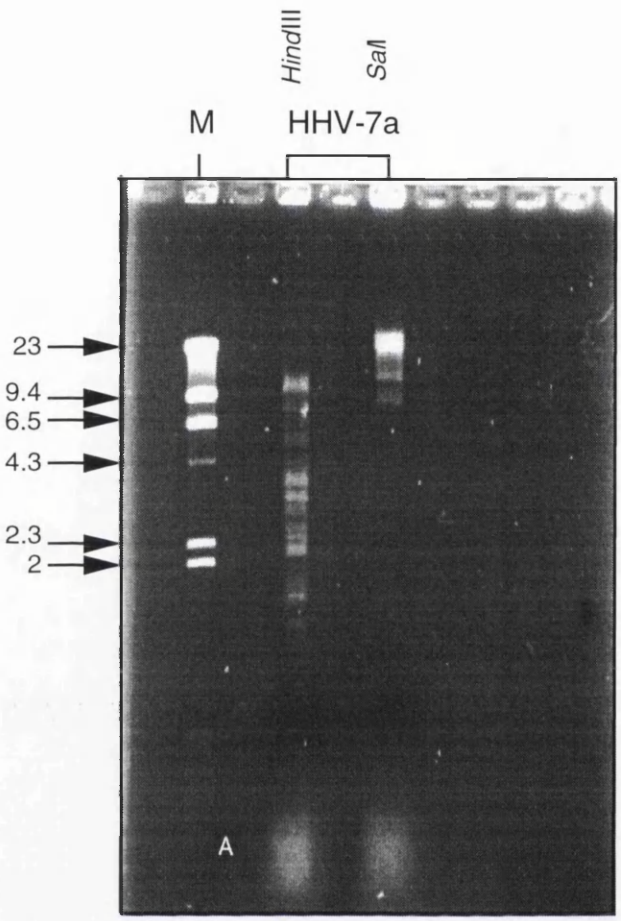


Fig. 7. *Hin*dIII and *Sal* I restriction profiles of HHV-7 RK genomic DNA .

The EtBr-stained 0.7% (w/v) agarose gel shows HHV-7a DNA digested with *Hin*dIII or *Sal*I. M is a *Hin*dIII digest of bacteriophage lambda DNA as size marker (kbp). A is smaller size DNA contamination.

(Fig. 8). Each sample also contained a much larger amount of smaller DNA fragments. HHV-7b and d displayed a smear of larger size DNA contamination (smaller than genome size, up to several kbp), and HHV-7d additionally contained smaller size (100 bp or less) fragments. HHV-7a and c showed no evidence of larger size DNA contamination, but did contain substantial amounts of smaller size fragments. Prior to purification, at least 50% of each HHV-7 DNA sample was estimated to comprise the contaminating smaller DNA fragments. The provenance of these fragments was not determined, and they may represent residual genomic or cellular DNA degraded during treatment of capsids with DNase.

Based on the predicted size of the HHV-7 genome (145 kbp; Black and Pellett, 1993), completion of the DNA sequence was calculated to require 5,000-5,500 clones (200-230 sequencing gels). A contamination level of 10% (involving an additional 560 clones) might have been acceptable, but 50% contamination would have doubled the size of the task. Thus, starting with DNA that was as pure as possible was fundamental to the efficiency of the project.

3.3.2. DNA purification

One sample of genomic DNA (HHV-7a; cytoplasmic fraction), which was heavily contaminated with fragments smaller than 100 bp (Fig. 8), was purified by precipitation using 8% polyethylene glycol (Section 2.2.1.5.). This step effectively removed the contaminating DNA (Fig. 9). An attempt was made also to purify the remaining samples (HHV-7b, c and d), by centrifugation on a glycerol gradient, as described by Van Zijl *et al.* (1988), but only a very small quantity of full-length DNA was finally recovered (data not shown).

3.3.3. Fragmentation of the DNA

It was important to start with fragmented DNA of an appropriate size (400-700 bp). If the fragments were too large they would not ligate efficiently into the M13 vector, and this would have increased the amount of effort needed to generate the required number of clones. Equally, if the fragments were too short, more M13 recombinants would have contained multiple inserts, and this would have increased the number of non-contiguous sequences in the database.

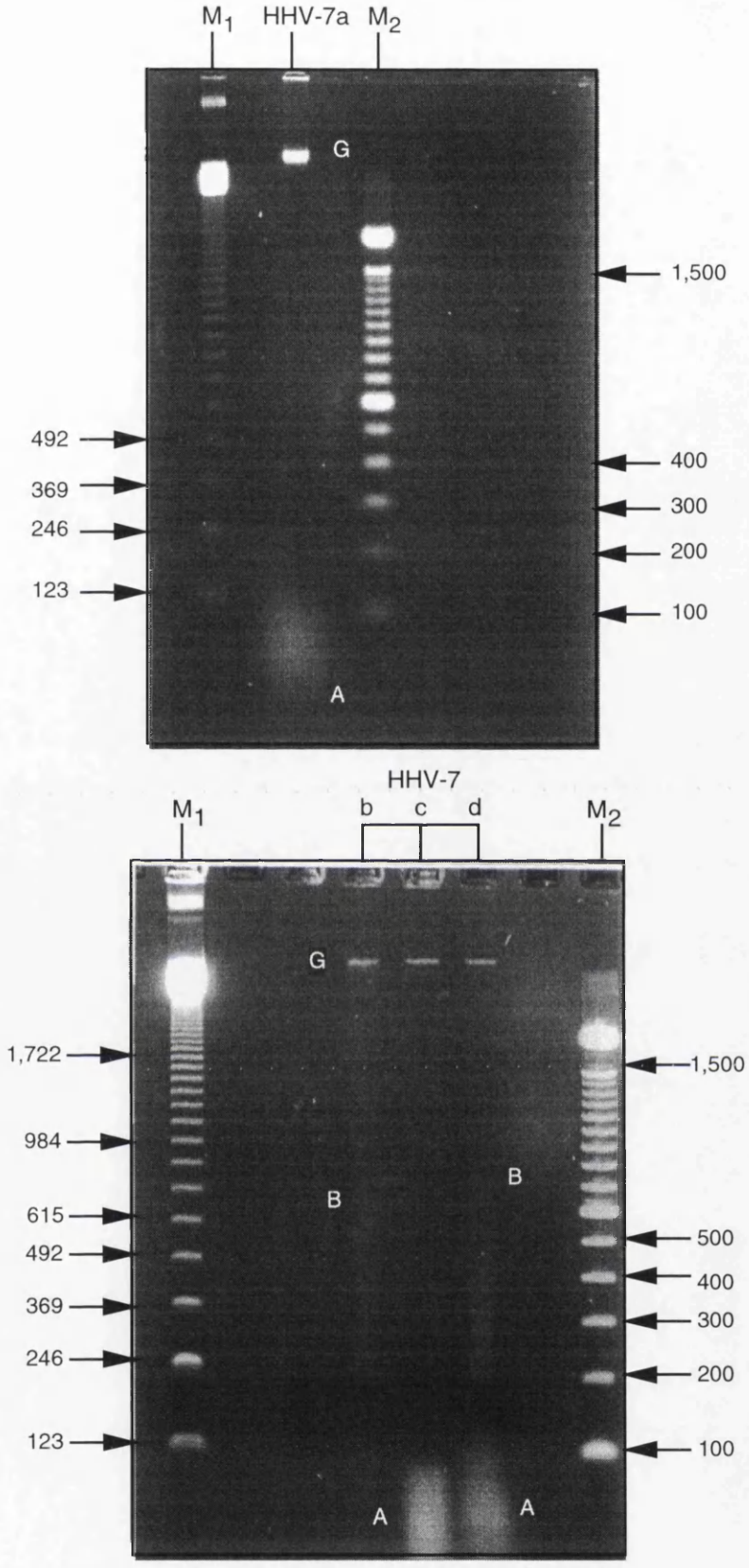


Fig. 8. Status of HHV-7 strain RK genomic DNA

The EtBr-stained 0.7% (w/v) agarose gels show untreated HHV-7 a to d. G is full length genomic DNA, A and B are smaller and larger size contaminants, respectively, M1 and M2 are 123 bp and 100 bp DNA ladders as markers.

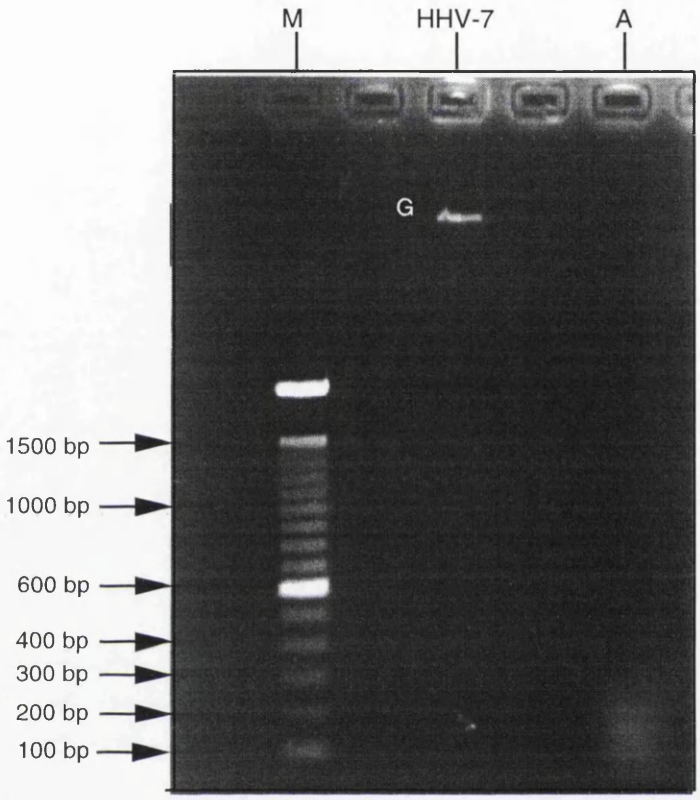


Fig. 9. Purification of HHV-7 RK genomic DNA.

The EtBr-stained 0.7% (w/v) agarose gel shows PEG-precipitated HHV-7 RK genomic DNA. G is full length genomic DNA precipitated by PEG, and A is small size contaminants. M is a 100 bp DNA ladder as markers.

50 µl of purified genomic DNA from sample HHV-7a was fragmented by sonication. 5 µl of sonicated DNA was visualised by electrophoresis to assess the extent of fragmentation, and a sufficient proportion of the DNA was found to be close to 600 bp in size (Fig. 10). The remaining 45 µl of DNA was electrophoresed on an agarose gel, utilising a 100 bp DNA ladder as a marker. The agarose gel was divided into two halves and the half containing the marker was stained with EtBr. The gel was reassembled and the marker visualised by long wavelength UV irradiation. Sonicated HHV-7 fragments in the range of 400-700 bp were excised from the gel (Fig. 11), and purified using a Geneclean II kit. The purified DNA fragments were resuspended in 50 µl of TE, and an aliquot of 5 µl was visualised by electrophoresis to assess recovery (Fig. 12).

3.3.4. End repair and ligation

In order to clone the sonicated DNA fragments into the vector it was necessary to repair any ragged ends. The fragments were treated with T4 DNA polymerase in the presence of the four dNTPs to produce flush ends, and ligated into dephosphorylated *Sma*I-cleaved bacteriophage M13 mp19 RF I DNA.

3.3.5. Transfection and electroporation

DNA templates were generated by transfection of ligated DNA into *E. coli* DH5αF', as described in Section 2.2.2.3.. A typical transfection yielded a mixture of blue (M13 phage lacking insert) and clear plaques (M13 phage containing HHV-7 insert DNA). Plaque numbers were invariably low, probably as a result of the small amount of HHV-7 DNA in the ligation. In an attempt to enhance the yield, electroporation was employed as an alternative. It was not, however, more efficient. In total, approximately 600 clones were made by electroporation.

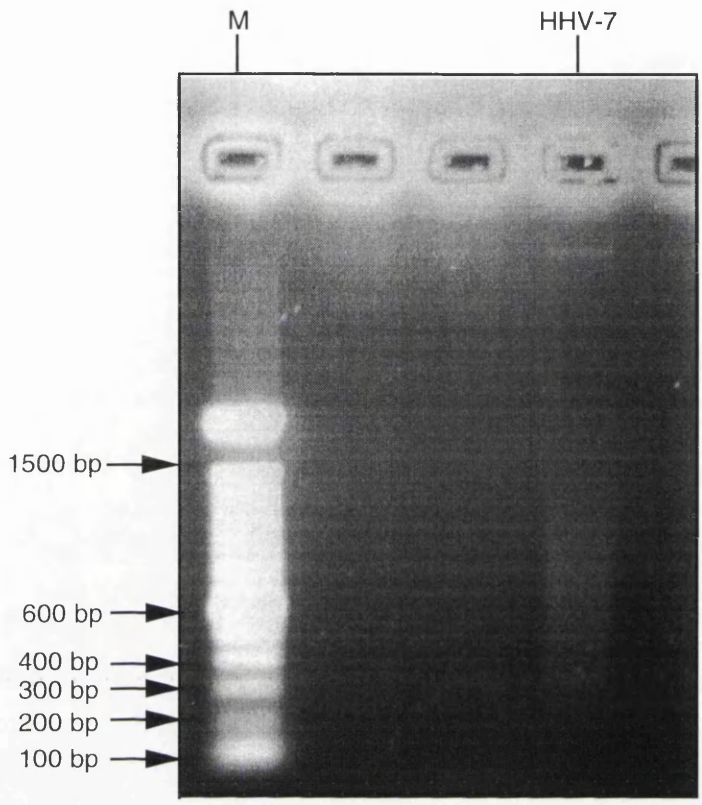


Fig. 10. Status of randomly fragmented HHV-7 RK genomic DNA.

The EtBr-stained 0.7% (w/v) agarose gel shows HHV-7 DNA randomly fragmented by sonication. M is a 100 bp DNA ladder as markers.

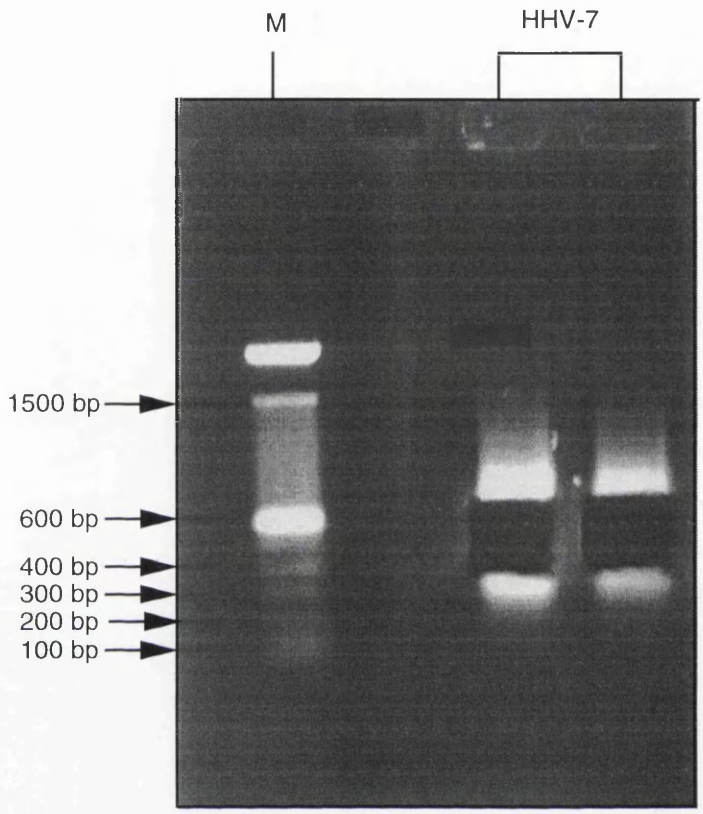


Fig. 11. Recovery of 400-700 bp fragments from randomly fragmented HHV-7 RK genomic DNA.

The EtBr-stained 0.7% (w/v) agarose gel shows randomly fragmented HHV-7 DNA following excision of the 400-700 bp fragments. M is a 100 bp DNA ladder as marker.

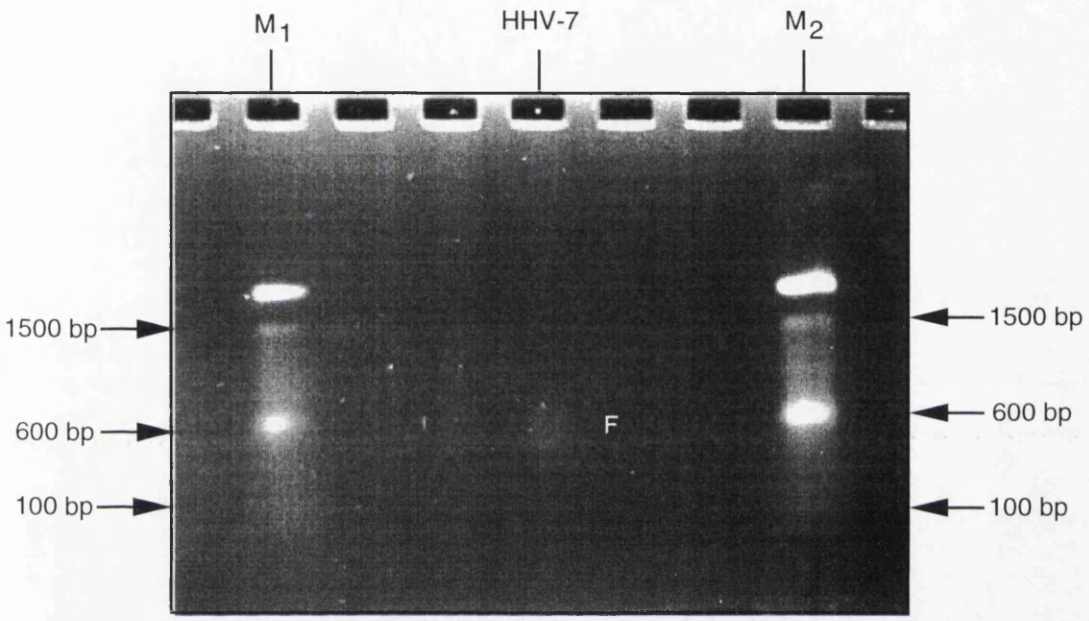


Fig. 12. Recovered 400-700 bp fragments of HHV-7 RK genomic DNA.

The EtBr-stained 0.7% (w/v) agarose gel shows 400-700 bp randomly generated fragments of HHV-7 RK DNA (F), M1 and M2 are 100 bp DNA ladders as markers.

3.3.6. Growth of recombinant phage

Growth and harvesting of clones was usually spread over three days, and clones were processed in sets of 144. On the first day, clear plaques were picked from agar plates, and recombinant phage was grown in small cultures of *E. coli* and precipitated overnight by PEG precipitation. On the second day, phage was pelleted and template DNA extracted using phenol, ethanol precipitated and resuspended.

Initially, 5,200 templates were generated, but this number proved to be insufficient owing to a significant number of poor quality templates. A further 1,800 templates were prepared and sequenced in order to bring the project to a successful conclusion. This brought the final number of templates to approximately 7,000.

3.3.7. Sequencing

Typically, templates were sequenced in batches of 96 or 192. Templates were loaded on gels in batches of 24, and the gels were run in groups of two, four or eight. In total, about 290 sequencing gels were involved in the project. Autoradiographs were read into a computer using a Summagraphics digitizer and an average of 207 nucleotides were obtained from each template, although sequences of over 300 bp were not obtained uncommonly. Rare templates with very short inserts (less than 30 bp) were ignored, along with poor quality sequences. The templates were stored in 96 well microtitre plates (74 in total). Each plate was divided into four sections (1-4) which corresponded to four sequencing gels, each accommodating 24 templates (1-24). Thus, in the database, reading 14_4.12 was for example the sequence generated from template 12 in the fourth quarter of plate 14. Thus, every sequence in the database was easily associated with the template and autoradiograph from which it was obtained.

3.4. ASSEMBLY OF THE DATABASE

The database was compiled with the aid of Staden's sequence assembly package (SAP; Staden, 1987), which facilitates computer-driven overlapping of individual complementary sequences. The general stages in the assembly of the sequence database are summarised below. Assembly of reiterated sequences presented unusual problems and is discussed in Section 4.2.7.

3.4.1. Initial assessment of DNA templates

In the initial stages of the project, a check was made on the quality and identity of the templates by entering approximately 200 sequences into SAP. A number of encouraging observations were made: the sequences were of an appropriate length; the number of overlaps between sequences was empirically similar to those found at the same stage in previous herpesvirus sequencing projects (Davison, 1991); several sequences contained a short reiteration (TAACCC) characteristic of the telomeric repeats of HHV-6 and HHV-7; and some sequences evidently originated from the HHV-7 major capsid protein gene (Mukai *et al.,* 1995). These observations provided reassurance that the majority of templates contained HHV-7 DNA.

3.4.2. Automatic input of sequence data

The progress of the assembly of the database is charted graphically on Fig. 13 and recorded in Table 4. Before being added to the database, sequences were checked for the presence of M13 vector sequence. Once entered into SAP, each batch of new sequences was automatically screened against those already in the database, with four potential outcomes. Each new sequence would: match and be joined to a stretch of contiguous sequences (contig) in the database; not match the database and be added as a single sequence contig; overlap a contig but fail to meet the entry criteria, and be rejected as a poor match; or match, and thus join, two contigs. Entry criteria included the minimum size of overlap between sequences (15 bp), the maximum percentage of non-matching nucleotides and number of padding characters needed to optimise the match. These criteria could be adjusted, but in the majority of the project were 15 bp, 12% and 8 characters, respectively. An example of entry of a single sequence resulting in the joining of two contigs is shown in Fig. 14. The new sequence overlapped two contigs and was added to the end of the contig with the closest match. The elongated contig was then compared with the second contig, in order to reassess the quality of the overlap, and the two contigs were joined.

A significant number of overlapping sequences were refused automatic entry into the database, even under the most relaxed criteria, owing to the accumulation of errors in the database. By necessity these sequences were entered manually (stage A in Table 4 and Fig. 13).

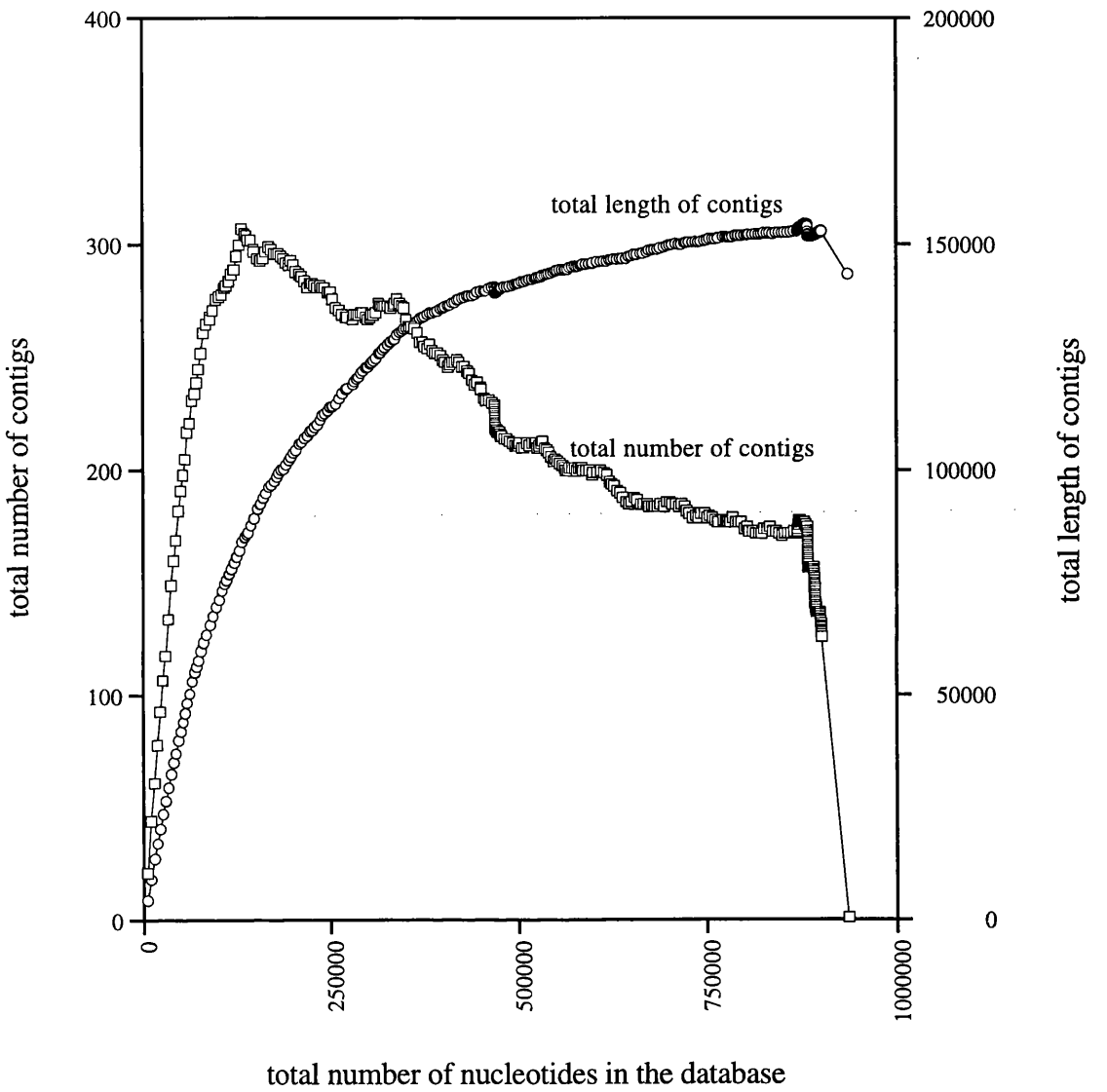


Fig. 13. Database assembly

Table 4. Stages in assembly of the HHV-7 database.

Stage	Number of Sequences	Number of Contigs	Total number of characters[1]	Total contig length
A) Automatic and manual input of sequences	4,285	170	882,813	153,712
B) Searching the database with contig ends	4,321	157	890,834	151,961
C) Entry of extended sequence data from manual reading, long gels and custom primers	4,401	126	901,286	152,952
D) Input of reiterated sequences	4,691	162[2]	-	-
E) Removal of contigs containing 1-2 sequences[3]	4,522	1	936,051	143,424[4]

[1] Total number of nucleotides and padding characters.
[2] The number of contigs increased because several sequences containing repeats were present in the database initially, and were duplicated when the completed repeat arrays were entered (Section 3.5.).
[3] Duplicated contigs were also removed.
[4] The final total contig length (prior to editing), representing the unique sequence and one copy of the direct repeat.

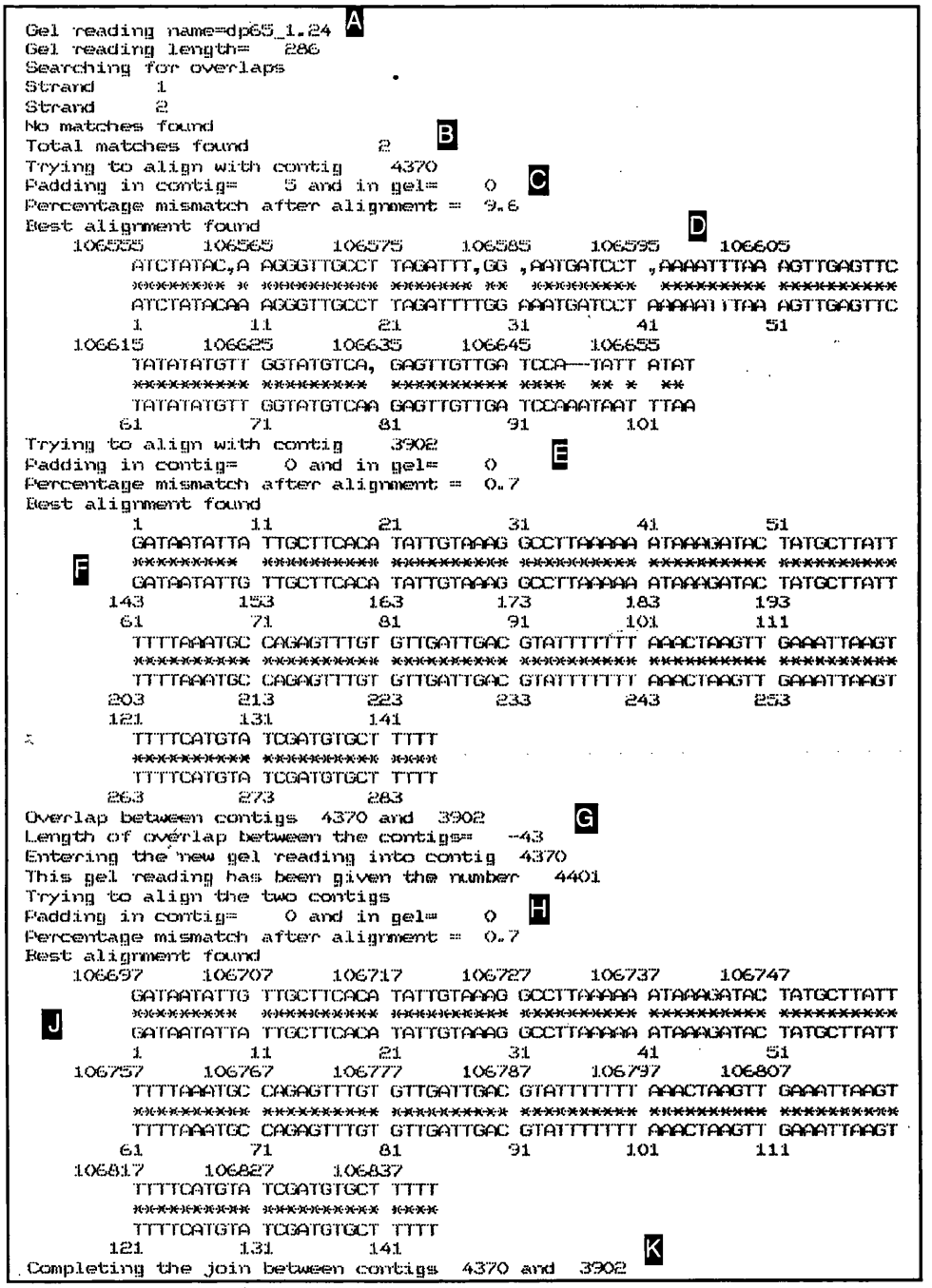


Fig. 14. Entry of a single sequence into the HHV-7 database

A- new sequence (dp65_1.24)
B- 1st contig (4370) matched by the new sequence
C- details of the closeness of the match
D- display of the match
E- 2nd contig (3902) matched, and details of the closeness of the match
F- display of the match
G- new sequence joined to the end of 1st contig
H- new sequence and 1st contig match 2nd contig
J- display of the match
K- contigs joined (by the new sequence)

Screening the contig ends - To reduce further the number of contigs in the database, sequences from the ends of each contig were compared against the whole database to ascertain whether any contigs could be joined (stage B in Table 4 and Fig. 13). Also, the autoradiographs from which the sequences at the contig ends had been obtained were carefully re-examined by eye with the aid of a light box, and many sequences were extended. These extended sequences were instrumental in promoting contig joining (stage C in Table 4 and Fig. 13).

Database assessment - By this stage, the database was in an advanced state. Although there were still a relatively large number of contigs, only eight contained more than two gel readings (multiple sequence contigs), and ranged in size from 300 bp to 46.5 kbp. The remainder (126 contigs) were all single sequence contigs except one, which contained two sequences. Although a few single sequence contigs overlapped poorly with the multiple sequence contigs, most did not match, even under the most relaxed conditions. In addition, the ends of three of the multiple sequence contigs and the middle of a fourth contained reiterated sequences. These were in various stages of organisation and were by no means complete.

Final joins - In order to assemble the multiple sequence contigs, several appropriate templates from the contig ends were resequenced and electrophoresed over an extended period on gels lacking a buffer gradient, in an attempt to produce elongated "reads" that might overlap other contigs. To extend these data, customised primers (Table 5) were also made to match sequences at the contig ends. The combined use of these approaches allowed the remaining multiple sequence contigs to be assembled into one large contig (Fig. 13 and stage D in Table 4).

Table 5. Custom oligonucleotide primers used to extend sequences in order to join contigs.

Primer	Sequence	Location*
P503.01	CCC AAT GAG TAC ATT	69970-84
P762.18	AAT GAC TGT GAC AAA	114893-907
P494.06	ACC GTA TTG GTA CAT	142939-53
P181.07a	GAT GAA AGA TTT CTA	115322-36
P181.07b	ACA TCG ATA CAT GAA	115184-98
P113.17a	GTC GGT GAA TGT AGA	25980-94
P113.17b	ATT ATA CTA GCA GAT	25752-66
P423.23a	GAT GTC TAT AGC ATC	46251-15
P423.23b	ATA GAT AGA GCT GAA	46440-54
P494.10	CAA TTT ACA CAT AGA	69769-83
P514.15	CAG TTA ATG TTG CCA	51777-91
P491.24a	TTG CCT TTA AAA GAA	25438-52
P491.24b	AGG AAG CAC TAC ACC	25563-77
P692.21	AAT AAC ATT GTT GAC	46829-43
P634.03	ATG GAT CAT AGA GTA	51015-29
P741.07	CAT ATC ACA GTG AGA	51355-69

*Location of the primer in the completed HHV-7 RK genome (bp), regardless of strand.

3.5. ASSEMBLY OF REITERATED SEQUENCES

Four regions containing sequences reiterated in head-to-tail orientation were observed within the database: R1, R2 and two sets of telomeric reiterations (T1 and T2). None of the regions assembled readily in the database and each had to be dealt with individually.

3.5.1. R2

R2 involved 71 sequences, each containing all or part of a 105 bp reiterated sequence. Although very similar, the reiterations were not identical and displayed nucleotide substitutions, deletions or insertions. Three approaches were attempted to determine the order of the reiterations. Firstly, a special database was created specifically for the R2 sequences. The SAP algorithm, however, lacked the sensitivity required to assemble automatically so many almost identical sequences. Secondly, each sequence was screened against the others to provide information about how they overlapped. These data were then transcribed onto a series of cards to produce a "jigsaw", each carrying one sequence name and the names of the sequences it could overlap. Eventually, most of the sequences were matched, by eye, to one of two groups. Unfortunately, the two halves of the "jigsaw" could not be joined and the arrangement of the reiteration remained incomplete. Eventually, assembly of the R2 sequences was completed by making sequence alignments by eye in a simple computer text file. Differences between the 105 bp reiterations made it possible to group overlapping sequences, and then to join groups together. During this process all of the sequences were checked against the autoradiographs and many sequences were extended. Also, templates were resequenced to lengthen some sequences. The sequences were assembled successfully and the arrangement of the 105 bp reiterations was determined. The completed sequence of RK reiteration R2 can be seen in Fig. 15, alongside its counterpart in JI.

3.5.2. R1

The R1 array has a structure like that of R2, but contains fewer reiterations and involved fewer sequences. Hence the sequence was readily determined in a manner similar to that used for R2. The completed sequence of RK reiteration R1 can be seen in Fig. 16, alongside its counterpart in JI.

RK

```
 1  ATTTAAATCTTTGAAA.CAA.CAGGAATTGCGGTTGTCTTCTCAGCAAGTTCTACACATTCCCACTCGCTTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 2  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTACGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTAAC *
 3  ATTTTAATATGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTACGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 4  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 5  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 6  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
 7  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGTCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
 8  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTACGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC -
 9  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGTCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC -
10  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
11  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGATTCTCACCCG.TTTAGGCAGGAAAGACCCCGACCACATAACCCACTGAC -
12  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TT.AGGCAGGAAAGACCCGAACCACATAACCCACTGAC -
13  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCATCCG.TTTAGGCAGGAAAGACCCCGACCACATAACCCACTGAC +
14  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGTAGGAAAGACCCAGACCGCATAACCCACTAAC -
15  ATTTTAATATGTGAAAATAAACAGGAACTGCGGTTATAGTTTCAGTACATTCTAGAGATTACATCCTG.CTAAGGCGAAAAACACTTTAACCGCAAAAGCCACTGAT *
16  TTTTAAACTTGTGAAAATAA.CAGGAA                                                                                  *
```

JI

```
 1  ATTTAAATCTTTGAAA.CAA.CAGGAATTGCGGTTGTCTTCTCAGCAAGTTCTACACATTCCCACTCGCTTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 2  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTACGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTAAC *
 3  ATTTTAATATGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTACGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 4  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 5  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC *
 6  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGTCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
 7  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
 8  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGTCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
 9  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
10  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
11  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGTCTTCTCAGCAGGTTTTATGAGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAACCCACTGAC +
12  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCCAACCACATAATCCACTGAC -
13  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAAGACCCGAACCACATAACCCACTGAC +
14  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCATCCG.TTTAGGCAGGAAAGACCCCGACCACATAACCCACTGAC +
15  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGCCTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGCAGGAAACACCCAGACCGCATAACCCACTGAC -
16  ATTTAAACCCGTGAAAACAA.CAGGAACTGCGGTTGACTTCTCAGCAGGTTTTATGGGTTCTCACCCG.TTTAGGTAGGAAAGACCCAGACCGCATAACCCACTAAC -
17  ATTTTAATATGTGAAAATAAACAGGAACTGCGGTTATAGTTTCAGTACATTCTAGAGATTACATCCTG.CTAAGGCGAAAAACACTTTAACCGCAAAAGCCACTGAT *
18  TTTTAAACTTGTGAAAATAA.CAGGAA                                                                                  *
```

Fig. 15. The structure of the R2 reiteration in HHV-7 strains RK and JI.

Nucleotides differing from the consensus 105 bp sequence are displayed as bold characters.

* Counterparts found in both strains, in identical positions.
\+ Counterparts found in both strains, in different positions.
\- No counterpart found in the other strain.

RK

ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGTAAATATTTGCACATGCTAATGTGTTCATGTGGGTAT
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCA**T**ATATTTGCACAT**A**CTAATGTGTTCATGTGGGTAT
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCA**T**ATATTTGCACAT**A**CTAATGTGTTCATGTGGGTAT
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGTAAATATTTGCACATGCTAAT

JI

ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGTAAATATTTGCACATGCTAATGTGTTCATGTGGGTAT
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCA**T**ATATTTGCACAT**A**CTAATGTGTTCATGTGGGTAT
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCA**T**ATATTTGCACAT**A**CTA**GCATATATA**TGT**AAT**
ATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGTAAATATTTGCACATGCTA

Fig. 16. The structure of the R1 reiteration in HHV-7 strains RK and JI.

Nucleotides differing those in the first repeat (84 bp) are displayed as bold characters.

3.5.3. T1 and T2

T1 and T2 comprise arrays of a simple hexameric element (TAACCC) which is similar to human telomeric sequences, plus a total of 34 types of related element, mostly hexanucleotides but some longer (Table 6). The telomeric arrays are quite different from R1 and R2. The reiterations are small and numerous and, to complicate matters further, are split between two loci near either end of DR. Neither T1 nor T2 was successfully assembled using SAP. As the telomeric elements varied in sequence, the approach used to resolve R1 and R2 was attempted. However, the large number (229) of separate sequences and reiterated elements therein made assembly by eye difficult. Thus, it was essential to simplify the data. This was done by representing each type of reiteration as a single character. The predominant element (TAACCC) was represented as a dash, and related elements were designated as letters or symbols (listed in Table 6). This made overlapping the sequences by eye much easier. As with R1 and R2, overlapping sequences were grouped, and then the groups were linked together.

Of the two telomeric reiterations, the region at the right end of DR (T2) is the smaller, containing 148 elements. A single solution which included all of the telomeric reiterations from this region was obtained quickly, and is displayed in Fig. 17. Unfortunately, assembly of the larger telomeric reiteration (T1) at the left end of DR, which apparently contains 663 elements, proved more difficult. Every sequence was checked manually and many were extended by eye or by resequencing before a solution was reached. The solution was the best that could be obtained from the available data, and is displayed in Fig. 18. It should be noted, however, that a few (less than ten) random sequences could not be incorporated. Differences between the telomeric reiterations of RK and JI are described later (Section 3.9.1.) .

Table 6. Telomeric repeat elements in HHV-7 RK, and the code used for their assembly.

Repeat[+]	Letter code
TAACCC	-
TAAGTC	B
TAGCTC	C
TGACCC	D
CAGCCC	E
CAACCC	G
TAAGCC	H
TAAACC	J
GAACCC	K
TAGCCC	L
TAAGTTCACCC	P
TAGGGCTG	Q
TAGGTC	R
CGGCCC	S
TAAATC	T
TAGCTGC	V
TACCCT	A
AATCCC	I
AACCCC	M
AATCCG	W
AATCTC	X
GATCCC	Y
TAATCC	Z
AAACCC	U
TAACCG TAACCT TAACCA TAGGCC TCACCC TTAGCCC CAATCC TAACTC TAAGTT TCACTG TAACAC	* (rare repeats)

[+]The repeat elements appear as they do in the genome sequence, and are complementary to the classic GGGTTA repeat.

RK

```
-*L*-L-C*C*D*-H-*CP--*DD-QS-L-ZZZL-QS-L--LQSJ---S--QS--QSKL--LQSK
L--LQS--LQSJ-LQS--QS--QS--QSK-----U-QS--QSK---QS--QS--*QS----QSKL
--DD--------------
```

JI

```
-*L*-L-C*C*D*-H-*CP--*DD-*S-L--ZZL-QS-L---QS--QSK----QS--QS--QSK-
--QS--QS--*QS----QSKL--DD----------------
```

Fig. 17. The structure of the telomeric array (T2) from the right end of DR in HHV-7 strains RK and JI.

The telomeric repeats are represented by a letter code, which is displayed alongside the repeats in Table 6.

RK

```
--C---C---B-D-B-D---C--C----L----------R---B--B------D-C---------B--B-*B--
--------B------B--------R---B--B------D-C----------------B---B------B--B--
--------B--------B--B--B------G--CH-G--CH-G--CB-G--CH-G--CH-G--CB-G--CPG--
CPG--CH-G--CH-G--CH-G--CB-G--CPG--CB-G--CB-G--CH-G--CH-G--CH-G--CH-G--CPG-
-CB-G--CB-G--CH-G--CB-G--CH-G--CB-G--CPG--CPG--CH-G--CPG--CH-G--CH-G--CH-G
--CH-G--CH-G-CPG--CH-G--CH-G--CJ-E---CB-G--CH-G--CH-G--CB-G--CB-G--CH-G--C
H-G--CH-G--CH-G--CH-G--CH-G--CB-G--CB-G--CB-G--CH-G--CH-G--CB-G--CPG--CH-G
--CB-G--CH-G--CH-G--CH-G--CH-G--CPG--CH-G--CPG--CPG--CH-G--CPG--CJ-E---CB-
G--CH-G--CB-G--CH-G--CH-E--CH-G--CPG--CB-U--CPGG-CB-U--CB-G-C*-E--C***C
```

JI

```
----CH-E--CH-G-C***C
```

Fig. 18. The structure of the telomeric array (T1) from the left end of DR in HHV-7 strains RK and JI.

The telomeric repeats are represented by a letter code, which is displayed alongside the repeats in Table 6.

3.6. EDITING THE DATABASE

3.6.1. Removal of single contigs

The sequence was completed without inclusion of the single sequence contigs. A very few short or poor quality single sequence contigs appeared to contain HHV-7 DNA inserts, as evidenced by partial matching of their sequences with the multiple sequence contig. However, the majority matched neither the multiple sequence contig nor each other, and were assumed to comprise cellular DNA. Thus, following entry of the completed reiterated sequences into the database (described above), the single sequence contigs were removed and the number of contigs was reduced to one (Fig. 13 and Table 4).

3.6.2. Editing the database

A paper copy of the single multiple sequence contig was printed, and every nucleotide in the sequence was checked (on both strands) by reference to the autoradiographs. There were many areas of uncertainty, and each was resolved by checking all the sequences in the region against the autoradiographs. Regions for which data were obtained on only one strand were examined particularly closely during editing. Finally, remaining ambiguities were resolved by analysis of specific templates from both strands using the universal or custom primers. Figs. 19 and 20 display the resolution of four typical errors in the database.

To correct the errors in the database, it was necessary to work from right to left along the contig, since following every correction the sequence to the right was renumbered by SAP. Manual correction of errors was time-consuming, and so corrections were entered using a batch file which could be left to run unattended. In all, the sequence was determined an average of 8.2 times per nucleotide, and 97% was determined on both strands. In addition, the proportion of random sequences obtained that did not match the final genomic sequence indicates that the DNA was 97% pure.

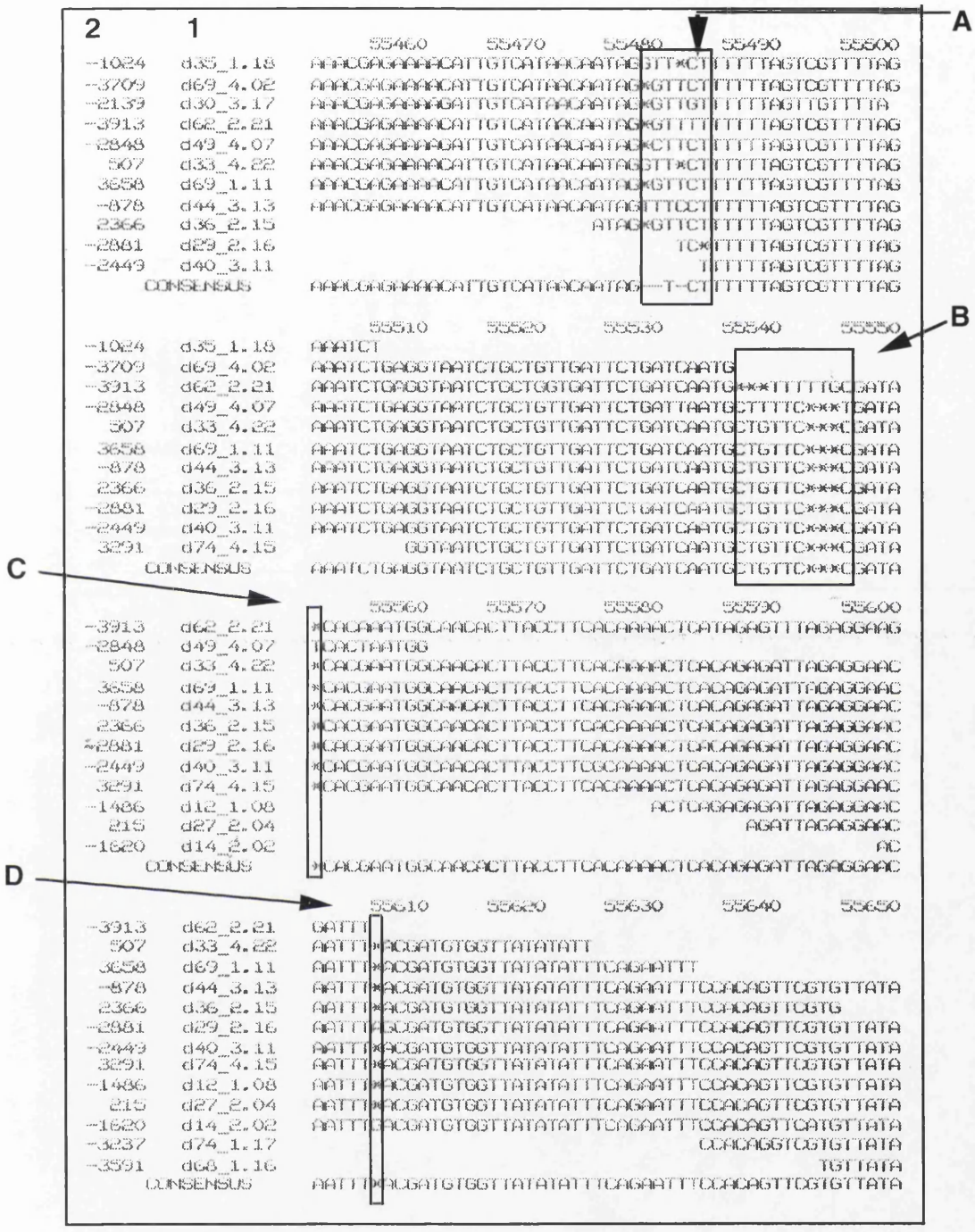


Fig. 19. Four unedited errors in the database.

The errors are all of a similar nature. A contains several sequences that are incorrect and B requires correction of two sequences, whereas C and D require single nucleotide deletions. 1 are the names of individual sequences and 2 are their SAP designations; sequences in the complementary strand are denoted with a dash. Asterisks are padding characters entered by the program.

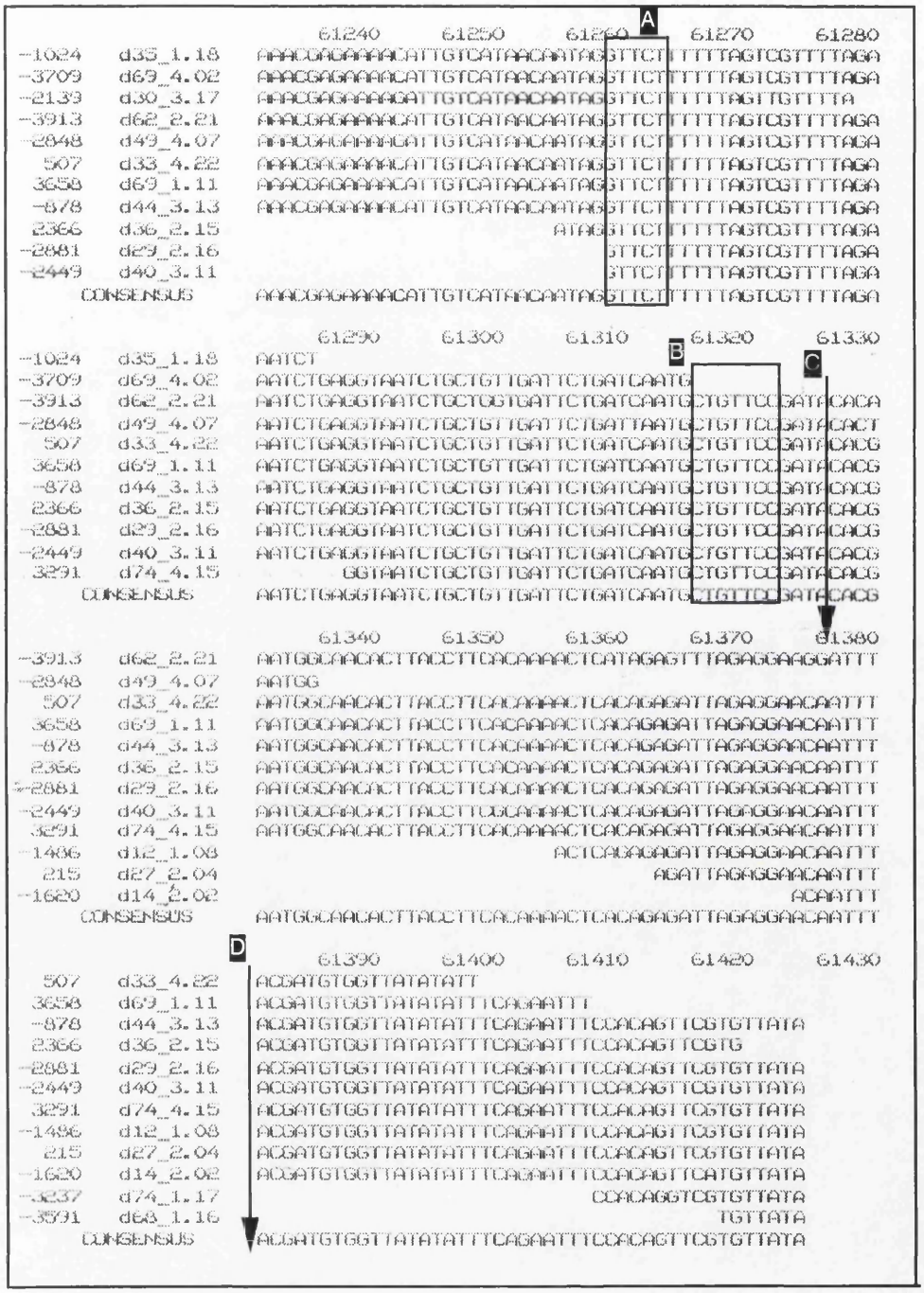


Fig. 20. Correction of four errors in the database (see Fig. 19).

3.6.3. Errors in the sequence

At this stage, and not before, the completed sequence was compared to that of HHV-7 strain JI (Nicholas, 1996), using the Gap program. The RK sequence differed from that of JI at approximately 200 positions. These positions were then reassessed carefully, by reference to autoradiographs as well as by resequencing templates when appropriate. As a result, 35 errors were identified in the RK sequence and corrected. The nature of these errors and their locations in the RK genome are listed in Table 7. Errors were located in regions of the genome that had been sequenced on one strand only or were covered by a single sequence, although several were located within compressions (aberrant migration of partially base paired sequencing products).

3.7. THE HHV-7 RK GENOME SEQUENCE

The structure of the HHV-7 genome comprises a long unique sequence (U) flanked by a direct repeat (DR). However, the database contained U flanked by portions of one copy of DR, since no evidence of differences between the DR sequences emerged when compiling the database. Thus, in order to construct the genome sequence from the database consensus, it was important to define the DR sequence. This was accomplished by locating by eye the ends of the genome, which also correspond to the U-DR and DR-U junctions. It is rare for the ends of two random sequences to map at the same nucleotide, in contrast to sequences from the ends of the genome. Thus, the genome ends, which are shown in Fig. 21, are denoted by a sharp change in contig depth at two specific points in the database.

The right genome end of RK (Fig. 21a) is identical to that of JI, but the left end is one base pair shorter, containing six rather than seven C residues (Fig. 21b). Most clones from the left terminus had six C residues and none had seven. As fragments were end-repaired prior to ligation, no information was obtained on unpaired nucleotides at the genome termini.

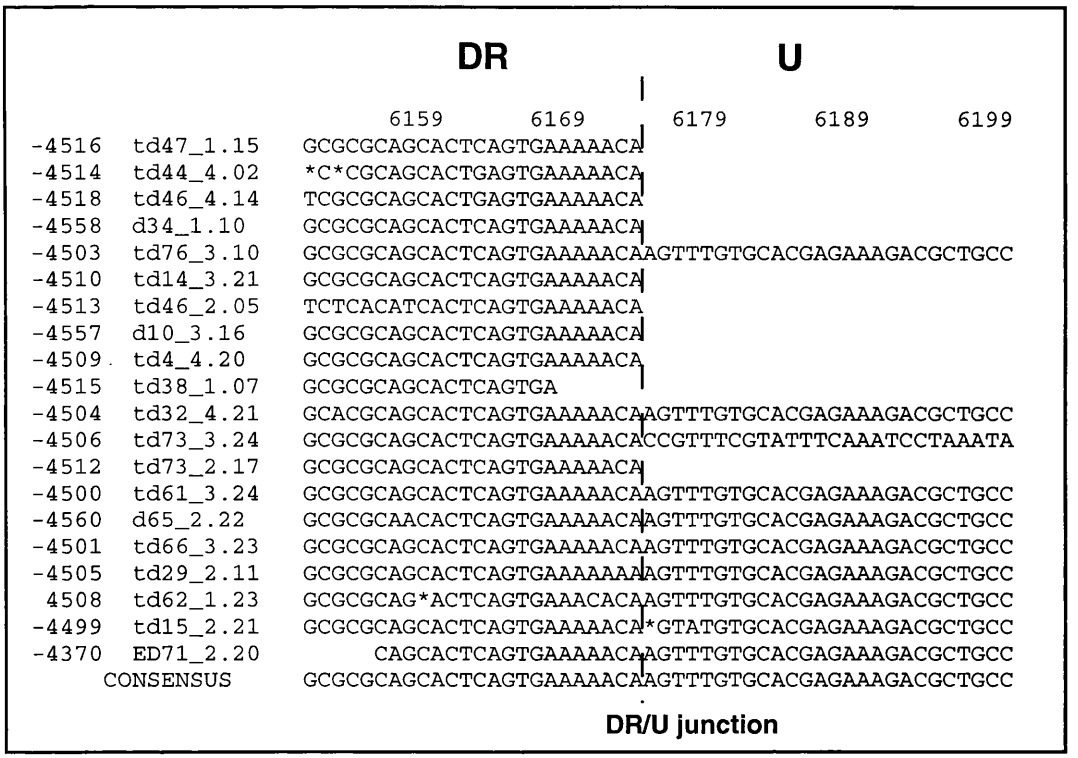
The HHV-7 RK genome sequence was reconstructed from the database consensus sequence using the Assemble program. The sequence is 153,080 bp in length, and comprises U (133,012 bp) flanked on each side by a copy of DR (10,034 bp). Overall, the genome contains 36% G+C,

Table. 7. Errors in the DNA sequence of HHV-7 strain RK

Position (bp)		Difference[a]	Sequence change		ORF
HHV-7 RK	HHV-7 JI		RK	JI	
8219 / 151251	/ 143316	ID	A	-	non-coding
8242 / 151274	/ 143338	ID	-	T	non-coding
14051	9830	ID	-	C	U4
18376	14156	S	G	T	U8
18996	14776	ID	-	G	U10
25580	21361	ID	-	G	U14
25589	21372	ID	-	A	U14
44347	40131	ID	-	G	U30
45463	41247	ID	-	A	U31
45493	41278	ID	-	A	U31
45499	41285	S	G	A	U31
56906	52692	ID	-	A	U38
60250	56044	ID	A	-	U39
60886	56670	ID	-	C	U39
67891	63674	ID	G	-	U42
67893-4	63676-7	MS	GG	AA	U42
73894	69677	S	A	C	U47
74201	69984	ID	G	-	U47
78657	74440	ID	-	A	U50
84842	80618	ID	-	G	U55A
91135	86913	ID	-	G	U57
106813	102592	ID	-	T	U73
106819	102598	ID	-	T	U73
110369	106150	S	C	G	U75
112891	108672	ID	-	T	U77
114913	110698	ID	-	A	U77
114931-4	110717-8	ID	- -	TG	U77
114936	110720	S	G	A	U77
123110	118897	ID	G	-	U85
125010	120784	ID	-	T	U86
129563	125349	S	T	G	non-coding

[a] ID is insertion or deletion, S is substitution and MS is multiple substitution.

A



B

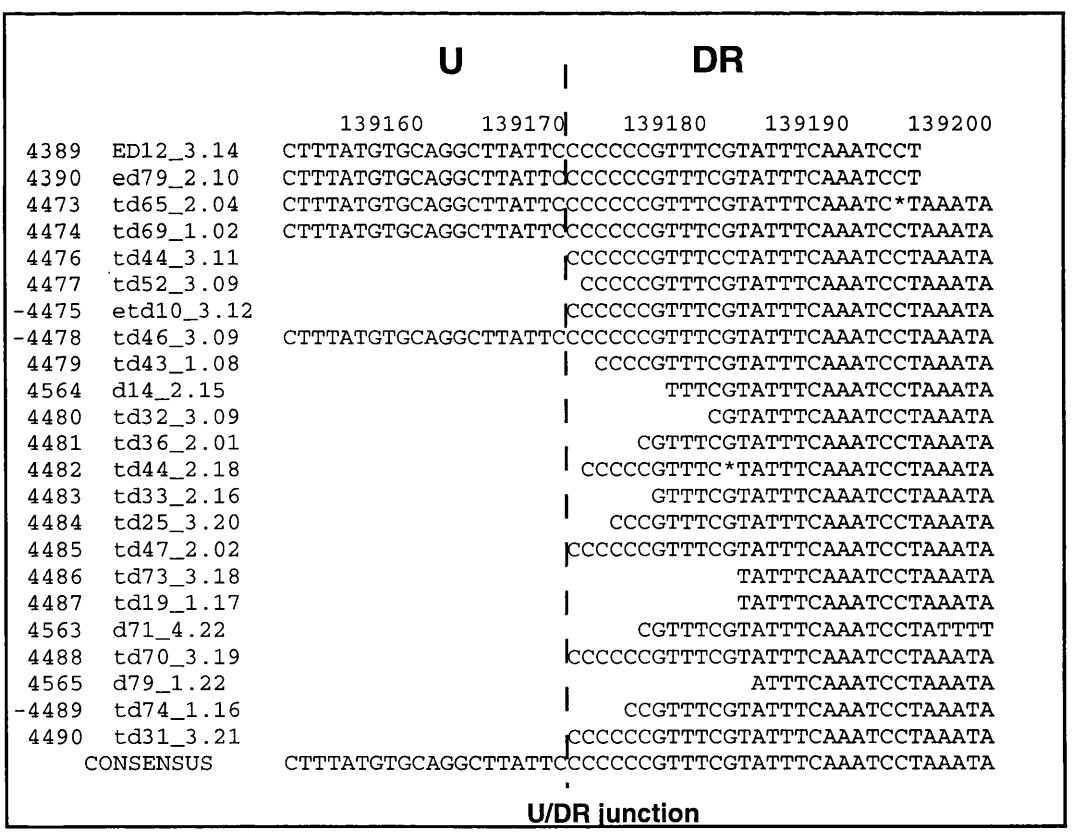


Fig. 21. The DR/U and U/DR junctions, or right and left ends of the genome (Figs. A and B), respectively, as represented in the database.

Sequences from the genome ends begin or terminate with DR, whereas those from within the genome span the DR/U junctions.

U and DR containing 34% and 52% G+C, respectively. The sequence has been deposited with the GenBank Data Library under accession number AF037218.

3.8. GENOME SEQUENCE COMPARISONS

Sequence comparisons were carried out using the HHV-7 RK sequence, the HHV-7 JI sequence (Nicholas, 1996; accession U43400, version 5, 22. May 1997) and the HHV-6 U1102 sequence (Gompels *et al.,* 1995; accession X83413, version 16, 17 February 1997). For certain comparisons, published sequences from other HHV-6 strains were utilised.

3.9. DIFFERENCES BETWEEN THE HHV-7 RK AND JI SEQUENCES

3.9.1. Reiterations

The sequences of HHV-7 RK and JI were compared using the Gap program and were found to differ at various locations throughout their lengths. The differences are greatest between the regions containing reiterated sequences R1, R2, T1 and T2. R1 and R2 are located toward the right end of U, whereas T1 and T2 are located near the beginning and end, respectively, of each DR (Fig. 22).

R1 in the RK sequence comprises three copies of a complex 84 bp element and one partial copy of 68 bp; the reiterations differ at two nucleotides only. The R1 arrays from RK and JI are very similar, as shown in Fig. 16. The first and second 84 bp elements in JI are identical to those in RK, but the third element is 3 bp shorter and contains several mismatches, also the partial copy is shorter (66 bp).

R2 consists of an array of 105 bp elements that differ at several positions. RK contains 15 copies of the element followed by a partial copy of 26 nucleotides, and JI has 17 elements plus the partial copy. The R2 arrays from RK and JI are compared in Fig. 15. Several reiterated elements are identical between the strains and some are positioned analogously, but other elements are unique. The arrangement of elements is similar in both strains at the extremities of the array, but differs in the central region. Interestingly, within each strain the elements closest to the edges of the array are the most imperfect.

T1 and T2 are similar in RK and JI in that they consist largely of the telomeric element TAACCC, but differ in the length of the arrays. T2 is the shorter reiteration, consisting of 148 elements and is displayed in Fig. 17, alongside the analogous reiteration in JI. T1 near the left end of DR consists of 633 elements (Fig. 18). It is likely, however, that the solution for T1 is not a correct representation of the actual sequence in this region. This aspect is taken up in the Discussion (Section 4.1.4.).

Bacterial clones containing T1 and T2 in the JI genome contain 20 and 106 copies of the repeated element, respectively (Secchiero *et al.*, 1995). Both are thus smaller than in RK, the former markedly so. Towards the edges of T2 the telomeric elements are arranged similarly between the strains, but the central portion of the RK array is absent in JI.

3.9.2. Rest of the genome

The two HHV-7 sequences also differ at many locations outside the reiterations. The validity of these differences was assessed carefully in the RK sequence, through reference to the database and the autoradiographs, in addition to resequencing several templates. Differences between the strains where thus identified as genuine differences, or errors in either the RK or JI sequences.

Errors in the RK sequence numbered 35 and are recorded in Table 7 (see Section 3.6.3.). Errors in the JI sequence were defined as deletions or insertions which caused frameshifts in coding regions identified by Nicholas (1996) that resulted in sub-optimal amino acid sequence conservation between HHV-7 and HHV-6. Only four errors were noted, two in ORF U38 (2 bp deletion at 57718 and a 1 bp deletion at 57733) and two in ORF U53 (1 bp deletion at 81361 and a 2 bp deletion, plus a 2 bp substitution at 81380), each causing a local shift from and then back into the correct reading frame (Table 8). It is possible that other differences may also result from errors in the JI sequence, but there was no evidence to support this. Therefore, all other disparities were classified as reflecting genuine strain variation. The types and distribution of the differences are listed in Table 9, and a summary is given in Table 10. Fig. 22 shows the layout of the differences between HHV-7 RK and JI, on the HHV-7 RK genome (the Fig. also includes the

Table. 8. Errors in the DNA sequence of HHV-7 strain JI

Position (bp)		Difference [a]	Sequence change		ORF
HHV-7 RK	HHV-7 JI		RK	JI	
57718	53497-8	ID	TG	--	U38
57733	53512	ID	C	-	U38
81361	77138	ID	A	-	U53
81380-3	77159-60	ID, S	CACG	--TC	U53

[a] ID is insertion or deletion and S is substitution.

Table. 9. Differences between HHV-7 strains RK and JI

Genome location (bp)		Difference[a]	Nucleotide change		ORF[b]	Amino acid change	
RK	JI		RK	JI		JI	RK
4130/147175	184/139231	SS	G	A	non-coding		
4178/147223	232/139279	SS	A	T	non-coding		
4761/147806	815/139862	SS	A	G	non-coding		
5211/148256	1255/140302	NS	G	A	DR2	S	N
5572/148617	1626/140673	SS	C	G	DR2		
5836/148881	1890/140937	SS	G	C	DR2		
5986/149031	2040/141087	SS	G	C	DR2		
6058/149103	2112/141159	ID	C	-	non-coding		
6282/149327	2335/141382	SS	A	G	non-coding		
6335-6/149380-1	2388-9/141435-6	MS	TG	CA	non-coding		
6433/149478	2486/141533	SS	C	G	non-coding		
6626/149671	2679/141726	NS	A	G	DR6	N	D
7227/150272	3280/142327	SS	A	G	DR7		
7876/150921	3929/142976	ID	A	-	non-coding		
7926/150972	3978/143025	ID	A	-	non-coding		
7995/151041	4045/143092	ID	A	-	non-coding		
8618/151664	4668/143715	ID	A	-	non-coding		
8828/151873	4877/143924	SS	T	G	non-coding		
10144	5914	SS	T	C	non-coding		
10397	6177	SS	C	G	non-coding		
10429	6209	SS	C	T	non-coding		
10533	6313	SS	A	G	non-coding		
10604	6384	NS	A	G	U2	F	S
10662	6442	SS	A	G	U2		
10826	6606	NS	T	C	U2	N	S
11314	7094	NS	T	A	U2	E	D
11329	7109	SS	G	A	U2		
11663	7443	ID	A	-	non-coding		
11782-84	7561-63	MS	GGG	AAA	non-coding		
11786	7565	SS	G	A	non-coding		
11877	7656	SS	G	A	U3		
12042	7821	SS	C	T	U3		
13986	9765	SS	A	G	U4		
14532	10311	SS	A	G	U4		
15120	10899	SS	G	A	U5		
16038	11817	SS	A	T	U5		
16941	12720	NS	A	T	U7	C	S
18747	14526	SS	T	A	non-coding		
19157	14937	NS	A	G	U10	E	G
19614	15693	SS	T	C	U10		
20310	16089	NS	C	T	U11	D	N
20893	16672	SS	G	A	U11		
20896	16675	SS	A	G	U11		
20956	16735	SS	A	G	U11		
21513	17292	NS	A	G	U11	Y	H
25119	20898	SS	C	T	U14		
25532	21311	NS	G	A	U14	R	H
26264	22043	SS	T	A	non-coding		
26525	22304	SS	G	A	U15		
28914	24693	SS	G	T	non-coding		
29880	25659	ID	-	A	non-coding		
30669	26449	SS	G	A	U19		
36169	31949	ID	T	-	non-coding		
36863	32642	NS	C	G	U26	M	I
37076	32855	NS	C	G	U26	M	I
41483	37262	NS	C	T	U29	R	K
43070	38849	SS	T	A	U30		
43847	39626	SS	G	A	U30		
44420	40199	SS	T	C	U31		
45038	40817	SS	G	A	U31		
45093	40872	NS	T	A	U31	S	T
45559	41340	NS	C	A	U31	A	D
46100	41879	SS	G	A	U31		
48014	43793	NS	G	A	U31	M	I
48188	43967	SS	A	G	U31		
51648	47427	SS	C	A	U33		
54801-2	50580-1	MS	AT	GC	U37	I	A

55352	51131	SS	C	T	U37		
55388	51167	SS	C	T	U37		
57272	53051	NS	T	C	U38	K	E
57480	53259	SS	A	G	U38		
57531	53310	SS	T	G	U38		
57857	53633	SS	A	G	U38		
58092	53878	SS	G	A	U38		
58889	54675	SS	T	C	U39		
59788	55564	NS	C	T	U39	D	N
59876	55652	SS	G	A	U39		
60539	56315	SS	G	A	U39		
60683	56459	SS	C	T	U39		
61228	57004	NS	T	G	U40	N	T
61732	57508	NS	C	T	U40	G	E
62949	58725	SS	A	G	U40		
64136	59912	SS	A	T	U41		
64163	59939	SS	A	G	U41		
64535	60311	SS	A	G	U41		
66238	62014	NS	G	A	U41	P	S
66970	62746	SS	G	C	non-coding		
67524	63300	SS	A	G	U42		
69466	65242	SS	T	A	U43		
72576	68352	SS	A	G	U45		
72583	68359	SS	G	A	U45		
74828	70604	SS	A	C	non-coding		
76742	72518	NS	T	G	U48	L	F
77170	72946	ID	-	A	non-coding		
78065	73842	NS	G	A	U50	R	H
78889	74666	NS	T	G	U50	L	V
78939	74716	SS	G	A	U50		
79103	74878	NS	T	C	U50	V	A
79152	74929	SS	G	A	U50		
79305	75082	SS	T	A	U50		
79820	75597	SS	G	A	U51		
81079	76856	NS	G	A	U52	L	F
81436	77210	NS	T	C	U53	V	A
82224	77998	NS	A	G	U53	M	V
82744-7	78514	ID	TCTG	-	U54		
84702	80472	NS	A	G	U55A	F	L
84813	80583	NS	C	G	U55A	V	L
84929	80699	NS	C	T	U55A	R	K
85653	81423	SS	G	A	U55B		
87406	83176	SS	C	T	U56		
88182	83952	SS	G	A	U57		
89460	85230	SS	C	T	U57		
89493	85263	SS	T	G	U57		
90090	85860	SS	C	A	U57		
90435	86204	SS	C	T	U57		
90927	86697	SS	A	G	U57		
90996	86766	SS	A	G	U57		
92542	88312	SS	C	T	U58		
92980	88750	NS	A	C	U58	R	S
97463	93233	NS	A	G	U64	I	M
97984	93754	NS	A	C	U64	Y	I
98138	93908	SS	C	G	U64		
98853	94623	NS	T	G	U65	I	M
99922	95692	SS	A	G	U66		
101182	96952	SS	C	T	U67		
101207	96977	NS	T	C	U67	L	S
105406	101176	SS	T	C	U72		
106411	102181	SS	C	T	U73		
107348	103118	NS	C	T	U73	L	F
107938	103708	SS	G	T	U73		
109107	104877	SS	C	T	U74		
111155	106925	SS	G	A	U76		
111725	107495	SS	T	C	U76		
112108	107878	NS	T	C	U76	N	D
112448	108218	SS	A	G	U76		
113231	109001	SS	A	G	U77		
113244	109014	NS	G	A	U77	A	N
113252	109022	SS	G	A	U77		
115353	111122	ID	A	-	non-coding		
120833	116602	NS	A	T	U82	N	K
121143	116912-4	ID	-	TTC	non-coding		

121627	117399	NS	C	T	U84	E	K
121685	117457	NS	A	G	U84	I	T
123077	118849	NS	G	T	U85	S	Y
123239	119011	SS	T	G	non-coding		
123249	119021	SS	A	T	non-coding		
123309	119081-2	ID	- -	TT	non-coding		
125104	120879	SS	C	T	U86		
125200	120975	SS	T	C	U86		
125389	121164	NS	G	T	U86	D	E
130991	126766	SS	T	C	U89		
131119	126892	NS	G	A	U89	P	S
131326	127099	NS	C	T	U89	A	T
134546	130318	ID	A	-	non-coding		
136728	132711-31	ID	-	TATTTA ATTATT ACTGTT TCA	non-coding		
137805	133809	NS	A	G	U95	D	G
140260	136264	SS	G	A	non-coding		
140569	136573	SS	T	C	non-coding		
140943	136947	SS	T	C	non-coding		
141306	137309	ID	A	-	non-coding		
141965	137968	NS	A	G	U100	C	R

[a] ID is insertion or deletion, SS is synonymous substitution, NS is non-synonymous substitution and MS is multiple substitution.
[b] Encoding potential is based on subsequent analysis in this thesis (Fig. 11).

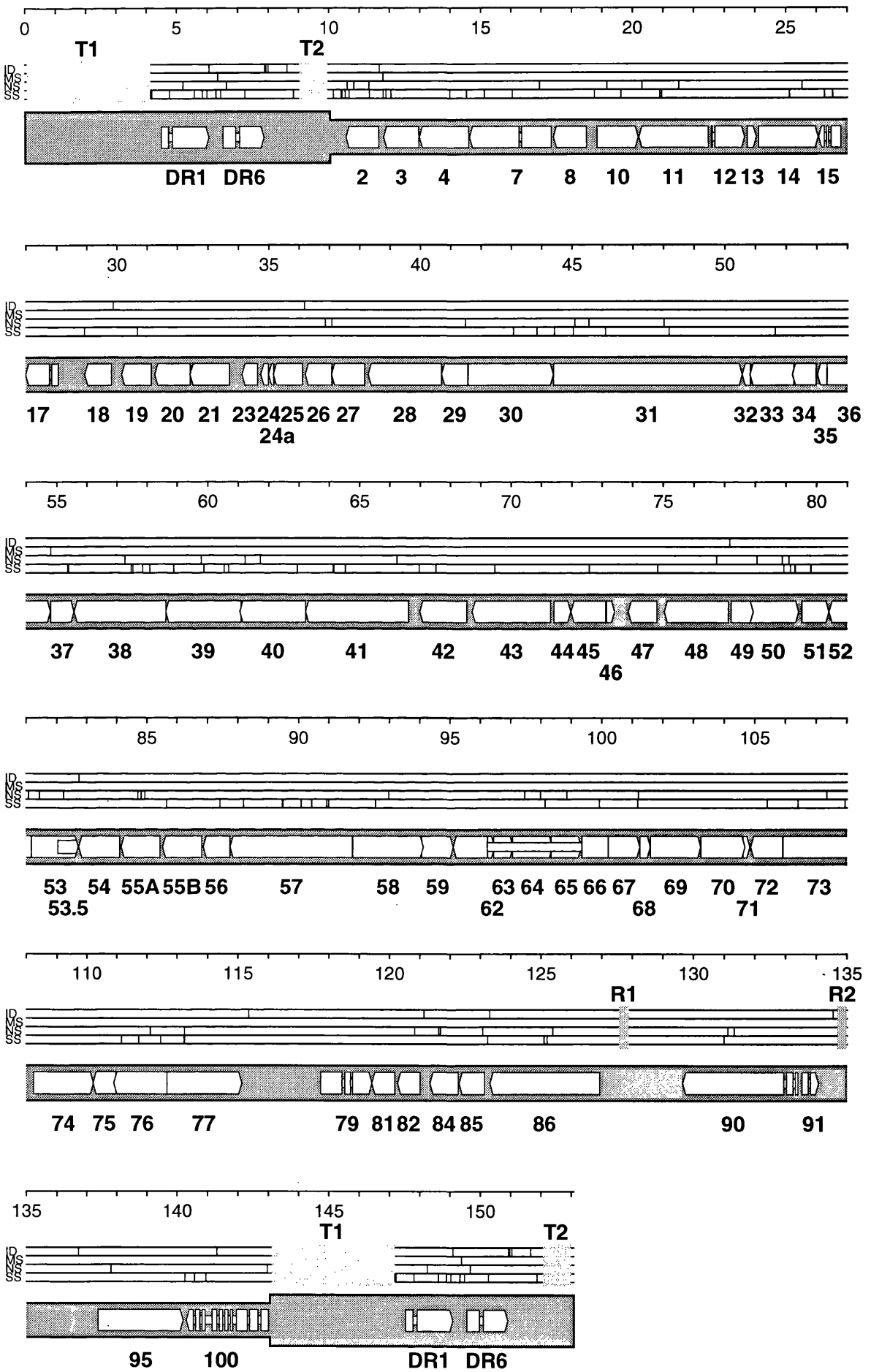
Table 10. Differences between the genome sequences of HHV-7 strains RK and JI (excuding tandem repeats)

type of difference	genome region		
	DR	**U**	**total[a]**
synonymous substitution	10	84	104
nonsynonymous substitution	2	46	50
multiple substitution	1	2	4
insertion/deletion	5	10	20
total	18	142	178

[a]U plus two copies of DR

Fig. 22 Distribution of sequence differences between the HHV-7 RK and JI genomes.

The RK genome is shaded, the thinner and thicker portions denoting U and DR, respectively. The scale is in kbp. Protein-coding regions (DR1, DR6, U2-U100; deduced in Section 4.10.2) are shown as open arrows above the gene nomenclature; the U prefix has been omitted. ORFs predicted to be expressed as spliced mRNAs are connected by open horizontal bars. The differences between the strains are categorised into insertion/deletion (ID), multiple substitution (MS), non-synonymous substitution (NS) or synonymous substitution (SS), and their locations are marked. Differences within the reiterated sequences are not given, and the reiterated sequences (R1, R2, T1 and T2) are shown as shaded rectangles above the genome.


0
5
10
15
20
25
T1
T2
ID
MS
NS
SS
DR1 DR6
2 3 4 7 8 10 11 12 13 14 15
30
35
40
45
50
17 18 19 20 21 23 24 25 26 27 28 29 30 31 32 33 34 36
24a
35
55
60
65
70
75
80
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
85
90
95
100
105
53 54 55A 55B 56 57 58 59 62 63 64 65 66 67 68 69 70 71 72 73
53.5
110
115
120
125
130
135
R1
R2
74 75 76 77 79 81 82 84 85 86 90 91
135
140
145
150
T1
T2
95 100 DR1 DR6

gene layout of HHV-7 which is discussed in Section 3.10.2.). As anticipated, the majority of differences do not affect coding capacity. Moreover, differences that do affect coding potential are scattered throughout the genome, and appear not to be clustered in particular genes. This indicates that no small subset of genes has been subject to unusual evolutionary pressures since divergence of the two strains.

3.10. THE GENETIC CONTENT OF HHV-7 AND HHV-6

3.10.1. Examination of coding potential

Although various criteria may be used to predict the genetic content of a nucleotide sequence, they are to some extent arbitrary. Thus, the results of such an analysis actually take the form of a spectrum of probabilities ranging from candidates with strong coding potential (usually ORFs with counterparts in other genomes), to others that are less certain (usually smaller ORFs lacking counterparts). This may lead to inclusion of false positives and false negatives into the overall picture. Also, specific factors may have a dominant effect on individual candidates regardless of other considerations; independent data on gene expression, for example.

The previous detailed predictions of the coding capacities of the HHV-7 JI and HHV-6 U1102 genomes were used as a starting point for the analysis of the genetic content of the RK sequence. An amino acid sequence counterpart in HHV-6 and a candidate polyadenylation signal close downstream or downstream from a subsequent ORF in the same orientation (i.e. expressed as part of a 3' coterminal family of genes), were the primary criteria for accepting potential HHV-7 genes. ORFs unique to each genome but substantial in size and possessing appropriate polyadenylation signals were then included. Similarities to proteins encoded by other herpesviruses and by other organisms, and potential splicing (Section 3.11.), were also taken into account. Some previously identified ORFs were excluded because they suffer from combinations of features indicating that they are less likely to encode proteins, such as small size, lack of suitable initiation codons, significant overlap with other genes, presence of reiterated sequences, absence of suitable polyadenylation signals and lack of sequence similarity to other proteins. The Codonpreference program was also used to assess whether codon usage patterns would aid identification of protein coding regions. However, codon usage

was found not to be significantly biased to help substantially. A more detailed account of the predicted coding capacities of RK sequence is given in the following subsection.

Throughout the analysis a conservative approach was taken and proposed coding was restricted to strong candidate genes only. However, this approach is biased towards the exclusion of ORFs that are unique to either HHV-6 or HHV-7 (false negatives), rather than the inclusion of ORFs that do not encode proteins (false positives). Nevertheless, each candidate gene was considered carefully in developing the following view of HHV-7 and HHV-6 gene content, both in terms of the exclusion of marginal ORFs and in identification of potentially spliced genes. Thus, this analysis is considered a more accurate representation of the genetic contents of HHV-7 and HHV-6 than were previously available.

3.10.2. The deduced layout of HHV-7 genes

The deduced layout of HHV-7 genes is shown in Fig. 23 and details are listed in Table 11. A loose copy of Fig. 23 is included at the back of this thesis. In addition, a listing of genetic features and an alignment of predicted amino acid sequences with the genome sequence is given in Appendix I. Also, evidence for splicing in several genes is discussed fully in Section 3.11.

The nomenclature for ORFs developed for HHV-6 and extended to HHV-7 JI was employed for RK, except that several ORFs potentially expressed by splicing are identified only by the 5'-proximal ORF. 19 different JI ORFs (Fig. 6, Table 3) have not been included by name in the analysis of RK (Fig. 23, Table 11). DR2, DR7, U5, U16, U17ex, U60, U80, U89, H6, H8, U98 and U99 now form exons in spliced genes (Section 3.11.2.). The remaining HHV-7 JI ORFs lacking counterparts in HHV-6 (H1, H2, H3, H4, H5 and H7), in addition to U17a, are considered unlikely to encode protein. H1 spans reiteration T1 and lacks an initiation codon. H2 is small (79 codons). H7 is also small (82 codons), and spans R2. H5 is 166 codons in length and lacks an initiation codon. Lastly, ORFs H3 and H4 are small (82 and 98 codons, respectively), lack candidate polyadenylation signals and importantly, are disrupted by insertions or deletions in RK. U17a is small (89 codons), and is not convincingly related to a corresponding HHV-6 ORF.

Fig. 23 Predicted layout of HHV-7 RK genes.

The genome is shaded, the thinner and thicker portions denoting the unique region and terminal direct repeats, respectively. The scale is in kbp. Protein-coding regions (DR1, DR6, U2-U100) are shown as open arrows above the gene nomenclature; the U prefix has been omitted. ORFs predicted to be expressed as spliced mRNAs are connected by open horizontal bars. Candidate polyadenylation signals (AATAAA, ATTAAA or AGTAAA) are indicated by vertical arrows in the appropriate strand. Reiterated sequences (R1, R2, T1 and T2) are shown as filled rectangles above the genome, and the minimal fragment shown to act as an origin of DNA replication (ori) is indicated (van Loon *et al.*, 1997).

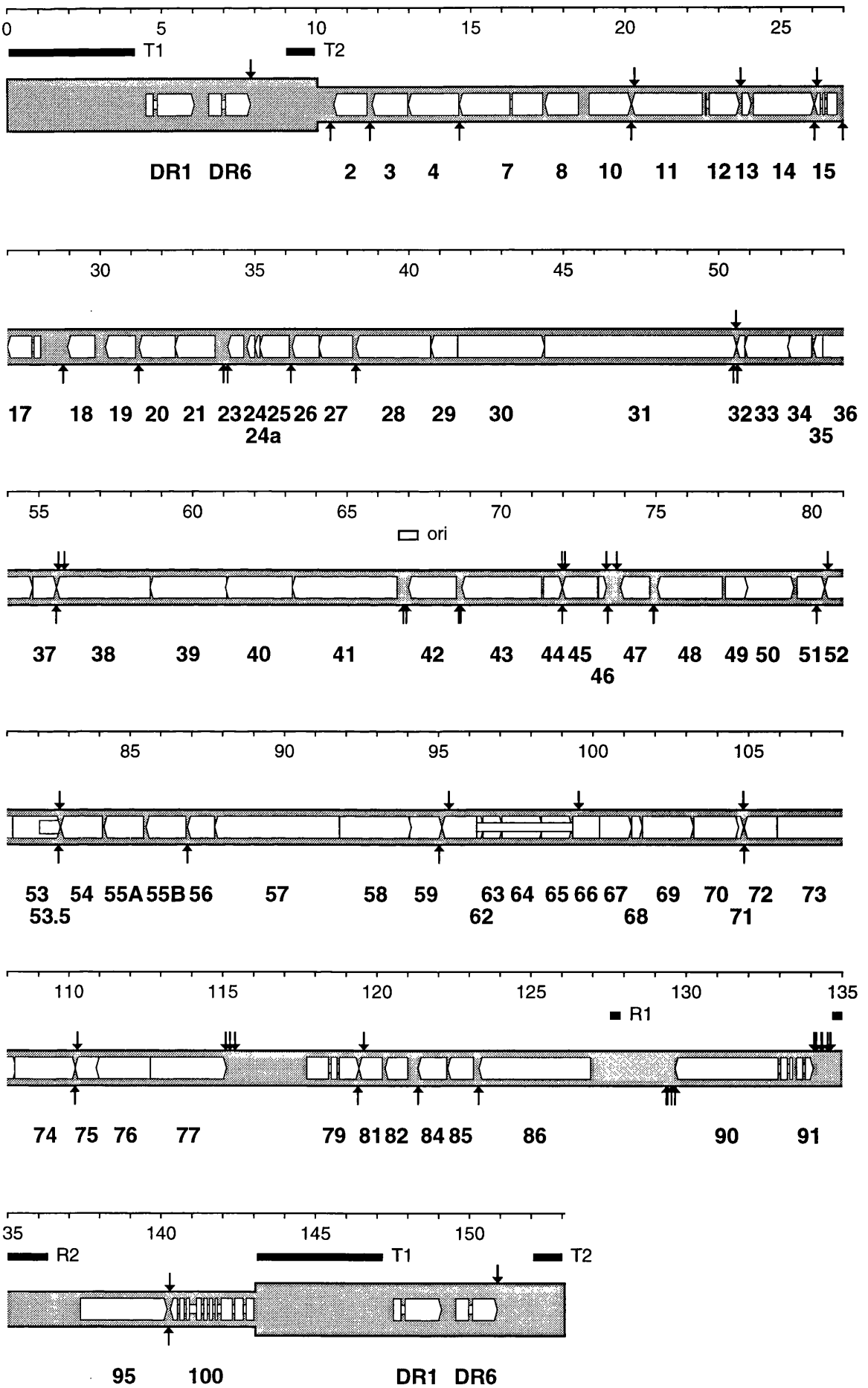

0
5
10
15
20
25
T1
T2
DR1 DR6
2 3 4 7 8 10 11 12 13 14 15
30
35
40
45
50
17 18 19 20 21 23 24 25 26 27 28 29 30 31 32 33 34 36
24a
35
55
60
65
70
75
80
ori
37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
85
90
95
100
105
53 54 55A 55B 56 57 58 59 63 64 65 66 67 69 70 72 73
53.5
62
68
71
110
115
120
125
130
135
R1
74 75 76 77 79 81 82 84 85 86 90 91
135
140
145
150
R2
T1
T2
95
100
DR1 DR6

Table 11. Features of HHV-7 (RK) genes.

GENE[a]		STRAND	START[b]	STOP[c]	CODONS	ID[d]	PROPERTIES[e]
DR1	ex1	+	4485	4728	475	33	HCMV US22 gene family; exon 1 starts at 4466
	ex2		4863	6046			
DR6	ex1	+	6509	6926	406	59	HCMV US22 gene family
	ex2		7055	7857			
U2		-	11637	10558	359	48	HCMV US22 gene family
U3		-	12953	11799	384	49	HCMV US22 gene family
U4		-	14603	12975	542	59	Related to U7 exon 2
U7	ex1	-	17324	16348	>871	54	HCMV US22 gene family (exon 1); related to U4 (exon 2);
	ex2		16266	14628			exon 1 starts at 17326
U8		-	18483	17395	362	51	HCMV US22 gene family
U10		+	18829	20184	451	50	
U11		-	22470	20203	755	31	Structural phosphoprotein
U12	ex1	+	22566	22598	333	47	G protein-coupled receptor
	ex2		22689	23657			
U13		+	23742	24038	98	33	
U14		+	24106	26052	648	50	HCMV UL25/UL35 gene family
U15	ex1	-	26785	26469	191	68	
	ex2		26393	26321			
	ex3		26249	26064			
U17	ex1	-	28057	27841	330	53	HCMV US22 gene family; IE-B[f] transactivator
	ex2		27768	26993			
U18		-	29821	28934	295	44	IE-B membrane glycoprotein
U19		-	31142	30165	325	34	IE-B protein
U20		-	32431	31256	391	22	Probable membrane glycoprotein
U21		-	33714	32422	430	31	Probable membrane glycoprotein
U23		-	34638	34123	171	-	Probable membrane glycoprotein
U24		-	34992	34744	82	28	Contains a hydrophobic domain
U24A		-	35166	34996	56	25	Contains a hydrophobic domain
U25		-	36118	35156	320	47	HCMV US22 gene family
U26		-	37090	36209	293	30	
U27		-	38172	37078	364	68	Processivity subunit of replicative DNA polymerase; [*UL42*]
U28		-	40705	38285	806	47	Ribonucleotide reductase large subunit [UL39]
U29		-	41568	40708	286	53	Capsid protein; component of intercapsomeric triplex [UL38]
U30		+	41583	44399	938	46	Tegument protein [*UL37*]
U31		+	44400	50579	2059	46	Very large tegument protein [UL36]
U32		-	50848	50576	90	66	Capsid protein; located on tips of hexons [*UL35*]
U33		-	52262	50829	477	59	Virion protein
U34		-	52989	52213	258	56	Membrane-associated phosphoprotein [*UL34*]
U35		-	53340	53026	104	58	Role in DNA packaging [UL33]
U36		+	53339	54796	485	58	Role in DNA packaging [UL32]
U37		+	54798	55577	259	62	[UL31]
U38		-	58625	55584	1013	67	Catalytic subunit of replicative DNA polymerase [UL30]
U39		-	61093	58625	822	56	Envelope glycoprotein gB [UL27]
U40		-	63221	61056	721	56	Role in DNA packaging [UL28]
U41		-	66619	63224	1131	68	Single-stranded DNA-binding protein [UL29]
U42		-	68546	66996	516	56	Post-translational regulator of gene expression [UL54]
U43		-	71310	68725	861	61	Component of DNA helicase-primase complex; primase [UL52]
U44		+	71367	71978	203	58	[*UL51*]
U45		-	73122	71983	379	50	[UL50][g]
U46		+	73154	73414	86	52	Membrane protein [UL49A]
U47		-	74803	73862	313	23	
U48		-	77133	75041	690	39	Envelope glycoprotein gH; complexes with gL [UL22]
U49		+	77226	77945	239	52	[UL24]
U50		+	77761	79425	554	55	Role in DNA packaging [UL25]
U51		+	79527	80411	294	35	G protein-coupled receptor
U52		-	81172	80408	254	56	
U53		+	81180	82721	513	52	N-terminal protease domain acts in capsid maturation and is a capsid protein; C-terminal domain is the minor capsid scaffold protein [UL26]
U53.5		+	82029	82721	230	51	Major capsid scaffold protein [UL26.5]
U54		-	84100	82736	454	42	Virion transactivator
U55A		-	85431	84148	427	33	Related to U55B
U55B		-	86807	85515	430	21	Related to U55A
U56		-	87741	86860	293	65	Capsid protein; component of intercapsomeric triplex [UL18]
U57		-	91781	87744	1345	68	Major capsid protein; forms hexons and pentons [UL19]
U58		+	91793	94120	775	61	
U59		+	94068	95111	347	38	
U62		+	96247	96474	75	45	

U63		+	96446	97081	211	68	
U64		+	97059	98378	439	41	Role in DNA packaging; tegument protein [*UL17*]
U65		+	98341	99333	330	59	Tegument protein [UL16]
U66	ex1	-	100215	99352	663	72	Role in DNA packaging; putative terminase [UL15]
	ex2		96235	95108			
U67		+	100214	101254	346	52	[*UL14*]
U68		+	101254	101598	114	48	
U69		+	101601	103241	546	53	Serine-threonine protein kinase; tegument protein [UL13]
U70		+	103243	104685	480	52	Deoxyribonuclease; role in maturation/packaging of DNA [UL12]
U71		+	104622	104843	73	53	Myristylated tegument protein [UL11]
U72		-	105906	104866	346	59	Envelope glycoprotein gM; role in virion envelopment [UL10]
U73		+	105923	108286	787	58	Origin-binding protein; helicase [UL9]
U74		+	108237	110216	659	41	Component of DNA helicase-primase complex [*UL8*]
U75		-	110973	110203	256	45	[UL7]
U76		-	112819	110897	640	59	Minor capsid protein; role in DNA packaging [UL6]
U77		+	112665	115127	820	75	Component of DNA helicase-primase complex; helicase [UL5]
U79	ex1	+	117733	118429	506	42	Probable role in DNA replication
	ex2		118522	118712			
	ex3		118788	119420			
U81		-	120179	119415	254	58	Uracil-DNA glycosylase [UL2]
U82		-	121009	120269	246	38	Envelope glycoprotein gL; complexes with gH [UL1]
U84		-	122271	121339	310	42	
U85		-	123141	122299	280	37	Probable membrane glycoprotein; related to OX-2
U86		-	126934	123317	1205	29	IE-A[h] protein
U90	ex1	-	133408	133323	1199	28	IE-A transactivator; exon 1 starts at 133434
	ex2		133244	133033			
	ex3		132949	129648			
U91	ex1	+	133539	133740	153	27	Probable membrane glycoprotein
	ex2		133817	134076			
U95		+	137378	140201	940	25	HCMV US22 gene family
U100	ex1	-	142997	142753	603	28	Envelope glycoprotein gp105; exon 1 starts at 143000
	ex2		142658	142377			
	ex3		142303	141947			
	ex4		141826	141749			
	ex5		141672	141565			
	ex6		141494	141381			
	ex7		141298	141144			
	ex8		140914	140815			
	ex9		140742	140599			
	ex10		140532	140304			
DR1	ex1	+	147531	147774	475	33	HCMV US22 gene family; exon 1 starts at 147512
	ex2		147909	149092			
DR6	ex1	+	149555	149972	406	59	HCMV US22 gene family
	ex2		150101	150903			

[a] Protein-coding exons (ex) are listed. Genes with counterparts in all mammalian herpesviruses are shaded.
[b] First exons: from first nucleotide of first complete codon (U7) or initiation codon (other ORFs). Subsequent exons: from first nucleotide.
[c] To last nucleotide of stop codon or exon.
[d] Percentage identical amino acid residues to the HHV-6 counterpart as determined by Gap at default values; U23 proteins did not align at these settings.
[e] Properties derived from the current analysis and summaries in Gompels *et al.* (1995), Nicholas (1996) and numerous other herpesvirus genome sequence papers. For genes with counterparts in all mammalian herpesviruses, the HSV-1 nomenclature is given in square parentheses; where identification is based on positional data, the HSV-1 gene is shown in italics.
[f] Immediate early B locus.
[g] Related to HSV-1 UL50 encoding deoxyuridine triphosphatase, but probably lacking enzymatic function.
[h] Immediate early A locus.

HHV-7 RK is thus considered to contain 84 different genes, two of which (DR1 and DR6) are present twice in DR, giving a total of 86 genes in the genome (Table 11, Fig. 23). The weakest candidate, U24a, is small and poorly conserved, but encodes a protein containing a highly hydrophobic domain in both HHV-7 and HHV-6. Indeed, all of the proposed HHV-7 genes but one have clear HHV-6 homologues. The exception, U55B, is a sizeable member of a 3' coterminal family, and has been shown previously to be related to an adjacent member, U55A, which is analogous to HHV-6 U55. In addition, all of the proposed HHV-7 genes have candidate polyadenylation signals (AATAAA, ATTAAA or AGTAAA) near the 3' end of the ORF or ORF family, as shown in Fig. 23.

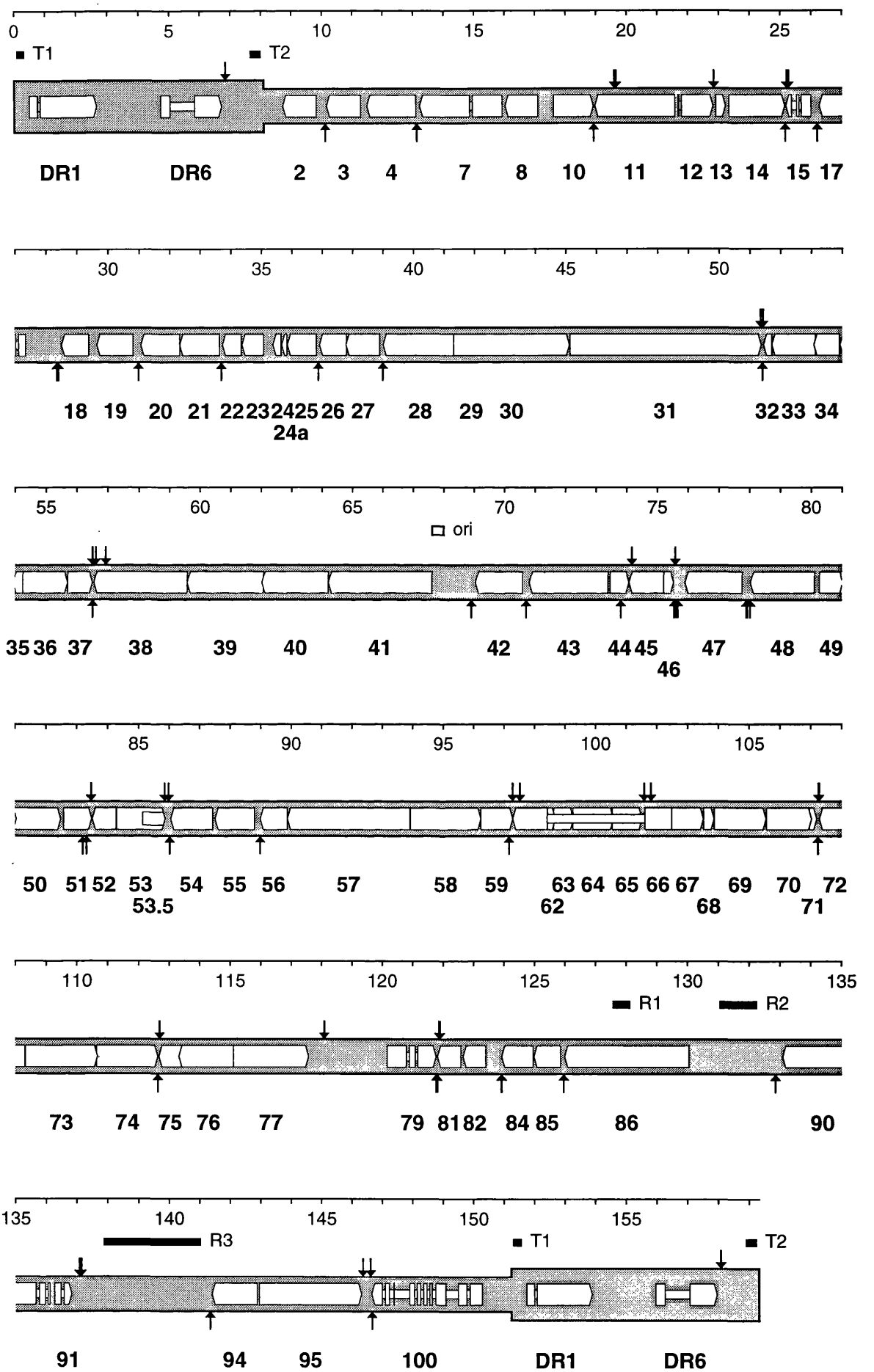
3.10.3. The deduced layout of HHV-6 genes

The criteria used above were applied to the HHV-6 U1102 sequence, and yielded the deduced gene layout, which is given in Fig. 24. The HHV-6 genome is predicted to contain 85 different genes, two of which (DR1 and DR6) are present twice in DR giving a total of 87 genes in the genome. All of the proposed genes have appropriate polyadenylation signals except U2, which has a CATAAA element instead. Of the 22 HHV-6 ORFs identified by Gompels *et al.* (1995) that lack HHV-7 counterparts, only two are strong candidates for encoding complete proteins: U22 a potential membrane glycoprotein and U94 a homologue of the adeno-associated virus type 2 *rep* gene. HHV-6 U87 now forms part of U86 owing to a sequence error by Nicholas (1994) that was corrected by Nicholas (1996). Most ORFs lacking counterparts in HHV-7 (LT1, DR3, DR4, DR5, DR8, LJ1, U1, U6, U9, U61, U78, U83, U88, U92, U93 and RJ1) are considered unlikely to encode proteins. U21ex, U96 and U97 are predicted to form exons in spliced genes, in addition to conserved ORFs DR2, DR7, U5, U16, U17, U60, U79, U89, U98 and U99.

Specific features responsible for excluding ORFs that lack HHV-7 counterparts from the genetic content of HHV-6 are as follows. U9, U61 and U83 (104, 115 and 97 codons, respectively) are small. DR5 (145 codons), DR8 and U78 (110 and 109 codons, respectively) are small and also lack candidate polyadenylation signals. U92 and U93 (147 and 197 codons, respectively), are larger but span R3 and lack suitable initiation codons. U88 is relatively large (413 codons), but spans R2. LT1 (112 codons) and RJ1 (143 codons) span reiteration T1; LT1 also lacks a

Fig. 24 Predicted layout of HHV-6 U1102 genes.

The genome is shaded, the thinner and thicker portions denoting the unique region and terminal direct repeats, respectively. The scale is in kbp. Protein-coding regions (DR1, DR6, U2-U100) are shown as open arrows above the gene nomenclature; the U prefix has been omitted. ORFs predicted to be expressed as spliced mRNAs are connected by open horizontal bars. Candidate polyadenylation signals (AATAAA, ATTAAA or AGTAAA) are indicated by vertical arrows in the appropriate strand. Reiterated sequences (R1, R2, R3, T1 and T2) are shown as filled rectangles above the genome, and the minimal fragment shown to act as an origin of DNA replication (ori) is indicated (Dewhurst *et al.*, 1993).


0
5
10
15
20
25
T1
T2
DR1
DR6
2
3
4
7
8
10
11
12 13
14
15
17
30
35
40
45
50
18
19
20
21
22
23
24
25
26
27
24a
28
29
30
31
32
33
34
55
60
65
70
75
80
ori
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
85
90
95
100
105
50
51
52
53
53.5
54
55
56
57
58
59
62
63
64
65
66
67
68
69
70
71
72
110
115
120
125
130
135
R1
R2
73
74
75
76
77
79
81
82
84
85
86
90
135
140
145
150
155
R3
T1
T2
91
94
95
100
DR1
DR6

candidate polyadenylation signal. LJ1 (321 codons) spans T2 and lacks a candidate polyadenylation site. DR3 (192 codons), DR4 (100 codons) and U6 (103 codons) overlap other genes; also, DR3 overlaps the 3' end of DR1 and lacks a consensus polyadenylation signal.

3.10.4. Gene organisation

About half of the genes in HHV-7 and HHV-6 have counterparts in all sequenced mammalian herpesviruses; these are termed core genes and are highlighted in Table 11. Gompels *et al.* (1995) and Nicholas (1996) pointed out that most of the non-core genes are located towards the genome termini. However, many non-core genes have counterparts in HCMV (Chee *et al.*, 1990) or MCMV (Rawlinson *et al.*, 1996). Several belong to the US22 gene family.

Where possible, amino acid sequence similarity between HHV-7 and HHV-6 was also used as a guide to locating initiation codons and possible splice sites (Section 3.11.1). This resulted in modifications to the locations of some HHV-7 genes from those reported for JI; details are incorporated in Tables 11 and 12. Also, modifications to the HHV-6 U1102 interpretation from comparisons with RK resulted in 5' truncation of ORFs U14, U27, U30, U44, U47, U67 and U76. All but one is reduced in size by approximately 5-10%; U47 is truncated to almost 50% of its original size. These modifications for HHV-6 genes are listed in Table 13; a few were stated or implied in Nicholas (1996), or in HHV-6 sequence papers predating Gompels *et al.* (1995).

3.11. GENE SPLICING IN HHV-7 AND HHV-6

Drawing largely on data from other herpesviruses (especially HCMV) and, in some cases, on experimental data from HHV-6, Gompels *et al.* (1995) and Nicholas (1996) indicated that splicing may occur in certain HHV-7 and HHV-6 genes. However, the detailed situation is unclear. Splice sites were predicted for HHV-6 (Lawrence *et al.*, 1990) and U17, U18 and U20 (Nicholas and Martin, 1994), and splicing patterns were deduced from transcript mapping data for U90, U91 and U100 (Schiewe *et al.*, 1994; Pfieffer *et al.*, 1995). Subsequently, Gompels *et al.* (1995) listed sites only for HHV-6 U12 and U17, with three of the four sites apparently incorrect, as described below. Nicholas (1996) indicated that splicing may occur in certain HHV-7 genes, but predicted splice sites only for U66 and U17. In reassessing this aspect of HHV-7 gene organisation, substantial support was found from sequence comparisons for splicing in

Table 12. Predicted splice sites in the HHV-7 and HHV-6 genomes

RK ORF		JI ORF	HHV-7 acceptor	HHV-7 donor	HHV-6 acceptor	HHV-6 donor
consensus			YYYYYYYYYYYNY**AG**G	MAG**GT**RAGT	YYYYYYYYYYYNY**AG**G	MAG**GT**RAGT
DR1	exon 1	DR1	TCTCTTCTATCAC**AG**A	CAT**GT**AAGC	CTATTCTTACTCT**AG**G	TAT**GT**GAGT
	exon 2	DR2	TTTGCTCTATCGC**AG**G	none	TTTGGTTTCCCTC**AG**G	none
DR6	exon 1	DR6	none	GCG**GT**GAGT	none	GCG**GT**GAGT
	exon 2	DR7	TCTACATCCCGGC**AG**C	none	CGCGTCCCATCAC**AG**C	none
U7	exon 1	U7	AAGTTAATTTTGC**AG**A	TAG**GT**ATGT	TTATGTTTCAAAC**AG**A	CAG**GT**GGGT
	exon 2	U5	GATGTTCTTTTTC**AG**A	none	TTTTTCTTACAAC**AG**G	none
U12	exon 1	-	none	CTG**GT**ATGA	none	CTG**GT**AAGT
	exon 2	U12	AACTTTTTTTCAC**AG**C	none	CAATATCTTTAAT**AG**C	none
U15	exon 1	U15	none	ACG**GT**GAGT	none	AAG**GT**GGGT
	exon 2	-	CTCTTTTATTTTC**AG**G	GCG**GT**AAGA	GTTTTTCTTTTAC**AG**G	GCG**GT**GAGT
	exon 3	-	TTTTTTTTTCTTT**AG**T	none	GTTTTTTTTTTTT**AG**T	none
U17	exon 1	U17	none	TAT**GT**AAGT	none	TGT**GT**AAGT
	exon 2	U16	TTGTTGTTTTCAT**AG**G	none	TTCCTTTTAAACA**AG**G	none
U66	exon 1	U66	none	CAC**GT**AAGT	none	CAC**GT**AAGT
	exon 2	U60	TCATTTTCTTCTC**AG**A	none	TCATTCCCCTCTC**AG**A	none
U79	exon 1	U79	none	AAG**GT**TAGT	none	ATG**GT**AAAT
	exon 2	H6	AACATGTTTTCTT**AG**A	CAG**GT**GGGT	TTGTTGCAATTTC**AG**A	CAG**GT**GGGT
	exon 3	U80	GTTTCTTTCTTTT**AG**G	none	ATCTTTTTATTTT**AG**G	none
U90	exon 1	-	GGTTTGTTATTGT**AG**G	TGA**GT**AGGT	TTCATTGGCTATC**AG**C	AGA**GT**AAGT
	exon 2	U90	TAAATTTTATTAC**AG**A	CAG**GT**ATTT	TATTTATACTTAC**AG**C	CTG**GT**AAGT
	exon 3	U89	TTCTTTAAATTCT**AG**C	none	TCTTTTACATCCT**AG**C	none
U91	exon 1	U91	none	CAG**GT**TTGT	none	CTG**GT**TAGT
	exon 2	-	TATTTTTTCTTGT**AG**A	none	ATGGTTTGTTTTT**AG**A	none
U100	exon 1	U100	AAAATCTCTTCGC**AG**A	ACA**GT**AAGT	CGAAATTTTCACA**AG**A	ACG**GT**AAGG
	exon 2	U99	TTTAATTCTTCTA**AG**G	ATG**GT**AAGC	TTTAATTTATCGC**AG**C	ATG**GT**GAGT
	exon 3	U98	GTACCCGCTTATT**AG**T	AGT**GT**AAGT	ATTTATTCACTCA**AG**T	CGT**GT**AAGT
	exon 4	-	TATTTTTTTTTTT**AG**A	AAT**GT**AAGA	TTCGTTTTTTGTC**AG**G	ACT**GT**AAGT
	exon 5	-	AATTGTGTTTCGC**AG**T	CAG**GT**AAAT	TTATGTTTCTAAC**AG**A	ACG**GT**GAGT
	exon 6	-	GCTTCTTCATCCT**AG**A	TTG**GT**AATT	TCCGGTTATGCAC**AG**C	ATG**GT**GAGC
	exon 7	-	TTTTTTTCATACC**AG**C	ACA**GT**GGAA	TTTCTTAATTTGC**AG**C	AGG**GT**GGGC
	exon 8	-	TTTTTTTAATTCT**AG**C	CAT**GT**GAGT	TTTCGACCTGCCT**AG**A[a]	AAT**GT**AAGT
	exon 9	-	ATTTCTCGTTCGC**AG**C	CAG**GT**GAGC	TTGATATTTGTTC**AG**T	TAG**GT**ATTA
	exon 10	H8	CATTTTCTCTTTT**AG**T	none	TGTTTTTTTTTTA**AG**T	none

[a] Another acceptor site (GCTACCGCTTTTT**AG**C) is located further upstream. It could extend coding similarity between HHV-7 and HHV-6 exon 8, but is in an inappropriate reading frame.

Table 13. Proposed modifications to HHV-6A(U1102) genes.

GENE[a]		STRAND	START[b]	STOP[c]	CODONS	REASON FOR CHANGE
DR1	ex1	+	501	759	689	Putative splicing; exon 1 starts at 500
	ex2		843	2653		
DR6	ex1	+	4725	5028	395	Putative splicing
	ex2		5837	6720		
U7	ex1	-	15921	14948	>872	Putative splicing; exon 1 starts at 15923
	ex2		14858	13214		
U12	ex1	+	21680	21712	351	Correction of splice sites
	ex2		21790	22812		
U14		+	23331	25145	604	5' truncation
U15	ex1	-	25992	25676	191	Putative splicing
	ex2		25602	25530		
	ex3		25364	25179		
U17	ex1	-	27349	27121	334	Definition of splice sites
	ex2		27034	26259		
U24A		-	35847	35674	57	Not previously identified
U27		-	38903	37797	368	5' truncation
U30		+	42325	45132	935	5' truncation
U44		+	73470	74087	205	5' truncation
U47		-	77768	75912	618	5' truncation
U53.5		+	85133	85867	244	Definition of coding region
U66	ex1	-	102486	101614	666	Definition of splice sites
	ex2		98415	97288		
U67		+	102485	103519	344	5' truncation
U76		-	115257	113317	646	5' truncation
U79	ex1	+	120164	120794	474	Definition of splice sites
	ex2		120891	121087		
	ex3		121170	121766		
U86		-	130044	125989	1351	Error correction
U90	ex1	-	136112	136054	941	Definition of splice sites; exon 1 starts at 136150
	ex2		135965	135772		
	ex3		135664	133092		
U91	ex1	+	136267	136477	153	Definition of splice sites
	ex2		136580	136830		
U100	ex1	-	150282	149873	656	Definition of splice sites; exon 1 starts at 150295
	ex2		149771	149490		
	ex3		149081	148746		
	ex4		148628	148551		
	ex5		148454	148347		
	ex6		148255	148142		
	ex7		148055	147895		
	ex8		147383	147374		
	ex9		147223	147095		
	ex10		146984	146642		
DR1	ex1	+	151735	151993	689	Putative splicing; exon 1 starts at 151734
	ex2		152077	153887		
DR6	ex1	+	155959	156262	395	Putative splicing
	ex2		157071	157954		

[a] Protein-coding exons (ex) are listed. Coordinates include an extra G residue at 128132 (in U86) as indicated by Nicholas (1996).
[b] First exons: from first nucleotide of first complete codon (U7) or initiation codon (other ORFs). Subsequent exons: from first nucleotide.
[c] To last nucleotide of stop codon or exon.

nine genes in addition to U66 and U17 (U7, U15, DR1, DR6, U12, U79, U90, U91 and U100), and splicing patterns were predicted.

3.11.1. Detection of splice sites

Regions of the RK genome in which splicing had been demonstrated experimentally in HHV-6, regions in which splicing was suspected, and regions lacking obvious coding potential, were examined for possible splicing. Selected regions of the RK genome and analogous sequences in HHV-6 were initially assessed for amino acid sequence similarity outwith the borders of previously predicted genes. In order to do this, conceptual translations of the HHV-6 and HHV-7 sequences were created (in all six reading frames) using the Ptrans program. The amino acid sequences were then examined by eye for similarity. In regions where similarity extended upstream from the proposed initiation codon or did not extend to the stop codon, or where local similarity was detected outwith established ORFs, sequences were analysed for the presence of consensus splice donor and acceptor sites (Krainer and Maniatis, 1988).

Splicing was proposed when candidate sites were located similarly in the HHV-6 and HHV-7 genomes, so that two exons could in principle be joined in the same register and express the conserved, and not the non-conserved, polypeptide regions. Table 12 summarises splice sites proposed from the analysis for the two genomes, and lists the ORFs previously recognised in HHV-7 JI that are proposed to form exons in spliced genes.

3.11.2. Candidate spliced genes

U66 - U66 is the only putatively spliced gene which has counterparts in all sequenced herpesviruses, and splice sites have been predicted previously for HHV-6 and HHV-7. Splice sites for this gene are located identically in all herpesviruses, and those in HSV-1 have been mapped experimentally (Dolan *et al.*, 1991). U66 is thus used as a benchmark for other HHV-7 spliced genes discussed below. Fig. 25 shows that significant amino acid sequence similarity between the exons in HHV-6 and HHV-7 does not extend beyond the proposed splice sites into the intron, that the splice sites are in the same register in HHV-7 and HHV-6, and that the exons are joined appropriately.

U66 internal splice

HHV-7

H E T I K S I A L F A S C Y N I H v s n k f f l p t y y a f t d w t
ACGAAACAATAAAAAGTACAGCACTGTTTGCTAGTTGTTACAACACT*CACgtaagt*aacaaatttttcttacctacttactatgcgtttacagactggac

g k e y - - y y i k t q s I y r I k l r s f s s q S I R G Q S
aggaaaggagtattaa.....tagtattacataaaaactcagtctttatatcgtttaaaactccgt *tcattttcttctcagA*GTATACGTGGTCAGAC

HHV-6

H Q N I K S I A L F A S C Y N I H v s t i n l t l f s e i k n -
ATCAAAATATAAAGAGCACAGCACTTTTCGCCAGCTGCTACAATACA*CACgtaagt*actatcaatttaactttatttttcagaaataaaaaattaa.....

- k g r l s f h s s q n r I l s r I t p n s f p s q S I R G Q S
.....taaaaaggtcgtctttcgtttcacagttctcaaaatcgtcttttatcccgtttaacgccgaat *tcattcccctctcagA*GTATACGCGGACAGAC

Fig. 25.

Sequence-derived evidence for splicing in the coding regions gene U66 (HHV-7 and HHV-6). Amino acid sequences extending to the stop codon (hyphen) defining the 3' end of the upstream ORF and from the stop codon defining the 5' start of the downstream ORF are shown. The dots indicate a sizeable region of the intron that is not shown. For this Fig. and Figs. 26-34 relevant DNA sequences are shown with appropriate conceptual translation products aligned from left to right. The DNA sequences begin and end at arbitrary points within the exons, except where stated otherwise. Conserved amino acid residues are underlined. Proposed intron sequences (flanked by splice donor and acceptor sites as listed in Table 12) are shown in lower case in the DNA sequence. Amino acid residues excluded from the protein predicted to be expressed *via* splicing are also shown in lower case. Consensus splice donor and acceptor sites are shown in italics, and previously identified initiation codons which are now thought to map internally are doubly underlined.

U7 - U7 was not previously suspected of being spliced. In HHV-6, U5 and U7 were regarded as substantial ORFs in different reading frames. In HHV-7 JI, U5 and U7 are in the same frame and are thus represented by a single, apparently fused ORF (U5/7). However, the data in Fig. 26 suggest that ORFs U7 and U5 form two exons of a single gene in both genomes (U7 in Table 12). Thus in this model, the spliced U7 gene is flanked at its 5' end by U8 which, like exon 1, contains US22 motifs, and at its 3' end by U4, which is related to exon 2 (Fig. 23). There is also a somewhat weaker indication that U7 may be spliced at its 5' end, since amino acid sequence similarity between HHV-7 and HHV-6 extends upstream from the proposed initiation codon. However, an upstream coding exon but was not identified. U8 is a possibility, but this arrangement would result in a protein with the unusual structure of two contiguous US22 domains.

U15 - U15 was not previously suspected of being spliced. Fig. 27 strongly supports the presence of three exons. Amino acid sequence conservation between HHV-6 and HHV-7 is high, and splice donor and acceptor sites are located identically in the two genomes. The ORFs previously identified as U15 in HHV-6 and HHV-7 JI contain exon 1 only.

DR1 and DR6 - Despite the lack of direct experimental data, splicing events were previously proposed for DR1, DR6, U12, U17 and U79. Nicholas (1996) reported that DR1 is related to DR6 and DR2 to DR7, and that each ORF contains motifs characteristic of the US22 gene family. He also noted that HHV-7 DR1 and DR2 lack initiation codons, in contrast to their HHV-6 counterparts, speculating that these might be supplied by splicing, and commented that splicing of DR1 to DR2 and DR6 to DR7 would result in proteins containing the US22 motifs in their usual order. Fig. 28 supports the view that DR1 and DR2 form exons in a spliced gene (Table 12). Similarly DR6 and DR7 are likely to be spliced (Table 12 and Fig. 29). Although DR1 can supply its own initiation codon it also contains a splice acceptor site at its 5' end, and thus may be spliced to an unidentified upstream non-coding region. In contrast DR6 does not have a splice acceptor site at its 5' end and is unlikely to be spliced upstream.

U12 - In the analysis of the HHV-6 genome, U12 was referred to as a spliced gene but the predicted donor and acceptor sites appear to be incorrect. Splicing in HHV-7 JI U12 was not

U7 internal splice

HHV-7

```
 E  N  L  N  R  M  L  N  G  E  S  P  I  L  R  K  K  P  R  H  M  Y  P  R  C  D  R  y  v  k  i  k  v
GAAAATTTGAACAGAATGTTAAATGGAGAGTCTCCGATTCTTCGGAAGAAACCACGGCATATGTATCCAAGGTGTGA TAGgtatgtaaagatcaaagttc

 I  I  f  g  i  I  y  t  i  I  I  I  m  v  p  w  m  f  f  f  r  L  L  K  N  M  P  S  I  L  F  A  V  I
ttcttttcggtattctctacacaatactcttactgatggttccatg gatgttctttttcagATTATTGAAGAACATGCCAAGTATTCTGTTTGCAGTGCF

 S  S  E  I  S  N  P  L  V  Q  S  V  I  K  F  L  H  P  I  I  I  P  N  G  D  T  E  L  K  Y  I  V  P
CTCATCTGAAATCAGTAATCCTCTTGTTCAAAGTGTTACAAAGTTTCTACATCCAATCATCATTCCAAACGGAGACACGGAACTGAAGTACATTGTTCCC

 V  T  E  S  R  L  I  N  G  L  Q  A  S  A  A  G  R  F  G  I  K  G  L  R  L  C  S  D  G  V  I  W  N
GTGACAGAATCCAGACTAATCAACGGCCTTCAAGCATCAGCTGCTGGACGCTTTGGAATAAAAGGGCTAAGATTATGTTCAGATGGTGTTATTTGGAATF

 R  L  I  D  Y  E  Y  E  M  F  K  Y  P  S  I  F  T  R  A  D  K  F  L  L  Q  L  R  D  L  K  F
GGTTGATAGACTATGAGTACGAAATGTTCAAGTATCCATCAACTTTCACAAGGGCTGATAAATTTCTTTTGCAGTTGCGCGATTTAAAATTT
```

HHV-6

```
                                                                  -  a  s  a  y  v  r  a  k  g  q  v  g  g  c  v  l  c
 E  N  L  N  R  M  L  N  G  E  L  P  V  L  R  S  K  P  R  H  M  C  V  R  K  D  R  w  v  d  v  f  c
GAGAATCTCAATAGGATGTTGAATGGGGAGTTGCCCGTTCTCCGTAGTAAGCCTCGGCATATGTGCGTGCGAAAGGA CAGgtgggtggatgtgttctgtg

 r  f  p  e  r  i  i  i  r  g  w  r  I  p  a  n  f  f  f  s  y  n  r  L  V  K  D  R  S  K  I  L  F
 v  v  f  q  r  g  -
tcgttttccagagaggataattattcgaggatggagactccccgctaacttttttttttcttacaacagGCTTGTGAAAGACCGTAGCAAAATTCTGTTT

 A  V  R  L  D  E  E  D  S  P  T  V  K  F  I  I  K  E  L  T  P  V  F  V  G  R  L  P  A  T  N  R  F
GCGGTGCGTCTGGATGAAGAGGATTCACCGACGGTTAAATTTATCACAAAATTTCTCACGCCGGTATTCGTCGGTCGACTGCCCGCTACCAACAGGTTTC

V  V  P  V  S  R  A  R  L  T  N  G  L  Q  G  T  A  A  A  R  F  G  I  K  G  L  H  P  S  S  D  C  L  L
TTGTTCCCGTCTCCCGCGCCAGGTTGACGAACGGTCTGCAGGGGACCGCAGCAGCTAGATTTGGCATTAAAGGACTACATCCCTCTTCGGACTGCTTGGT

 W  N  I  L  V  D  Y  E  Y  E  T  Y  K  Y  P  S  I  Y  I  R  A  D  Q  I  A  D  M  V  K  D  L  K  F
GTGGAACATACTGGTGGATTATGAGTATGAAACTTACAAGTATCCTTCCACCTACATCAGAGCCGATCAGATTGCGGACATGGTGAAAGACTTAAAATTT
```

U7 upstream splice

HHV-7

```
 -  s  q  v  n  f  a  e  I  V  P  V  C  N  D  V  C  I  E  P  T  P  D  L  P  S  F  D  F  A  V  E  L
taatcacaagttaattttgcagAAATTGTTCCTGTGTGCAATGATGTTTGCATTGAACCGACTCCGGATTTACCGTCTTTTGATTTTGCCGTTGAGTTGC

 L  L  S  S  Y  G  E  G  M  E  I  V  R  N  G  I  K  C  C  L  A  W  P  P  N  Y  V  L  I  F  G  E  F  Y
TGCTGTCTTCATATGGAGAAGGCATGGAGATAGTTCGCAATGGTATAAAGTGCTGTCTTGCTTGGCCGCCAAACTATGTGCTTATTTTTGGCGAATTTTF

```

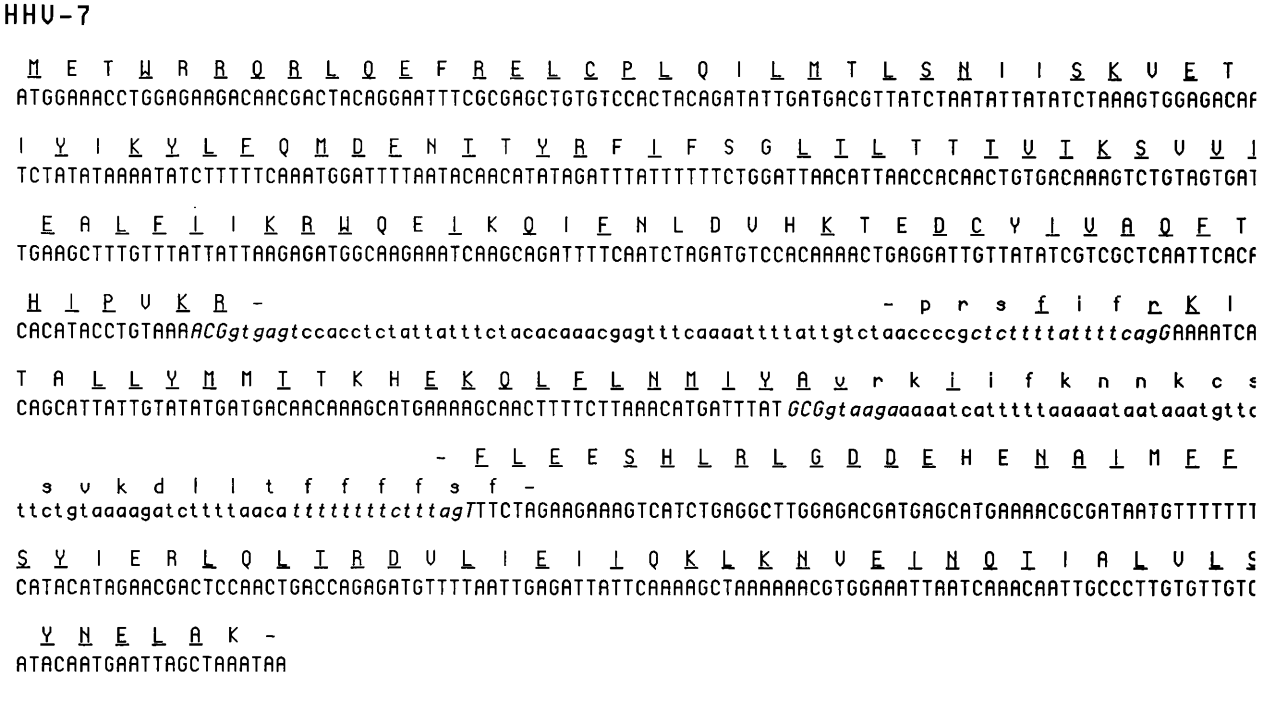
HHV-6

```
 -  I  i  l  e  t  e  g  g  e  r  v  l  c  e  s  i  l  l  y  s  v  f  m  f  q  t  e  I  V  P  V  C
tgactgatacttgaaacggagggcggagagcgagtcttgtgtgaaagtatttttactttattctgtgt ttatgtttcaaacagAGATCGTACCGGTCTGCA

 N  D  E  F  A  L  S  S  C  V  P  T  L  D  F  D  V  D  V  L  S  A  A  Y  G  D  G  L  E  I  S  S  P  L
ACGACGAGTTCGCCTTGTCATCATGTGTGCCTACTTTGGATTTTGATGTGGACGTTCTGTCTGCAGCTTACGGAGACGGGTTGGAAATCAGTTCTCCGGC

 L  R  C  C  I  A  W  P  P  M  Y  A  L  T  L  G  E  F  Y
CTTGCGATGTTGTATCGCCTGGCCTCCCATGTATGCATTAACCCTGGGTGAGTTTTA
```

Fig. 26.
Sequence-derived evidence for splicing U7. Details of the layout are the same as in Fig. 25.

U15 internal splices

HHV-7



HHV-6

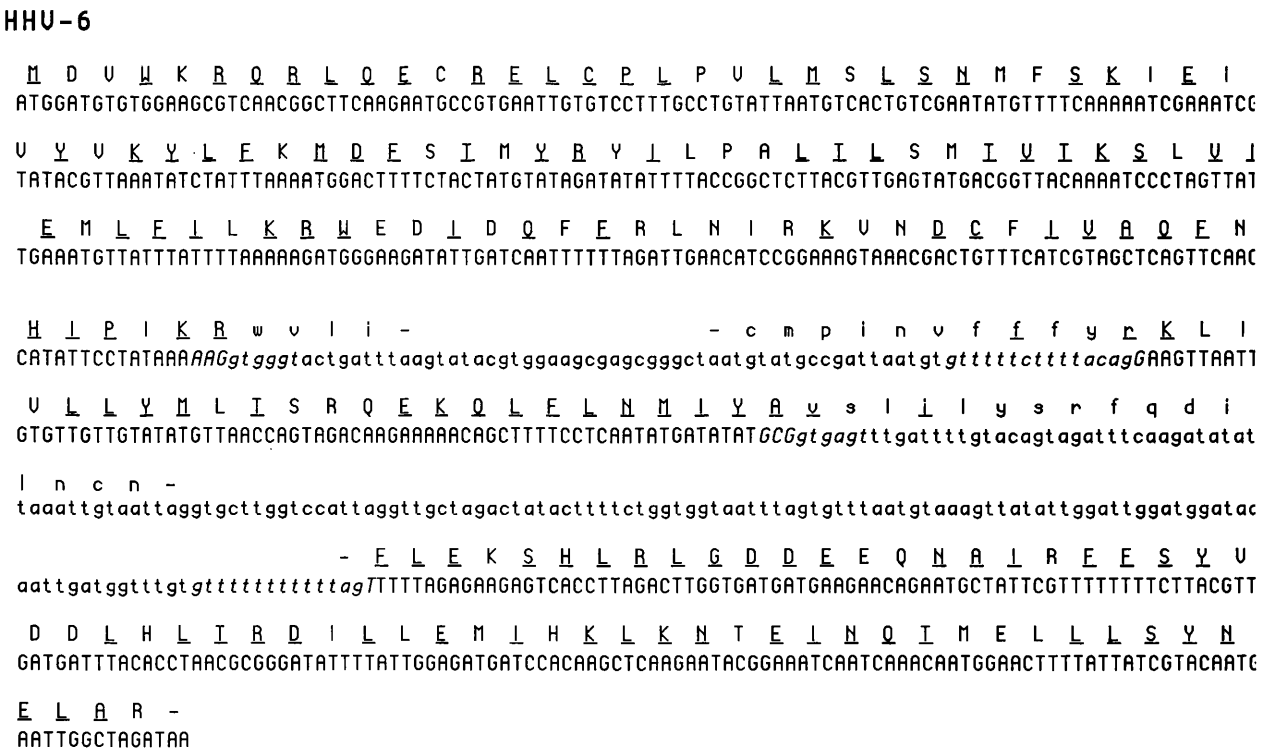
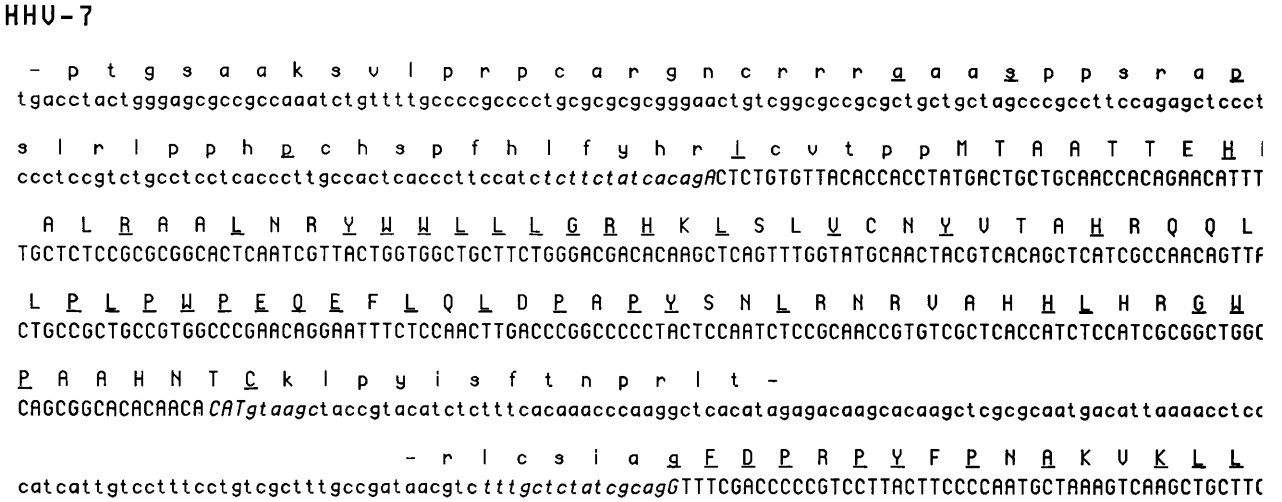


Fig. 27.
Sequence-derived evidence for splicing U15. The whole putative U15 gene is shown, the DNA sequences extend from the proposed initiation codon (ATG) in exon 1 to the stop codon (TAA) in 3. Details of the layout are the same as in Fig. 25.

DR1 upstream and internal splices

HHV-7



HHV-6

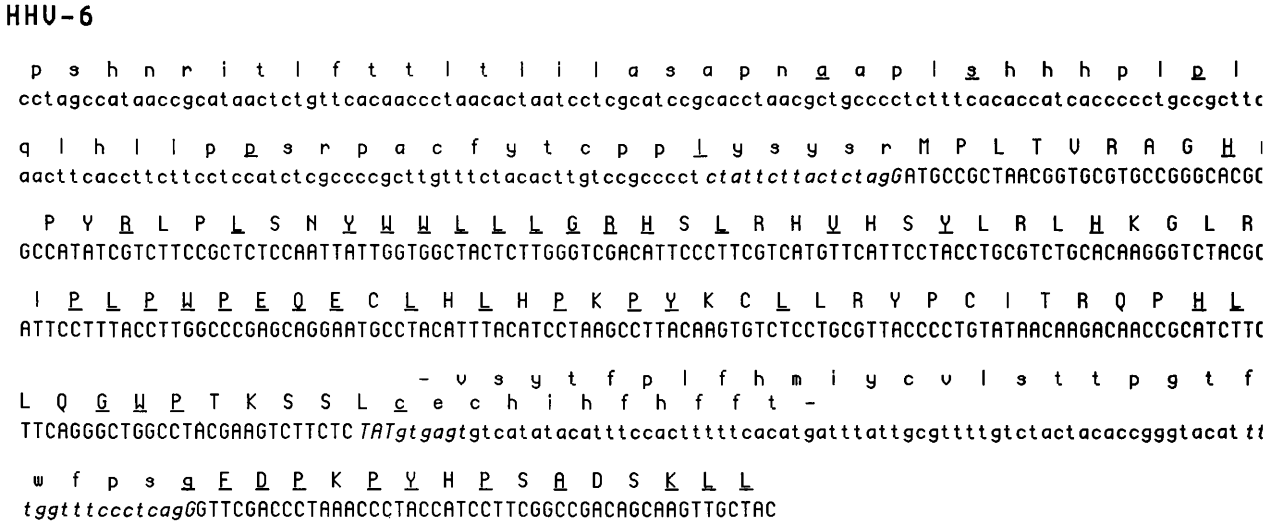
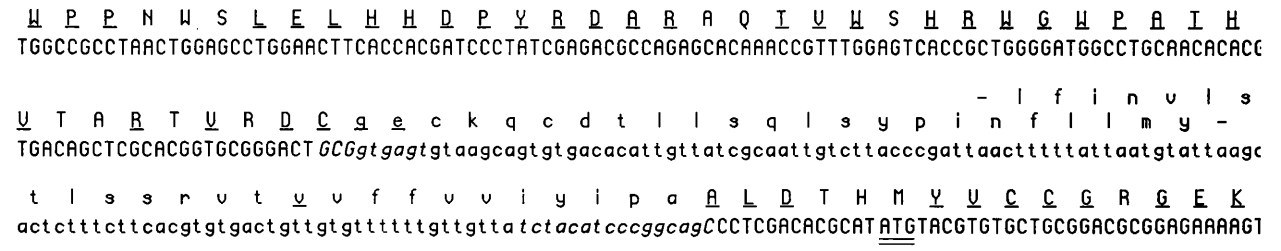


Fig. 28.
Sequence-derived evidence for splicing DR1. The 5' end of DR1 supplies its own initiation signal, and also contains a splicing site to an upstream non-coding region. Details of the layout are the san as in Fig. 25.

DR6 internal splice

HHV-7



HHV-6

Fig. 29.
Sequence-derived evidence for splicing DR6. Details of the layout are the same as in Fig. 25.

mentioned. Evidence that U12 contains two exons (U12 is the second exon) is given in Fig. 30, and the locations of the acceptor and donor sites are recorded in Table 12.

U17 - U17 was predicted previously to be a spliced gene with two exons (U16 and U17) in HHV-6 and HHV-7 JI. The analysis of HHV-7 RK supports alignments made for splice sites in HHV-7 JI and for the acceptor, but not the donor in HHV-6. Evidence for splicing in U17 is shown in Fig. 31.

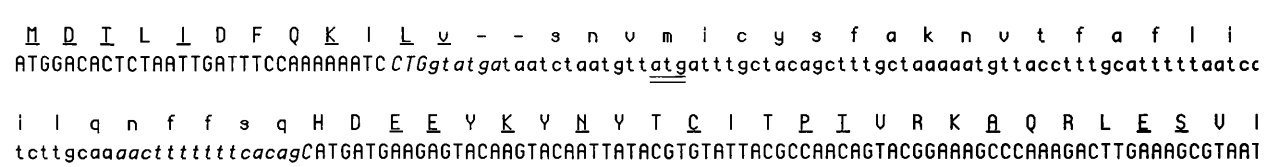
U79 - U79 comprises three exons in HHV-7 RK (Fig. 32). This spliced gene was previously represented as U79 (exon 1), H6 (exon 2), and U80 (exon 3) in HHV-7 JI and as U79 (exons 1 and 2) and U80 (exon 3) in HHV-6. However, splice donor and acceptor site locations were not established in these analyses.

U90 - Schiewe *et al.* (1994) reported from mRNA transcript mapping that U90 is spliced in HHV-6. Although splicing was mentioned for these ORFs in the HHV-7 JI and HHV-6 sequence analyses, no details were given. As can be seen from the analysis of HHV-7 RK in Fig. 33 and Table 12, U90 comprises three exons: exon 1 was previously unidentified; and exons 2 and 3 correspond to U90 and U89 in the HHV-6 and HHV-7 JI sequences, respectively.

U91 - Schiewe *et al.* (1994) also reported that the U91 gene contains two exons. Exon 1 was noted in the HHV-7 JI sequence and exon 2 in the HHV-6 sequence, each recorded as U91. In both cases the other exon was not recognised. Evidence of splicing in U91 is given in Fig. 34. The putative HHV-6 and HHV-7 proteins contain two pronounced hydrophobic regions (one near the N terminus and the other about two-thirds through the protein), as illustrated in Fig. 35. This indicates that the U91 protein is likely to be associated with membranes. In addition, both proteins contain a single consensus site for N-linked glycosylation in the region between the hydrophobic domains, and the HHV-7 protein has two additional sites near the C terminus. Thus, U91 may encode a previously unrecognised membrane glycoprotein.

U12 internal splice

HHV-7



HHV-6

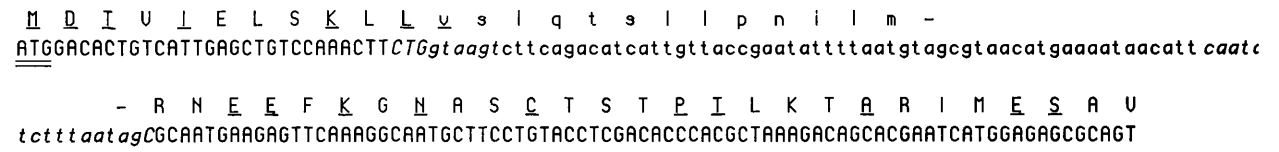


Fig. 30.

Sequence-derived evidence for splicing U12; DNA sequences extend from the proposed initiation c (ATG) in exon 1. Details of the layout are the same as in Fig. 25.

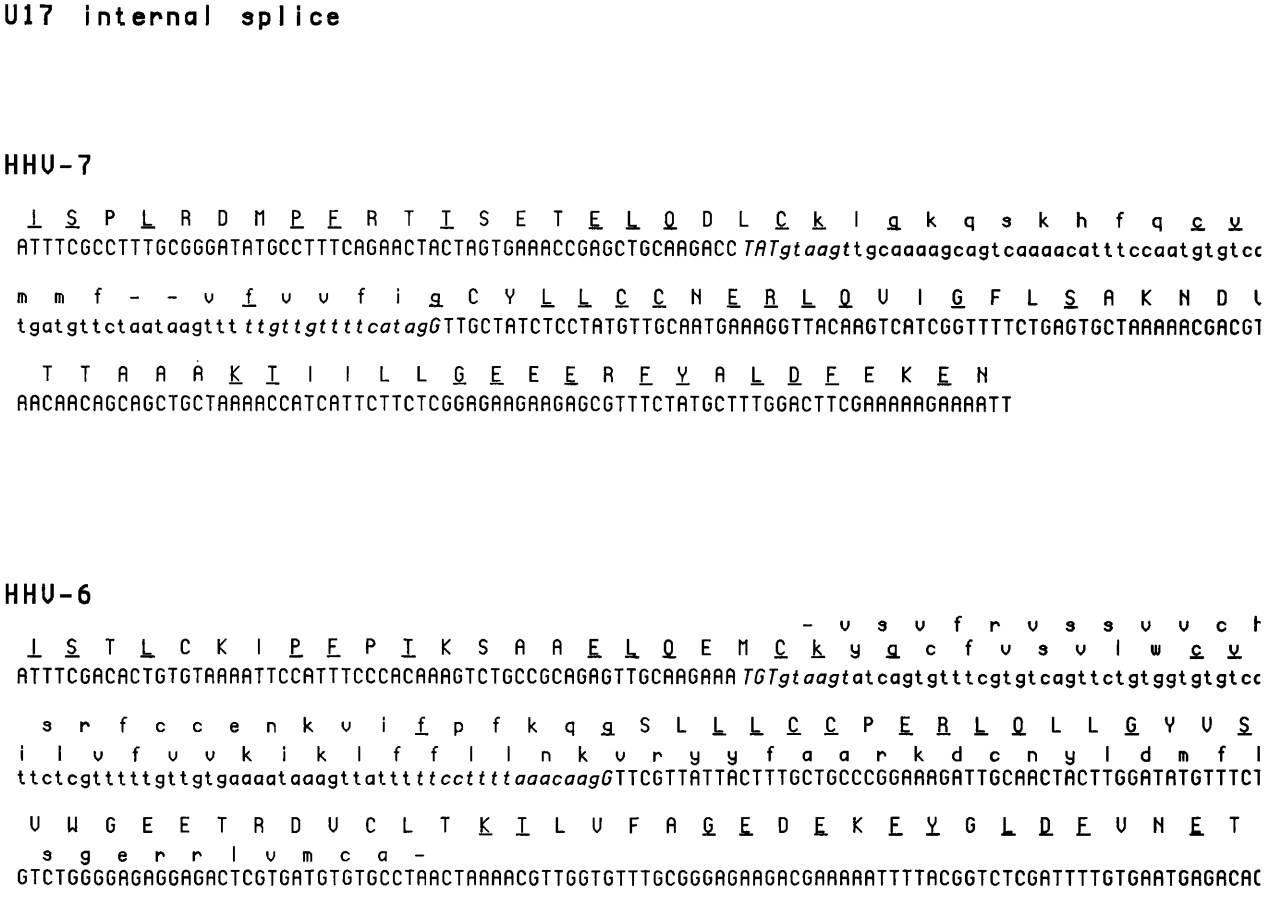


Fig. 31.
Sequence-derived evidence for splicing U17. Details of the layout are the same as in Fig. 25.

U79 internal splice

HHV-7

```
 E  D  D  R  R  N  E  K  Q  D  L  E  q  -                                                        -  c  l
GAAGATGACCGTCGAAACGAAAAACAAGACCTCG AAGgttagttgccttttttctaaatcactagtcctaaatatgacttgttgaatttttgtagtgtct

 r  f  l  h  k  h  v  f  l  D  A  S  K  E  K  R  M  K  V  H  H  E  K  R  H  A  E  E  Q  A  N  E  E
aaggtttttacataaacatgttttcttagATGCATCCAAAGAAAAAAGGATGAAAGTGCATCACGAAAAACGTCATGCGGAAGAACAAGCAAACGAAGAG

 V  A  S  S  S  Q  L  S  S  R  I  P  E  G  A  L  S  P  T  I  S  I  D  L  Q  E  Y  Q  E  F  E  D  F
GTTGCTTCTTCGAGTCAGTTATCAAGTAGAATACCAGAGGGTGCGTTATCGCCCACTATTTCTATTGATCTTCAGGAATATCAAGAATTTGAGGATTTTC

 D  K  R  I  C  G  Q  v  q  q  v  l  g  l  -                                                  -  G  l
ACAAGCGCATTTGTGGGCAGgtgggtggagttttgggtttatgatgcatttttattgcagtttttatgtggttaaagatgtttctttctttttagGGGAA

 N  Q  D  A  V  C  K  K  V  Q  S  D  E  S  F  C  I  N  K  P  L  E  Q  F  R  E  K  L
GAATCAAGACGCTGTATGCAAGAAAGTCCAAAGTGATGAAAGTTTTTGTATAAATAAACCGTTAGAGCAGTTTAGAGAGAAACTAA
```

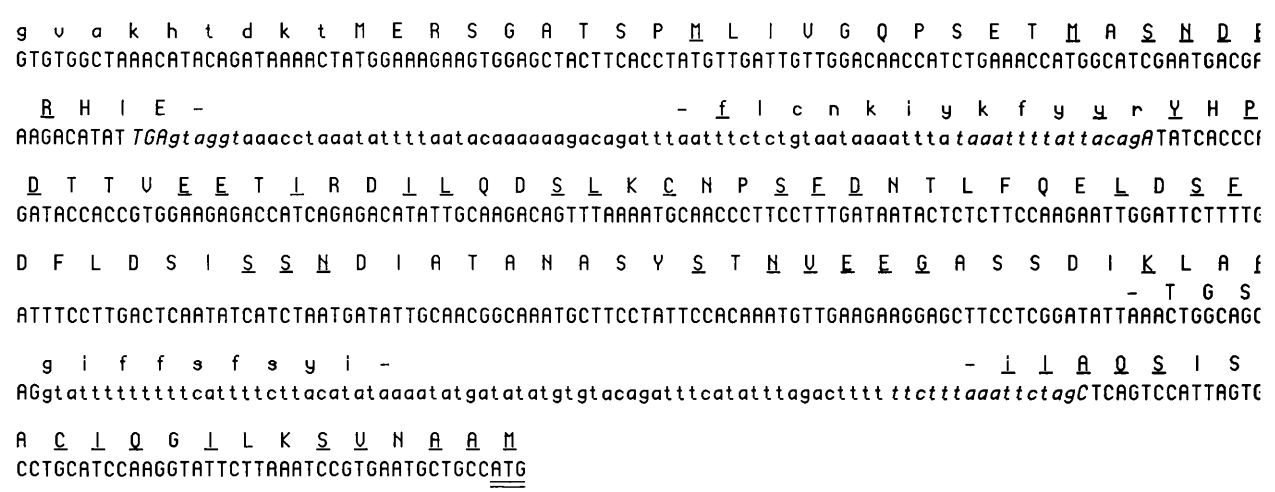
HHV-6

```
 E  D  K  Q  P  D  K  K  E  E  S  D  q  k  y  k  c  i  v  v  v  n  f  y  f  s  c  v  t  v  v  d  f
GAAGACAAGCAACCGGACAAAAAGGAGGAATCGG ATGgtaaataataaatgtatagttgttgtaaatttttatttttcgtgtgtcacggtcgtggacttttt

 f  l  n  s  i  c  f  v  a  i  s  E  L  P  K  E  K  R  Q  K  Y  H  D  M  K  R  N  L  E  E  Q  S  H  E
ttttaaatagtatctgtt ttgttgcaatttcagAACTTCCTAAGGAGAAGAGACAGAAGTACCATGACATGAAACGTAATTTGGAAGAACAAAGTCACGA

 D  G  I  T  L  T  S  T  T  L  V  N  G  A  V  E  G  A  L  P  P  C  I  S  I  D  N  H  E  D  Q  Q  H
GGATGGCATAACACTAACTTCTACGACACTCGTAAATGGAGCGGTTGAGGGTGCGTTACCGCCCTGTATTTCTATTGATAATCACGAAGATCAACAACAT

 D  E  L  D  K  R  A  Y  A  Q  v  q  q  v  s  g  s  p  k  p  r  s  s  g  s  l  l  c  v  s  k  a  d
GATGAATTAGACAAGCGCGCTTATGCG CAGgtgggtggagtttcgggttcgccaaaacctaggtcttcggggtctttgttgtgcgtatctaaagctgatc

          -  G  T  N  R  E  G  L  S  N  E  D  N  Y  G  N  F  R  L  N  K  S  L  E  Q  L  R  A  K  L
 l  f  i  l  g  d  e  p  r  r  s  i  q  -
ttttttattttagGGGACGAACCGAGAAGGTCTATCCAATGAAGATAATTATGGGAATTTTCGGCTTAATAAGTCTTTAGAACAGCTTAGGGCCAAACTTC
```

Fig. 32.
Sequence-derived evidence for splicing U79. Details of the layout are the same as in Fig. 25.

U90 upstream and internal splices

HHV-7



HHV-6

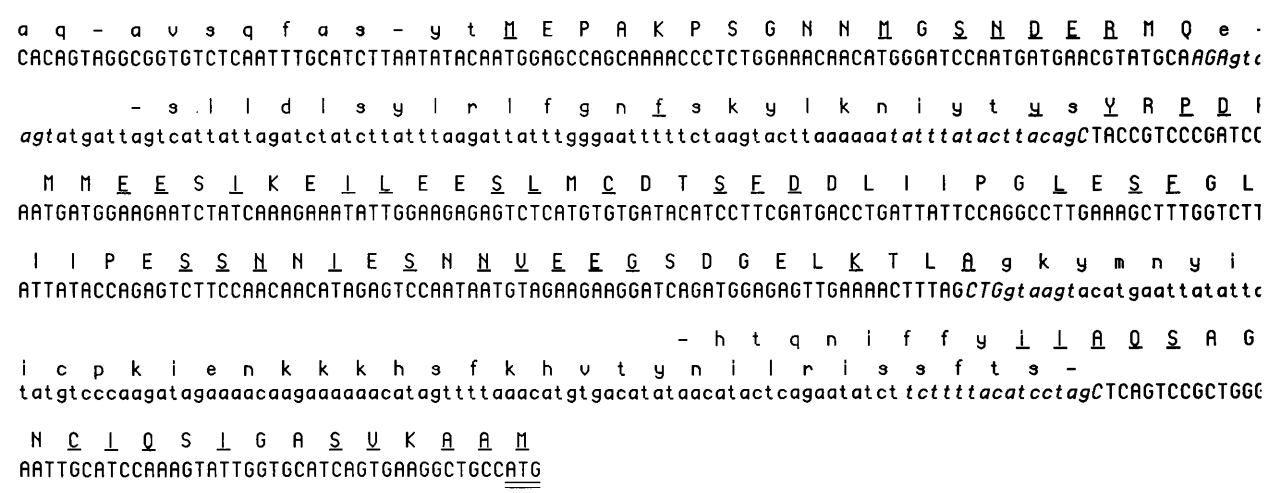


Fig. 33.
Sequence-derived evidence for splicing U90. Details of the layout are the same as in Fig. 25.

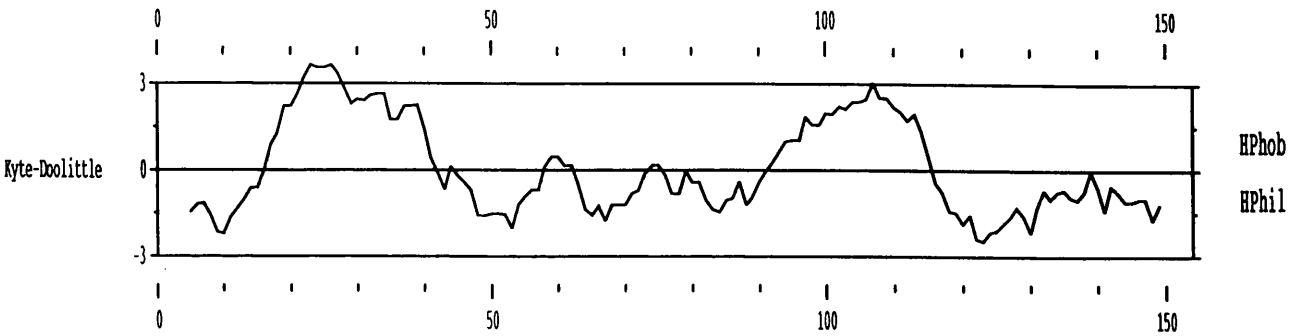
U91 internal splices

HHV-7

HHV-6

Fig. 34.
Sequence-derived evidence for splicing U91. Details of the layout are the same as in Fig. 25.

HHV-7 U91



HHV-6 U91

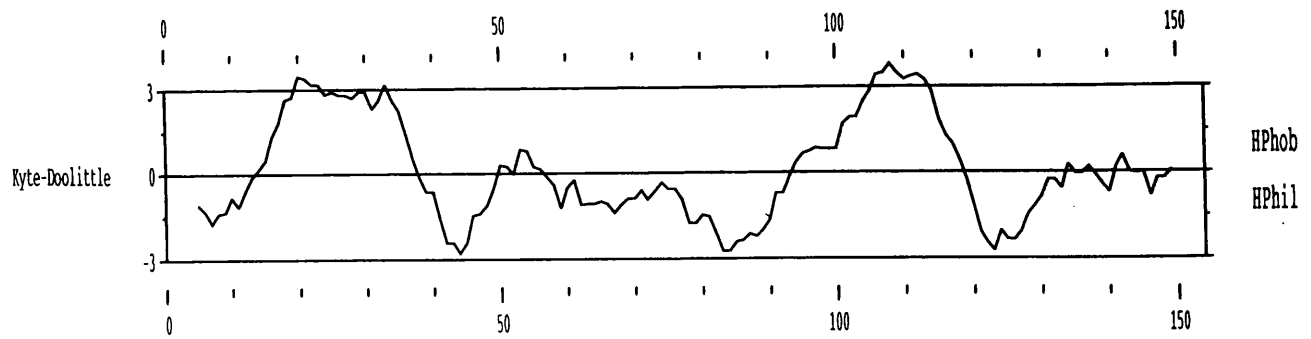


Fig. 35

Hydrophobicity profiles generated using Pepplot from the U91proteins potentially expressed by splicing.

U100 - Of the spliced genes recognised in HHV-6 and HHV-7, U100 is the most complex. The region at the right end of U contains several short ORFs which are expressed in HHV-6 as a multiply spliced mRNA encoding the virion envelope glycoprotein gp105 (Pfeiffer *et al.,* 1995). The mRNA contains ten coding exons (here termed exons 1-10) and two upstream non-coding exons. The likelihood of splicing in this gene was pointed out in the HHV-7 JI sequence analysis, but no splicing pattern was deduced. In the HHV-6 sequence, U100 is represented by U100 (exon 1), U99 (exon 2), U98 (exon 3), U97 (exon 7) and U96 (exon 10), and in this part of the HHV-7 genome, three of the four ORFs were reported to be counterparts of those in HHV-6. A scheme for U100 splicing in HHV-7 RK is given in Table 12. An amino acid alignment of the putative U100 proteins from HHV-6 and HHV-7 is shown in Fig. 36.

The amino acid similarity for the most part is convincing, although less so in the regions encoded by exons 1 and 10. In addition, it was necessary to propose that the exon 8 and exon 10 acceptor sites are not located similarly in HHV-7 and HHV-6, although they are in the same register in each genome. Indeed, it is curious that use of an alternative acceptor site for HHV-6 exon 8 located upstream from the site mapped by Pfeiffer *et al.* (1995) could significantly extend the similarity with the HHV-7 exon 8 polypeptide; details (see Fig.36) The site is an incorrect reading frame, however, so this point remains unsolved. Of the 18 cysteine residues in the HHV-7 U100 protein, 14 are conserved in HHV-6. Three of the remainder are encoded by ORF 8. U100 in HHV-7 JI corresponds to exon 1, U99 to exon 2, U98 to exon 3 and H8 to exon 10. The status of upstream non-coding exons identified in HHV-6 by Pfeiffer *et al.* (1995) is unknown.

Other possible splice sites were noted during the analysis, including those consistent with splicing of U12 exon 1 to U13 and the 3' end of U19 to an unidentified downstream exon. However, the evidence in these instances was weak, and they have not been included in Table 12. Moreover, in a few regions where splicing might reasonably be suspected, such as that containing U24 and U24a, no supporting evidence was forthcoming.

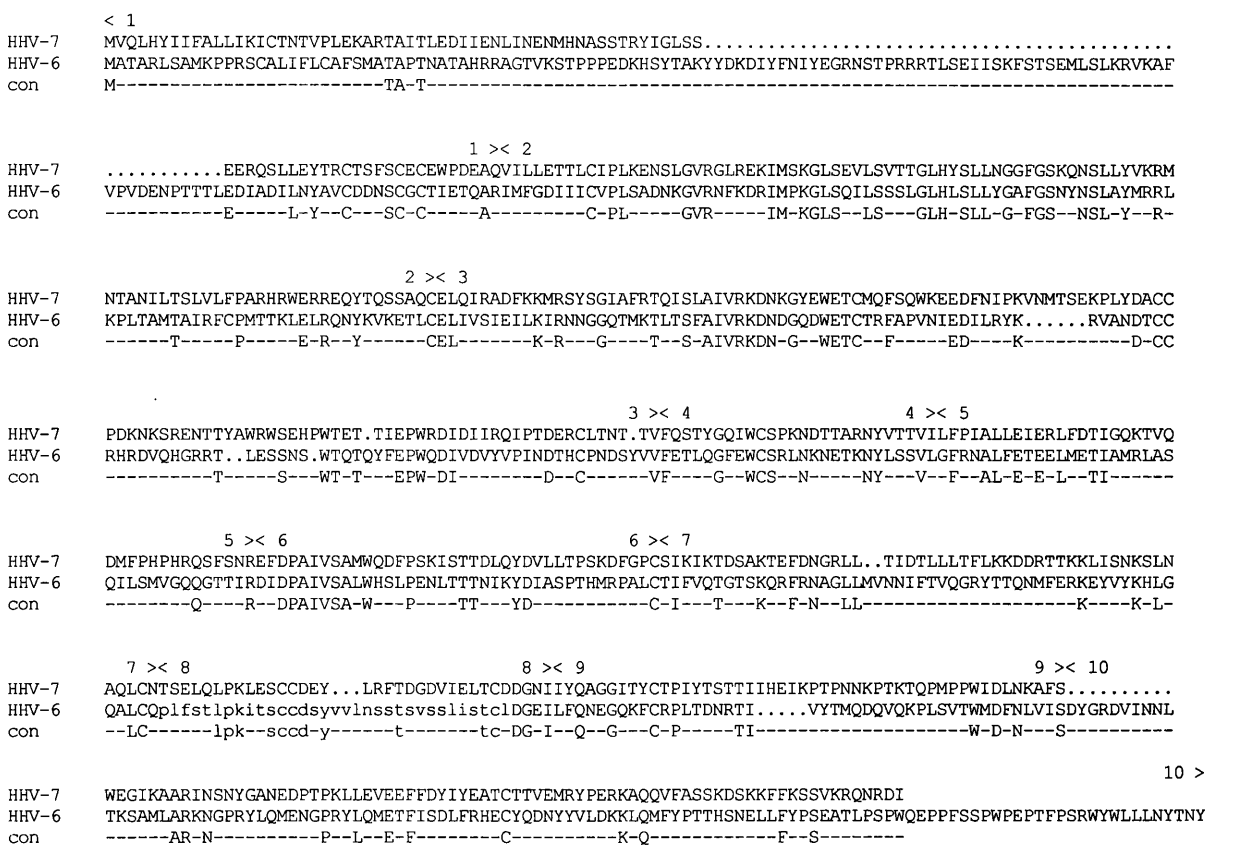


Fig. 36.

Sequence-derived evidence for the splicing of U100, in the form of alignments of putative amino acid sequences in HHV-7 and HHV-6. Exons are numbered 1-10 and their junctions are indicated by arrows. Dots indicate padding characters. Conserved amino acid residues are shown in upper case. Lower case residues in HHV-6 are potentially encoded by the HHV-6 DNA sequence upstream of exon 8, but lack a suitable splice acceptor site and are not encoded by anmRNA analysed by Pfeiffer *et al.* (1995); see text for details.

Chapter 4

Discussion

4.0. DISCUSSION

This chapter allows me the opportunity to comment on the implications of several aspects of the results. These include: the relationship between HHV-7 strains RK and JI; the putative functions of the telomeric elements; the coding potential and importance of splicing in the HHV-7 genome; and the phenotypic differences between HHV-7 and HHV-6.

4.1. THE DNA SEQUENCES OF HHV-7 RK AND JI

4.1.1. Strain divergence

McGeoch and Cook (1994) estimated an overall rate of sequence change of 1 x 10^{-7} synonymous substitutions per synonymous site per year in the gB gene, based on cospeciation of herpesviruses with their hosts over periods up to 60 Myears. Although this value was claimed as an order of magnitude estimate only, it was considered compatible with estimates derived by analysis of HSV-1 isolates from human populations established over a much shorter period (0.1 Myear). On this basis, the presence of 84 synonymous substitutions (Table 10) in the U regions of HHV-7 RK and JI (which, unlike DR, is not potentially subject to intragenomic homologous recombination) indicates that the lineages resulting in the two strains diverged from a common ancestor of the order of 10,000 years ago. In contrast, based on comparisons between the amino acid sequence of the HHV-7 gB gene and existing alignments derived from several other gB genes (McGeoch *et al.,* 1995), HHV-7 and HHV-6 are estimated to have diverged 50-60 million years ago (D.J. McGeoch, personal communication).

4.1.2. Nucleotide differences

This is the first occasion when complete sequences from two strains of one herpesvirus have each been compared in their entirety. As anticipated, the majority of nucleotide differences between HHV-7 RK and JI do not affect coding capacity (Table 9 and 10). Moreover, differences that do affect coding potential are scattered throughout the genome and appear not to be clustered in particular genes (Fig. 22). This indicates that no small subset of genes has been demonstrably subject to unusual evolutionary pressures since divergence of the two strains.

4.1.3. Genome ends

Sequences containing the genome termini were identified in the database since they share one end. Most of the clones from the left genome terminus of HHV-7 RK had six C residues and none had seven (Fig. 21), although the possibility of heterogeneity was suggested by the presence of less than six C residues in a few terminal clones. As fragments were end repaired (using T4 DNA polymerase) prior to ligation, no information was obtained on unpaired nucleotides at the termini. The genome termini of another HHV-7 strain (R-2), also sequenced from end-repaired DNA, are identical to those of RK (Secchiero *et al.,* 1995). Sequences from the termini of HHV-7 JI have not been determined directly. However, Secchiero *et al.* (1995) sequenced regions of HHV-7 JI concatameric DNA (consisting of head to tail copies of the genome) which span the DR-DR junction and thus correspond to a fused version of the termini, including any unpaired residues normally present at the termini. Sequences corresponding to the left genome terminus contain between four and six C nucleotides, and in some instances one or two additional residues at the DR-DR junction itself. This supports the length variability of the stretch of C nucleotides present at the left terminus of RK. Similarly, sequence variation in terminal base composition has previously been observed in HHV-6 (Thomson *et al.,* 1994). The right terminus of the published JI genome sequence is identical to that of RK (Fig. 22a), but the left terminus is one base pair longer, containing seven rather than six C nucleotides (Fig. 22b). The origin of this additional nucleotide is unclear.

4.1.4. Reiterations

Singer and Frenkel (1997) used Southern blot hybridisation of *Bam*HI-cleaved HHV-7 RK genomic DNA to determine the approximate length of the BamHI fragments from the left and right ends of the genome. These fragments contain the telomeric arrays T1 and T2. They estimated the restriction fragment from the left end (containing T1) to be 4.5 kbp in length, and that of the right end (containing T2) to be 3.8 kbp. From the RK DNA sequence, the *Bam*HI fragment at the right end is 3811 bp in length, correlating very closely with the experimentally determined size. However the length of the fragment containing T1 in the DNA sequence is about 1.7 kbp longer than that determined experimentally. Thus, it is likely that the solution for T1 sequence is not a correct representation of the genome sequence. This may reflect either the intractability of this region or sequence heterogeneity occurring in the genome or in M13

recombinants. Variability in the genome itself is supported by the observation that the terminal *Bam*HI fragment identified by Singer and Frenkel (1997) is heterogeneous in size. Therefore, it must be emphasised that although the size of the RK genome as derived from the sequence is 153080 bp, the actual size of the major species is likely to be a little less than 150 kbp.

The corresponding sequences in JI genomic DNA are likely to be different, however, since they vary in size in different plasmid clones (Ruvolo *et al.*, 1996). In support of the view that HHV-7 does not have unique genome size, Frenkel and Roffman (1996) reported that the genomes of different HHV-7 strains range in size from 140 to 150 kbp, and suggested that the main cause of variation involves the telomeric reiterations.

4.1.5. Resolution of the T1 reiteration

If, as appears probable, the T1 array is heterogeneous in sequence even within a single HHV-7 strain, this would somewhat reduce the imperative to realise a single solution for T1. If further investigation of the array were required, longer sequences could be generated from M13 clones using an automatic sequencer. Such sequences would still be unable to span the T1 array, but might prove less difficult to overlap. However, even longer sequences from a heterogeneous target would not provide a single solution.

An alternative would be to sequence DNA from a viral clone. This would at least provide a representative array from one variant in a heterogeneous population. Cloned purification of HHV-7 is likely to be problematic, however, given the growth properties of the virus.

4.2. FUNCTIONS OF THE TELOMERIC REITERATIONS

The human telomere-like elements in HHV-7 (and HHV-6) comprise arrays of a tandem repeat $(TAACCC)_n$, and related sequences (Martin *et al.*, 1991a; Secchiero *et al.*, 1995). In addition to HHV-7 and HHV-6, human telomere-like repeats have been reported in the genomes of other lymphotropic herpesviruses, including alphaherpesviruses (MDV) and gammaherpesviruses (EHV-2) (Kishi *et al.*, 1991, 1988; Telford *et al.*, 1995).

Whether these elements are functionally related to those at human telomeres is a matter of speculation, and resemblance of the repeated elements to human telomeric repeats may be coincidental. Nevertheless, it seems unlikely that different herpesviruses would independently generate (or capture) tandem arrays of the same element unless functionally driven. Furthermore, such small reiterated elements are relatively uncommon in herpesvirus genomes. It is likely, then, that telomeric elements have been acquired and retained in order to perform a function of benefit to the viruses. Consequently, several authors have suggested possible functions in lytic phase replication or latency.

It is perhaps significant that in HHV-6, HHV-7 and MDV (but not EHV-2) the telomeric arrays are located near the genome ends and adjacent to the conserved pac-1 and pac-2 sites, which direct cleavage and packaging of replicated DNA to unit-length genomes. In view of this, several authors (Secchiero *et al.,* 1995; Frenkel and Roffman, 1996; Thomson *et al.,* 1994; Gompels and Macaulay, 1995; Gompels *et al.,* 1995; Martin *et al.,* 1991) have proposed that the telomeric reiterations could play a role in the cleavage of concatemers (Liu and Gilbert, 1994) or DNA packaging during lytic phase replication (Deiss *et al.,* 1986).

In addition, it has been suggested that the telomeric reiterations may play a role in maintenance of the latent state. Positioned at the herpesvirus genome termini, the telomeric elements may enable the linear genome to survive as a distinct mini-chromosome in the dividing host cell and also achieve segregation during cell division (Thomson *et al.,* 1994; Gompels *et al.,* 1995; Gompels and Macaulay, 1995). Alternatively, Torelli *et al.* (1995) proposed that herpesviruses containing telomeric elements may achieve latency *via* a different route. The arrays may enable viral DNA to integrate into the telomeres of host cell chromosomes. However, the ability to segregate during cell division, is likely to be advantageous to the viral genome only if it is normally latent in dividing cells. The site(s) of persistent infection for HHV-7 and HHV-6 are not known, although both display tropism for non-dividing T-cells. Additionally, evidence in support of HHV-7 or HHV-6 integration into the host genome is scarce, nor is there any expectation of integration from what is known of other herpesviruses.

Secchiero *et al.* (1995) speculated that the telomeric arrays in HHV-6 and HHV-7 might be involved in the regulation of viral gene expression. Additionally, Gompels and Macaulay (1995) observed that monomers of the telomeric element are distributed around HHV-6 ori, suggesting a possible role in lytic phase replication. In contrast, telomeric elements are not located near HHV-7 RK ori. However, this does not completely rule out a role for the telomeric elements in HHV-6 lytic replication since the structure of its origin is different from that of HHV-7.

4.3. CODING POTENTIAL OF HHV-7

As a result of this study, HHV-7 RK is considered to contain 84 different genes (Table 11) and HHV-6 is predicted to contain 85 different genes. Two genes (DR1 and DR6) are present twice in each genome. The deduced layouts of HHV-7 and HHV-6 genes are shown in Figs. 23 and 24, respectively. In addition, HHV-7 genetic features and an alignment of predicted amino acid sequences with the genome sequence are given in Appendix I.

As expected, the HHV-7 RK contains approximately 40 core genes, and in addition approximately 30 more genes have counterparts in HCMV. All but one of the remainder have direct counterparts in HHV-6. Even the one exception gene, U55B, is related to U55A, and is thus also homologous to HHV-6 U55. HHV-7 U55B and U55A are also homologous to HCMV UL84, which may encode an essential replication function (Iskenderian *et al.,* 1996). Whether this is an example of the HHV-7 genome generating diversity by gene duplication or, alternatively, expulsion of genetic material from HHV-6 and HCMV is unclear; the former seems more likely.

HHV-6 contains only two genes that are absent from HHV-7: U22 which is a putative membrane glycoprotein, and a homologue of the AAV-2 *rep* gene, U94 (Gompels *et al.,* 1995). The function of U94 is unknown.

Gene U91 was not previously recognised as encoding a putative membrane glycoprotein in HHV-7 and HHV-6. The amino acid sequences of the two proteins contain two pronounced hydrophobic regions (one near the N terminus and the other about two-thirds through the

protein), as well as a single consensus N-linked glycosylation site of the form NXS or NXT located between the two hydrophobic domains. In addition, U91 in HHV-7 has two further potential N-linked glycosylation sites near the C terminus. Each hydrophobic domain is likely to anchor in the host cell, and possibly the virion membrane. If the central region is situated outside the cell, each protein could be N-glycosylated. Alternatively, if the the central region resides within the cell, only HHV-7 U91 is likely to be N-glycosylated.

4.4. SPLICING

4.4.1. Splicing patterns in HHV-7 and HHV-6

Drawing largely on the correspondence between the HHV-6 and HHV-7 gene arrangements, and in some cases or experimental data from HHV-6, it was possible to determine splicing patterns in 11 out of the total of 84 HHV-7 genes. Sites in 9 of these genes represent new predictions.

The comparative aspects of the analysis ensured that substantial confidence can be placed in the predicted splicing patterns. Some reservations should be stated, however. Heavy reliance on homology between HHV-7 and HHV-6 will have restricted recognition of spliced genes to those which are conserved in both viruses. Splicing in genes that are unique to one virus, and splicing involving non-coding exons are unlikely to have been identified from DNA sequences alone. Also, examples of complex splicing (such as that in U100) or alternative splicing are likely to have remained undetected. Thus, although splicing is likely to be common in HHV-7 and HHV-6, it is expected that the full extent of splicing has not yet been appreciated.

Some of the genes potentially spliced in HHV-7 and HHV-6 also have counterparts in HCMV and MCMV. Indeed, it was evidence for splicing in certain HCMV genes that led to initial suspicions of splicing in their HHV-7 and HHV-6 counterparts (e.g. U17). However, the evolutionary distances between HHV-7 or HHV-6 and HCMV or MCMV precluded the easy identification of corresponding splicing patterns in HCMV or MCMV.

4.4.2. Confirmation of splicing

Identifying spliced genes using sequence data only is a speculative occupation. It is notable that experimental data relating to splicing are limited for HHV-6 and completely absent for HHV-7. Consequently, further experimental investigation into splicing in both genomes would be useful. Transcript mapping by Northern blotting, S1 nuclease analysis, primer extension and cDNA cloning are the obvious routes.

4.4.3. Origins of splicing

HHV-7 U66 is the only core gene that is spliced in all of the sequenced herpesviruses. In contrast, examples of splicing have been documented among α-, β- and γ- non-core genes, particularly within the genomes of HCMV and EBV. Several HCMV non-core genes, including UL36/UL37, UL122/UL123 and UL112, are spliced, and some make use of alternative temporally regulated splicing (Mocarski, 1996). In the EBV genome all of the EBNA genes, plus LMP-2, which are latency associated and specific to EBV, express highly spliced transcripts. In contrast, the lytic cycle genes of EBV, many of which are core genes, are either unspliced or contain one or two short introns (Sample, 1994).

Thus, it is notable that non-core genes are more likely to exhibit splicing than core genes. Many of these genes probably originated from the host genome by gene capture, or were generated de novo in relatively recent evolutionary times (Davison and McGeoch, 1995; McGeoch, 1989). It is unlikely that splicing in recently acquired genes reflects splicing in the host genome, since gene capture seems to occur by reverse transcription processes. However, splicing (particularly alternative splicing) may be introduced to modify the expression of genes that are recently established in the viral genome. In addition, although gene duplication is a method of diversity well established in the herpesviruses (e.g. the US22 family), splicing may also be employed as an aid to divergence, particularly since it requires less DNA (Krainer and Maniatis, 1988).

4.5. PHENOTYPIC DIFFERENCES BETWEEN HHV-6 AND HHV-7

Although HHV-6 and HHV-7 have very similar genetic contents, their long evolutionary divergence (50-60 Myears) is reflected in disparate DNA and encoded protein sequences.

Indeed, the level of amino acid sequence similarity is significant but, for may genes, modest (Table 11). Therefore, the hope of identifying the genetic causes of phenotypic differences by sequence comparisons appears forlorn at present. These differences are, in any case, likely to be multifactorial.

Phenotypically HHV-7 and HHV-6 are very similar, but the two viruses do differ in several ways. Both replicate optimally in $CD4^+$ T-cells, but whereas HHV-6 has a wider host range, that of HHV-7 is more limited (Ablashi *et al.,* 1991; Frenkel *et al.,* 1990). The cellular receptor used by HHV-6 is not known, but is believed not to be CD4. By contrast, CD4 appears to be a critical part of the cell receptor used by HHV-7. Moreover, HHV-7 infection selectively down-modulates surface expression of cellular CD4, but HHV-6 enhances CD4 presentation during infection (Lusso *et al.,* 1989, 1991, 1994). HHV-6 infection occurs early in childhood and 80% of children develop an antibody response by 13 months. HHV-7 infection also occurs in childhood, but later than HHV-6 infection, perhaps taking up to three years from birth to develop a similar level of infection (Briggs *et al.,* 1988; Brown *et al.,* 1988; Okuno *et al.,* 1989; Wyatt *et al.,* 1991). The B variant of HHV-6 is widely accepted to be a causative agent of ES (Yamanishi *et al.,* 1988), but the A variant and HHV-7 have no proven involvement in any disease.

The relationship between the two variants of HHV-6 (A and B) is much closer than that between HHV-6 and HHV-7. The variants share approximately 95%-99.5% DNA sequence identity and approximately 95%-100% amino acid sequence identity (Ablashi *et al.,* 1993). Yet HHV-6 A and B also differ phenotypically in several ways, including: differential tropism for different $CD4^+$ T cell lines; and the etiologic link of the B variant to ES. Therefore, it is perhaps surprising that the phenotypes of HHV-6 and HHV-7 do not differ more significantly.

4.6. CONCLUSION

The determination of the complete sequence of HHV-7 RK has facilitated detailed genetic comparisons between two strains of the virus and allowed re-evaluation of the genetic content of HHV-7 and HHV-6. Thus, the study has achieved several goals: facilitated the study of microevolution over the whole genome length; reinforced our appreciation of the variability of reiterated sequences in the HHV-7 genome and its implications in genome size; improved our

understanding of the genetic content of HHV-7 and HHV-6; led to detailed predictions of splicing in HHV-7 and HHV-6; and categorised U91 as a spliced, putative membrane glycoprotein in HHV-7 and HHV-6. This analysis thus lays the foundation for further experimental studies of gene expression in HHV-7.

References

Ablashi, D.V., Agut, H., Berneman, Z., Campadelli-Fiume, G., Carrigan, D., Ceccerini-Neilli, L., Chandran, B., Chou, H., Collandre, H., Cone, R., Dambaugh, T., Dewhurst, S., DiLuca, L., Foa-Tomasi, L., Fleckenstein, B., Frenkel, N., Gallo, R., Gompels, U., Hall, C., Jones, M., Lawrence, G., Martin, M., Montagnier, L., Neipel, F., Nicholas, J., Pellett, P., Razzaque, G., Torrelli, G., Thomson, B., Salahuddin, S., Wyatt, L. and Yamanishi, K. (1993). Human herpesvirus 6 strain groups: a nomenclature. *Arch. Virol.*, **129**, 363-366.

Ablashi, D.V., Balachandran, N., Josephs, S.F. Hung, C.L., Krueger, G.R.F., Kramarski, B., Salahuddin, S.Z. and Gallo, R.C. (1991). Genomic polymorphism, growth properties, and immunologic variations in human herpesvirus 6 isolates. *Virology* **184**, 545-552.

Ablashi, D.V., Berneman, Z.N., Kramarsky, B., Whitman, J., Asano, Y. and Pearson, G.R. (1995). Human herpesvirus 7 (HHV-7): current status. *Clin. Diag. Virol.* **4**, 1-13.

Ablashi, D.V., Lusso, P., Hung, C-L., Salahuddin, S.Z., Josephs, S.F., Liana, T., Kramarsky, B., Biberfeld, P., Markham, P.D. and Gallo, R.C. (1988). Utilization of human hematopoietic cell lines for the propagation and characterization of HBLV (human herpesvirus 6). *Int. J. Cancer,* **42**, 787-791.

Ablashi, D.V., Salahuddin, S.Z., Josephs, S.F., Imam, F., Lusso, P., Gallo, R.C., Hung, C., Lemp, J. and Markham, P.D. (1987). HBLV (or HHV-6) in human cell line. *Nature* **329**, 207.

Agut, H., Aubin, J.T. and Huraux, J.M., (1991) Homogeneous susceptibility of distinct human herpesvirus 6 strains to antivirals *in vitro. J. Infect. Dis.* **163**, 1382-1383.

Agut, H., Guetard, D., Collandre, H., Dauguet, C., Montagnier, L., Miclea, J.M., Baurman, H. and Gessain, A. (1988). Concomitant infection by human herpesvirus 6, HTLV-1, and HIV-2. *Lancet* **1**, 712.

Agut, H., Huraux, J.M. and Collandre, H., (1989). Susceptibility of human herpesvirus 6 to acyclovir and ganciclovir. *Lancet* **2**, 626.

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

Alford, C.A. and Britt, W.J. (1990). Cytomegalovirus. In: *Virology* (2nd edn), pp 1981-2010. B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman (eds), Raven Press, New York.

Anders, D.G. and Gibson, W. (1988). Location, transcription analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1. *J. Virol.* **62**, 1364-1372.

Anders, D.G. and Punturieri, S.M. (1991). Multicomponent origin of cytomegalovirus lytic-phase DNA replication. *J. Virol.* **65**, 931-937.

Anders, D.G., Kacica, M.A., Pari, G. and Punterieri, S.M. (1992). Boundaries and structure of human cytomegalovirus oriLyt, a complex origin for the lytic-phase DNA replication. *J. Virol.* **66**, 3373-3384.

Ando, Y., Kakimoto, K., Ekuni, Y. and Ichijo, M. (1992). HHV-6 infection during pregnancy and spontaneous abortion. *Lancet* **340**, 1289.

Appereley, J.F. and Goldman, J.M. (1988). Cytomegalovirus: biology, clinical features and methods for diagnosis. *Bone Marrow Transplant* **3**, 253-264.

Asano, Y., Nakashima, T., Yoshikawa, T., Suga, S., Yazaki, T., Kajita, Y. and Ozaki, T. (1991). Severity of human herpesvirus 6 viremia and clinical findings in infants with exanthem subitum. *J. Pediatr.* **118**, 891-895.

Asano, Y., Suga, S., Yoshikawa, T., Yazaki, T. and Uchikawa, T. (1995). Clinical features and viral excretion in an infant with primary human herpesvirus 7 infection. *Prediatrics* **95**, 187-190.

Asano, Y., Yoshikawa, T., Kajita, Y., Ogura, R., Suga, S., Yazaki, T., Nakashima, T., Yamada, A. and Kurata, T. (1992). Fatal encephalitis/encephalopathy in primary human herpesvirus-6 infection. *Arch. Dis. Child.* **67**, 1484-1485.

Asano, Y., Yoshikawa, T., Suga, S., Kobayashi, I., Nakashima, T. and Yazaki, T. (1994). Clinical features of infants with primary human herpesvirus 6 infection (exanthem subitum, roseola infantum). *Pediatrics* **93**, 104-108.

Asher, Y. Heller, M. and Becker, Y. (1969). Incorporation of lipids into herpes simplex virus particles. *J. Gen. Virol.* **4**, 65-76.

Aubin, J-T., Agut, H., Collandre, H., Yamanishi, K., Chandran, B., Montagnier, L. and Huraux, J-M. (1993). Antigenic and genetic differentiation of the two putative types of human herpes virus 6. *J. Virol. Methods,* **41**, 223-234.

Aubin, J.T., Collandre, H., Candotti, D., Ingrand, D., Rouzioux, C., Burgard, M., Richard, S., Huraux, J.M. and Agut, H. (1991). Several groups among the human herpesvirus 6 strains can be distinguished by southern blotting and polymerase chain reaction. *J. Clin. Microbiol.,* **29**, 367-372.

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

Balachandran, N., Amelse, R.E., Zhou, W.W. and Chang, C.K. (1989). Identification of proteins specific for human herpesvirus 6-infected human T cells. *J. Virol.* **63**, 2835-2840.

Balachandran, N., Tirwatnapong, S., Pfeiffer, B., Ablashi, D.V. and Salahuddin,S.Z. (1991). Electophoretic analysis of human herpesvirus 6 polypeptides immunoprecipitated from infected cells with human sera. *J. Infect. Dis.* **163**, 29-34.

Bankier, A.T., Beck, S., Bohni, R., Brown, C.M., Cerny, R., Chee, M.S., Hutchison III, C.A., Kouzarides, T., Martignetti, J.A., Preddie, E., Satchwell, S.C., Tomlinson, P., Weston, K.M. and Barrell, B.G. (1991). The DNA sequence of the human cytomegalovirus genome. *DNA sequence* **2**, 1-12.

Baringer, J.R. and Swoveland, P. (1973). Recovery of herpes simplex virus from human trigeminal ganglions. *N. Engl. J. Med.*, **288**, 648-650.

Bastian, F.O., Rabson, A.S., Yee, C.L. and Tralka,T.S. (1972). Herpesvirus hominis: isolation from human trigeminal ganglion. *Science* **178**, 306-307.

Beck, S. and Barrell, B. (1991). An HCMV reading frame which has similarly with both the V and C regions of the TCR gamma chain. *DNA sequence* **2**, 33-38.

Beck, S. and Barrell, B.G. (1988). Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens. *Nature* **331**, 269-272.

Becker, W.B., Engelbrecht, S., Becker, M.L.B., Piek, C., Robson, B.A., Wood, L. and Jacobs, P. (1988). Isolation of a new human herpesvirus producing a lytic infection of helper (CD4) T-lymphocytes in peripheral blood lymphocyte cultures - another cause of acquired immunodeficiency? *S. Am. Med. J.* **74**, 610-614.

Becroft, D.M.O. (1981). Prenatal cytomegalovirus infection: epidemiology, pathology, and pathogenesis. In Rosenberg HS and Berstein J (eds) *Perspective in Pediatric Pathology,* pp 203-241.

Bernard, J. and Mercier, A. (1993). Sequence of two Eco RI fragments from Salmonis herpesvirus 2 and comparison with Ictalurid herpesvirus 1. *Arch. Virol.* **132**, 437-442.

Berneman, Z.N., Ablashi, D.V., Li, G., Eger-Fletcher, M., Reitz, M.S. Jr., Hung, C.-L., Brus, I., Komaroff, A.L., and Gallo, R.C. (1992a). Human herpesvirus 7 is a T-lymphotropic virus and is related to, but significantly different from, human herpesvirus 6 and human cytomegalovirus. *Proc. Natl. Acad. Sci. USA* **89**, 10552-10556.

Berneman, Z.N., Gallo, R.C., Ablashi, D.V., Frenkel, N., Katsafanas, G., Kramarsky, B., and Brus, I. (1992b). Human herpesvirus 7 (HHV-7) strain JI: independent confirmation of HHV-7. *J. Infect. Dis.* **166**, 690-691.

Biberfeld, P., Karmarsky, B., Salahuddin, S.Z. and Gallo, R.C. (1987). Ultrastructural characterisation of a new human B-lymphotropic DNA virus (HBLV) isolated from patients with lymphoproliferative disease. *J. Natl. Cancer Inst.*, **79**, 933-941.

Birney, E., Kumar, S. and Krainer, R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803-5816.

Black, J. B., Inoue, N., Kite-Powell, K., Zaki, S. and Pellett, P.E. (1993). Frequent isolation of human herpesvirus 7 from saliva. *Virus Res.*, **29**, 91-98.

Black, J.B. and Pellett, P.E. (1993). Human herpesvirus 7. *Rev. Med. Virol.*, **3**, 217-223.

Black, J.B., Sanderlin, K.C., Goldsmith, C.S., Gary, H.E., Lopez, C. and Pellett, P.E. (1989). Growth properties of human herpesvirus 6 strain Z29. *J. Virol. Methods* **26**, 133-145

Bogner, E., Reschke, M., Reis, B., Mockenhaupt, T. and Radsak, K. (1993). Identification of the gene product encoded by ORF UL56 of the human cytomegalovirus genome. *Virology* **196**, 290-293.

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

Boppana, S.B., Pass, R.F., Britt, W.J., Stagno, S. and Alford, C.A. (1992). Symptomatic congenital cytomegalovirus infection: neonatal morbidity and mortality. *Ped. Infect. Dis. J.* **11**, 93-99.

Bornkamm, G.W., Delius, H., Fleckenstein, B., Werner, F.J. and Mulder, C. (1976). Stucture of herpesvirus saimiri genomes: arrangement of heavy and light sequences in the M genome. *J. Virol.* **19**, 154-161.

Borst, J., van de Griend, R.J., van Oostveen, J.W., Ang, S-L., Melief, C.J., Seidman, J.G. and Bolhuis, R.L.H. (1987). A T-cell receptor γ/CD3 complex found on cloned functional lymphocytes. *Nature* **325**, 683-688.

Boshoff, C., Whitby, D., Hatziionnou, T., Fisher, C., van der Walt, J., Hatzakis A., Weiss, R. and Schultz, T. (1995). Kaposi's sarcoma-associated herpesvirus in HIV-negative Kaposi's sarcoma. *Lancet* **345**, 1043-1044.

Bovenzi, P., Mirandola, P., Secchiero, P., Strumia, R., Cassai, E. and Di Luca, D. (1993). Human herpesvirus 6 (variant A) in Kaposi's sarcoma. *Lancet* **341**, 1288-1289.

Brenner, M.B., McLean, J., Scheft, H., Riberdy, J., Ang, S-L., Seidman, J.G., Devlin, P. and Krangel, M.S. (1987). Two forms of the T-cell receptor γprotein found on peripheral blood cytotoxic T lymphocytes. *Nature* **325**, 689-694.

Briggs, M., Fox, J. and Tedder, R.S. (1988). Age prevalence of antibody to human herpesvirus 6. *Lancet,* **1**, 1058-1059.

Britt, W.J. (1996). Human cytomegalovirus overview: the virus and its pathogenic mechanisms. *Bailliere's clinical infectious diseases* **3**, 307-325.

Brown, H.A., Sumaya, C.V., Liu, C.R., Ench, Y., Kovacs, A., Coronesi, M. and Kaplan, M.H. (1988). Fall in human herpesvirus 6 seropositivity with age. *Lancet,* **1**, 396.

Buchwald, D., Cheney, P.R., Peterson, D.L., Henry, B., Wormsley, S.B., Geiger, A., Ablashi, D.V., Salahuddin, S.Z., Saxinger, C., Biddie, R., and Komaroff, A. (1992). A chronic illness characterised by fatigue, neurologic and immunologic disorders, and active human herpesvirus type 6 infection. *Ann. Intern. Med.* **116**, 103-113.

Buckmaster, E.A., Scott, S.D., Saunderson, M.J., Boursnell, S.M. and Ross, N.L.J. (1988). Gene sequence mapping data from Marek's disease virus and herpesvirus of turkey's: Implications for herpesvirus classification. *J. Gen.Virol.* **69**, 2033-2042.

Burns, J.C., Newburger, J.W., Sundell, R., Wyatt, L.S. and Frenkel, N. (1994). Seroprevalence of human herpesvirus 7 in patient with Kawasaki disease. *Pediatr. Infect. Dis. J.* **13**, 168-169.

Burns, W.H. and Sandford, G.R. (1990). Susceptibility of human herpesvirus 6 to antivirals *in vitro. J. Infect. Dis.* **162**, 634-637.

Cameron, J.M. (1993). New antiherpes drugs in development. *Rev. Med. Virol.,* **3**, 225-236.

Campadelli-Fiume, G. (1994). Herpes simplex viruses: Viral glycoproteins. In: *Encyclopedia of virology*, vol 2. Edited by R.G. Webster and A. Granoff. Academic Press.

Carrigan, D.R. and Knox, K.K. (1994). Human herpesvirus 6 (HHV-6) isolation from bone marrow: HHV-6-associated bone marrow suppression in bone marrow transplant patients. *Blood* **84**, 3307-3310.

Carrigan, D.R., Drobyski, W.R., Russler, S.K., Tapper, M.A., Knox, K.K. and Ash, R.C. (1991). Interstitial pneumonitis associated with human herpesvirus 6 infection after marrow transplantation. *Lancet* **338**, 147-149.

Caserta, M.T., Hall, C.B., Schnabel, K., McIntyre, K., Long, C., Costanzo, M., Dewhurst, S., Insel, T. and Epstein, L.G. (1994). Neuroinvasion and persistence of human herpesvirus 6 in children. *J. Infect. Dis.* **170**, 1586-1589.

Cebrian, J., Berthelot, N. and Laithier, M. (1989). Genome structure of cottontail rabbit herpesvirus. *J. Virol.* **63**, 523-531.

Cha, T-A., Tom, E., Kemble, G.W., Duke, G.M., Mocarski, E.S. and Spaete, R.R. (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **70**, 78-83.

Challberg, M.D. (1991). Herpes simplex virus DNA replication. *Semin. Virol.* **2**, 247-256.

Challoner, P.B., Smith, K.T., Parker, J.D., MacLeod, D.L., Coulter, S.N., Rose, T.M., Schultz, E.R., Bennett, L., Garber, R.L., Chang, M., Schad, P.A., Stewart, P.M., Nowinski, R.C., Brown, J.P. and Burmer, G.C. (1995). Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc. Natl. Acad. Sci. USA.* **92**, 7440-7444.

Chang, Y., Cesarman, E., Knowles, D. and Moore, P. (1995). Identification and partial characterisation of a new human herpesvirus associated with Kaposis-sarcoma. *Aids Res. Human Retroviruses.* **11**, s68.

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

Chee, M.S., Bankier, A.T., Beck, S., Bohni, R., Brown, C.M., Cerny, R., Horsnell, T., Hutchison, C.A., Kouzarides, T., Martignetti, A.J., Preddie, E., Satchwell, S.C., Tomlinson, P., Weston, K.M. and Barrell, B.G. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. In: *Current topics in microbiology and immunology.* **154**, 125-169. J.K. McDougall (ed). Springer-Verlog, Berlin/Heidelberg.

Chen, H., Pesce, A.M. and Carbonari, M., (1992). Absence of antibodies to human herpevirus 6 in patients with slowly progressive human immunodeficiency virus type 1 infection. *Eur. J. Epidemiol.* **8**, 217-221.

Chen, M., Popescu, N., Woodworth, C., Berneman, Z., Corbellino, M., Lusso, P., Ablashi, D.V. and DiPaolo, J.A. (1994b). Human herpesvirus 6 infects cervical epithelial cells and transactivates human papillomavirus gene expression. *J. Virol.* **68**, 1173-1178.

Chen, M., Wang, H., Berneman, Z., Lusso, P., Ablashi, D.V. and DiPaolo, J.A. (1994a). Detection of human herpesvirus 6 and human papillomavirus 16 in cervical carcinoma. *Am. J. Pathol.* **145**, 1509-1516.

Cherrington, J.M., Miner, R. Hitchcock, M.J.M., Lalezari, J.P. and Drew, W.L. (1996). Susceptibility of human cytomegalovirus to cidofovir is unchanged after limited *in vivo* exposure to various clinical regimens of drug. *J. Infect. Dis.* **173**, 987-992.

Chrisp, P. and Clissold, S.P. (1991). Foscarnet. A review of its antiviral activity, pharmacokinetic properties and therapeutic use in immunocompromised patients with cytomegalovirus, and cell-mediated immunity. *Amer. J. Med.* **23**, 593-601.

Clark, D.A., Alexander, F.E., McKinney, P.A., Roberts, B.E., O'Brien, C., Jarrett, R.F., Cartwright, R.A. and Onions, D.E. (1990). The seroepidemiology of human herpesvirus 6 (HHV-6) from a case-control study of leukaemia and lymphoma. *Int. J. Cancer* **45**, 829-833.

Clark, D.A., Freeland, J.M.L., Mackie, P.L.K., Jarrett, R.F. and Onions, D.E. (1993). Prevalence of antibody to human herpesvirus 7 by age . *J. Infect. Dis.*, **168**, 251-252.

Collier, A.C., Meyers, J.D., Corey, L., Murphy, V.L., Roberts, P.L. and Handsfield, H.H. (1987). Cytomegalovirus infection in homosexual men. Relationship to sexual practices, antibody to human immunodeficiency virus, and cell-mediated immunity. *Am. J. Med.* **23**, 593-601.

Comps, M. and Cochennec, N. (1993). A herpes-like virus from the European oyster ostraeedulis L. *J. Invertebrate Pathology* **62**, 201-203.

Cone, R.W., Hackman, R.C., Huang, M.L., Bowden, R.A., Meyers, J.D., Metcalf, M., Zek, J., Ashley, R. and Corey, L. (1993). Human herpesvirus 6 in lung tissue from patients with pneumonitis after bone marrow transplantation. *N. Engl. J. Med.*, **329**, 156-161.

Costa, R.H., Draper, K.G., Devi-Rao, G., Thomson, R.L. and Wagner, E.K. (1985). Virus-induced modification of the host cell is required for expression of the bacterial chloramphenicol acetyltransferase gene controlled by a late herpes simplex virus protein (VP5). *J. Virol.* **56**, 19-30.

Craighead, J., Kanich, R.E. and Almeida, J.D. (1972). Nonviral microbodies with viral antigenicity produces in cytomegalovirus-infected cells. *J. Virol.* **10**, 766-775.

Cuende, J.I., Ruiz, J., Civeira, M.P. and Prieto, J. (1994). High prevalence of HHV-6 DNA in peripheral blood mononuclear cells of healthy individuals detected by nested-PCR. *J. Med. Virol.* **43**, 115-118.

Dale, J.K., Strauss, S.E., Ablashi, D.V., Salahuddin, Z.S., Gallo, R.C., Nishibe, Y. and Inoue, Y.K. (1989). The Inoue-Melnick virus, human herpesvirus type 6, and the chronic fatigue syndrome. *Ann. Intern. Med.* **110**, 92-93.

Dambaugh, T. and Hennessy, K., (1986). In: *The Epstein-Barr Virus,* p 1.Epstein M.A., Achong, B.G. (eds). Heineman Medical, London.

Dargan, D.J., Jamieson, F.E., MacLean, J., Dolan, A., Addison, C. and McGeoch, D.J. (1997). The published DNA sequence of human cytomegalovirus strain AD169 lacks 929 base pairs affecting genes UL42 and UL43. *J. Virol.* **71**, 9833-9836.

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

Davison, A.J. (1993). Herpesvirus genes. *Rev. Med. Virol.* **3**, 237-244.

Davison, A.J. and Clements, J.B. (1996). Herpesviruses: General properties. In: *Topley and Wilsons, Microbiology and Microbial Infections*, 9th edition. Edited by L. Collier. In press.

Davison, A.J. and Davison, M.D. (1995). Identification of structural proteins of channel catfish virus by mass spectrometry. *Virology* **206**, 1035-1043.

Davison, A.J. and McGeoch, D.J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *J. Gen. Virol.* **67**, 597-611.

Davison, A.J. and McGeoch, D.J. (1995). Herpesviridae. In: *Molecular evolution of viruses.* pp 290-309. A.J. Gibbs, F. Garcia-Arenal, C. Callisher (eds). Cambridge University Press.

Davison, A.J. and Scott, J.E. (1986). The complete DNA sequence of varicella-zoster virus. *J. Gen. Virol.* **67**, 1754-1816.

Davison, A.J. and Taylor, P.(1987). Genetic relations between varicella-zoster virus and Epstein-Barr virus. *J. Gen. Virol.* **68**, 1067-1079.

Davison, A.J. and Wilkie, N.M. (1983). Inversion of the 2 segments of the herpes simplex virus genome in intertypic recombinants. *J. Gen. Virol.* **64**, 1-18.

De Clercq, E. (1993). Therapeutic potential of HPMPC as an antiviral drug. *Rev. Med. Virol.* **3**, 85-96.

Deatly, A.M., Spivack, J.G., Lavi, E, O'Boyle, D.R. and Fraser, N.W. (1988). Latent herpes simplex virus type 1 transcripts in peripheral and central nervous system tissue of mice map to similar regions of the viral genome. *J. Virol.* **62**, 749-756.

DeMarchi, J.M. (1981). Human cytomegalovirus DNA: restriction enzyme cleavage maps and map location for immediate early, early, and late RNA. *Virology* **114**, 23-38.

Demmler, G.J. (1991). Infectious Diseases Society of America and Centers for Disease Control. Summary of a workshop on surveillance for congenital cytomegalovirus disease. *Rev. Infect. Dis.* **13**, 315-329.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of computer programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

Dewhurst, S., McIntyre, K., Schnabel, K. and Hall, C.B. (1993). Human herpesvirus 6 (HHV6) varient B accounts for the majority of symptomatic primary HHV6 infections in a population of US infants. *J. Clinical Microbiol.* **31**, 416-418.

Di Luca, D., Dolcetti, R., Mirandola, P., De Re, V., Secchiero, P., Carbone, A., Boiocchi, M. and Cassi, E. (1994). Human herpesvirus 6: a survey of presence and variant distribution in normal peripheral lymphocytes and lymphoproliferative disorders. *J. Infect. Dis.* **170**, 211-215.

Di Luca, D., Katsafanas, G., Schrimer, E.C., Balachandran, N. and Frenkel, N. (1990). The replication of viral and cellular DNA in human herpesvirus 6-infected cells. *Virology* **175**, 199-210.

Dieterich, D.T., and Rahmin, M. (1991). Cytomegalovirus colitis in AIDS: presentation in 44 patients and a review of the literature. *J. of the Acquired Immunodeficiency Virus* **4**, S29-S35.

Dolan, A., Arbuckle, M. and McGeoch, D.J. (1991). Sequence analysis of the splice junction in the transcript of herpes simplex virus type 1 gene UL15. *Virus Res.* **20**, 97-104.

Downing, R.G., Sewankambo, N., Serwadda, D., Honess, R.W., Crawford, D., Jarrett, R. and Griffin, B.E. (1987). Isolation of human lymphotropic herpesviruses from Uganda. *Lancet* **2**, 390.

Drew, W.L., Mintz, L., Miner, R.C., Sands, M. and Ketterer, B. (1981). Prevalence of cytomegalovirus in homosexual men. *J. Infect. Dis.* **143**, 188-192.

Drew, W.L., Sweet, E.S., Miner, R.C. and Mocarski, E.S. (1984). Multiple infections by cytomegalovirus in patients with acquired immunodeficiency syndrome: documentation by Southern blot hybridization. *J. Infect. Dis.* **150**, 952-953.

Drobyski, W.R., Dunne, W.M., Burd, E.M., Knox, K.K., Ash, R.C., Horowitz, M.M., Flomenberg, N. and Carrigan, D.R. (1993). Human herpesvirus 6 (HHV-6) infection in allogeneic bone marrow transplant recipients: evidence of a marrow-suppressive role for HHV-6 *in vivo*. *J. Infect. Dis.* **167**, 735-739.

Drobyski, W.R., Knox, K.K., Majewski, D. and Carrigan, D.R. (1994). Fatal encephalitis due to variant B human herpesvirus 6 infection in a bone marrow-transplant recipient. *N. Engl. J. Med.* **330**, 1356-1360.

Dunne, W.M. and Demmler, G.J. (1992). Serological evidence for congenital transmission of human herpesvirus 6. *Lancet,* **340**, 121-122.

Dunne, W.M. and Jevon, M. (1993). Examination of human breast milk for evidence of human herpesvirus 6 by polymerase chain reaction. *J. Infect. Dis.,* **168**, 250.

Dupin, N., Grandadam, M., Calvez, V., Gorin, I., Aubin, J.T., Harvard, S., Lamy, F., Leibowitch, M., Huraux, J.M., Escande, J.P. and Agut, H. (1995). Herpesvirus-like DNA in patients with Mediterranean Kaposi's sarcoma. *Lancet* **345**, 761-762.

Efstathiou, S., Lawrence, G.L., Brown, C.M. and Barrell, B.G. (1992). Identification of homologues of the human cytomegalovirus US22 gene family in human herpes virus 6. *J. Gen. Virol.* **73**, 1661-1671.

Efstathiou, S., Minson, A.C., Field, H.J., Anderson, J.R. and Wildy,P. (1986). Detection of herpes simplex virus-specific DNA sequences in latently infected mice and in humans. *J. Virol.* **57**, 446-455.

Emery, V.C. and Griffiths, P.D. (1990). Current status review: Molecular biology of cytomegalovirus. *Int. J. Exp. Path.* **71**, 905-918.

Ensoli, B., Lusso, P., Schachter, F., Josephs, S.F., Rappaport, J., Negro, F., Gallo, R.C. and Wong-Staal, F. (1989). Human herpesvirus 6 increases HIV-1 expression in co-infected T-cells via nuclear factors binding to the HIV-1 enhancer. *EMBO. J.* **8**, 3019-3028.

Ensser, A., Pflanz, R. and Fleckenstein, B. (1997). Primary structure of the alcelaphaline herpesvirus 1 genome. *J. Virol.* **71**, 6517-6525.

Epstein, M.A. and Achong, B.G. (1986). *The Epstein-Barr virus: recent advances.* Ed. by W. Heinemann. William Heinemann Medical Books, London.

Epstein, M.A., Achong, B.G. and Barr, Y.M. (1964). Virus particles in cultured lymphoblasts from Burkitt's lymphoma. *Lancet* **1**, 702-703.

Ertl, P.F. and Powell, K.L. (1992). Physical and functional interaction of human cytomegalovirus DNA polymerase and its accessory protein (ICP36) expressed in insect cells. *J. Virol.* **66**, 4126-4133.

Essers, S., Schwinn, A., ter Meulen, J., von Lips, H., Dietz, K., Mhalu, F.S., Shao, J. and ter Meulen, V. (1991). Seroepidemiological correlations of antibodies to human herpesviruses and human immunodeficiency virus type 1 in African patients. *Eur. J. Epidemiol.* **7**, 658-664.

Evans, A.S. and Neiderman, J.C. (1991). Epstein-Barr virus. In: *Viral infections of humans, epidemiology and control*, 3rd Edn. Chr 10, 265-292.. Edited by A.S. Evans. Academic press.

Everett, R.D. (1987). The regulation of transcription of viral and cellular genes by herpesvirus immediate-early gene products (review). *Anticancer Res.* **7**, 584-604.

Farr, T.J., Harnett, G.B., Pietroboni, G.R. and Bucens, M.R. (1990). The distribution of antibodies to HHV-6 compared with other herpesviruses in young children. *Epidemiol. Infect.* **105**, 603-607.

Faulds, D. and Heel, R.C. (1990). Gancyclovir. A review of its antiviral activity, pharmacokinetic properties and therapeutic efficacy in cytomegalovirus infections. *Drugs* **39**, 597-638.

Fenner, F., Bachmann, P.A., Murphy, F.A., Studdert, M.J. and White, D.O. (1987). Herpesviridae. In: *Veterinary virology*, Chr 19, 339-373. Academic press.

Flamand, L., Gosselin, J., D'Addario, M., Hiscott, J., Ablashi, D.V., Gallo, R.C. and Menezes, J. (1991). Human herpesvirus 6 induces interleukin 1β and tumor necrosis factor α, but not interleukin 6, in peripheral blood mononuclear cell cultures. *J. Virol.* **65**, 5105-5110.

Flamand, L., Stefanescu, I., Ablashi, D.V. and Menezes, J. (1993). Activation of the Epstein-Barr virus replicative cycle by human herpesvirus 6. *J. Virol.* **67**, 6768-6777.

Fowler, K.B., Stagno, S., Pass, R.F., Britt, W.J., Boll, T.J. and Alford, C.A. (1992). The outcome of congenital cytomegalovirus infection in relation to maternal antibody status. *New Engl. J. Med.* **326**, 663-667.

Fox, J., Briggs, M. and Tedder, R.S., (1988). Antibody to human herpesvirus 6 in HIV-1 positive and negative homosexual men. *Lancet* **2**, 396-397.

Fox, J.D. Briggs, M., Ward, P.A. and Tedder, R.S. (1990). Human herpesvirus 6 in salivary glands. *Lancet,* **336**, 590-593.

Fraser, N.W., Lawrence, N.C., Wroblewska, Z., Gilden, D.H. and Koprowski, H. (1981). Herpes simplex virus type 1 DNA in human brain tissue. *Proc. Natl. Acad. Sci. USA* **78**, 6461-6465.

Frenkel, N. and Rapaport, D. (1995). Detection by PCR of human herpesviruses 6 and 7. In: "Frontiers of Virology. PCR: protocols for diagnosis of human and animal virus diseases" (Becker, Y. and Darai, G. Eds.), pp. 243-251.

Frenkel, N. and Wyatt, L.S. (1992). HHV-6 and HHV-7 as exogenous agents in human lymphocytes. *Dev. Biol. Stand.*, **76**, 259-265.

Frenkel, N., and Roffman, G. (1996). Human herpesvirus 7. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 3rd edition, pp. 2609-2622. Lippincott-Raven, Philadelphia.

Frenkel, N., Schirmer, E.C., Wyatt, L.S., Katsafanas, G., Roffman, E., Danovich, R.M. and June, C.H. (1990). Isolation of a new herpesvirus from human CD4 T cells. *Proc. Natl. Acad. Sci. USA* **87**, 748-752.

Furlong, G.D., Swift, H. and Roizman, B. (1972). Arrangement of herpesvirus deoxyribonucleic acid in the core. *J. Virol.* **10**, 1071-1074.

Furukawa, M., Yasukawa, M., Yakushijin, Y. and Fujita, S. (1994). Distinct effects of human herpesvirus 6 and human herpesvirus 7 on surface molecule expression and function of $CD4^+$ T cells. *J. Immunol.* **152**, 5768-5775.

Gallant, J.E., Moore, R.D., Richman, D.D., Keruly, J. and Chaisson, R.E. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. *J. Infect. Dis.* **166**, 1223-1227.

Gelb, L.D. (1990). Varicella-zoster virus. In: *Virology,* 2nd edn, pp 2011-2054. B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman (eds), Raven Press New York.

Geng, Y.Q., Chandran, B. and Joseph, S.F., (1992). Identification and characterization of a human herpesvirus 6 gene segment that trans-activates the human immunodeficiency virus type 1 promoter. *J. Virol.* **66**, 1564-1570.

Gerard, C. and Gerard, N.P. (1994). C5A Anaphylatoxin and its seven transmembrane seqment receptor. *Annu. Rev. Immunol.* **12**, 775-808.

Gibson, M.G. and Spear, P.G. (1983). Insertion mutants of herpes simplex virus have a duplication of the glycoprotein D gene and express two different forms of glycoprotein D. *J. Virol.* **48**, 396-404.

Gibson, W. (1981). Structural and nonstructural proteins of strain Colburn cytomegalovirus. *Virology* **111**, 516-537.

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

Gibson, W., McNally, L.M. and Welch, A.R. (1993). Cytomegalovirus maturational proteinase: site-directed mutagenesis used to probe enzymatic and substrate domains. In Michelson S and Plotkin SA (eds) *Multidisciplinary approach to understanding cytomegalovirus disease,* pp 21-25. Amsterdam: Elsevier.

Gold, D., Bowden, R., Sixbey, J., Riggs, R., Katon, W.J., Ashley, R., Obrigewitch, R.M. and Corey, L. (1990). Chronic fatigue. A prospective clinical and virologic study. *J. Amer. Med. Assoc.* **264**, 48-53.

Gompels, U.A. and Macaulay, H.A. (1995). Characterization of human telomeric repeat sequences from human herpesvirus 6 and relationship to replication. *J. Gen. Virol.* **76**, 451-458.

Gompels, U.A., Nicholas, J. Martin, M.E.D., Lawrence, G., Thomson, B., Efstathiou, S., Teo, I., Craxton, M.A., Barrell, B. and Honess, R.W. (1992). The DNA sequence of human herpesvirus 6 strain U1102: a comparative analysis of gene organisation and coding content. Abstract, *17th International Herpesvirus Workshop.* 43.

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

Griffin, A.M. (1989). The nucleotide sequence of the glycoprotein gB gene infectious laryngotracheitis virus: analysis and evolutionary relationship to the homologous gene from other herpesviruses. *J. Gen. Virol.* **72**, 393-398.

Hall, C.B., Long, C.E., Schnabel, K.C., Caserta, M.T., McIntyre, K.M., Costanzo, M.A., Knott, A., Dewhurst, S., Insel, R.A. and Epstein, L.G. (1994). Human herpesvirus 6 infection in children - A prospective study of complications and reactivation. *New Engl. J. Med.* **331**, 432-438.

Hall, J.D., Furman, P.A., St. Clair, M.H. and Knopf, C.W. (1985). Reduced *in vivo* mutagenesis by mutant herpes simplex DNA polymerase involves improved nucleotide selection. *Proc. Natl. Acad. Sci. USA.* **82**, 3889-3893.

Hamzeh, F.M., Lietman, P.S., Gibson, W. and Hayward, G.S. (1990). Identification of the lytic origin of DNA replication in human cytomegalovirus by a novel approach utilising ganciclovir-induced chain termination. *J. Virol.* **64**, 6184-6195.

Handsfield, H.H., Chandler, S.H., Caine, V.A., Meyers, J.D., Corey, L., Medeiros, E. and McDougall, J.K. (1985). Cytomegalovirus infection in sex partners. Evidence for sexual transmission. *J. Infect. Dis.* **151**, 344-348.

Harnett, G.B., Farr, T.J., Pietroboni, G.R. and Bucens, M.R. (1990). Frequent shedding of human herpesvirus-6 in saliva. *J. Med. Virol.* **30**, 128-130.

Harrison, P.T., Thompson, R. and Davison, A.J. (1991). Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinase. *J. Gen. Virol.* **72**, 2583-2586.

He, J., McCarth, M., Zhou, Y., Chandran, B. and Wood, C. (1996). Infection of primary human fetal astrocytes by human herpesvirus 6. *J. Virol.* **70**, 1296-1300.

He, Y.S., Xu, L. and Huang, E.S. (1992). Characterization of human cytomegalovirus UL84 early gene and identification of its putative protein product. *J. Virol.* **66**, 1098-1108.

Hidaka, Y., Lui, Y., Yamamoto, M., Mori, R., Miyazaki, C., Kushuhara, K., Okada, K. and Ueda, K. (1993). Frequent isolation of human herpesvirus 7 from saliva. *Virus Res.*, **29**, 91-98.

Hidaka, Y., Okada, K., Kusuhara, K., Miyazaki, C., Tokugawa, K. and Ueda, K. (1994). Exanthem subitum and human herpesvirus 7 infection. *Ped. Infect. Dis. J.* **13**, 1010-1011.

Hill, T.J., Blyth, W.A. and Harbour, D.A. (1978). Trauma to the skin causes recurrence of herpes simplex virus in the mouse. *J. Gen. Virol.* **39**, 21-28.

Hirai, K. and Watanabe, Y. (1976). Induction of alpha type DNA polymerases in human cytomegalovirus-infected WI-38 cells. *Biochim. Biophys. Acta.* **447**, 328-339.

Hirai, K., Furukawa, T. and Plotkin, S.A. (1976). Induction of DNA polymerase in WI-38 and guinea pig cells infected with human cytomegalovirus (HCMV). *Virology* **70**, 251-255.

Ho, M., Suwansirikul, S., Dowling, J.N., Youngblood, L.A. and Armstrong, J.A. (1975). The transplanted kidney as a source of cytomegalovirus infections. *New Engl. J. Med.* **293**, 1109-1112.

Holmes, G.P., Kaplan, J.E., Stewart, J.A., Hunt, B., Pinsky, P.F. and Schonberger, L.B. (1987). A cluster of patients with a chronic mononucleosis-like syndrome. *J. Amer. Med. Assoc.* **257**, 2297-2302.

Honess, R.W., Gompels, U.A., Barrell, B.G., Craxton, M., Staden, R., Chang, Y.N. and Hayward, G.S. (1989). Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *J. Gen. Virol.* **70**, 837-855.

Honess, R.W. (1984). Herpes simplex and "The herpes complex": Diverse observations and a unifying hypothesis. *J. Gen. Virol.*, **65**, 2077-2107.

Honess, R.W. and Roizman, B. (1975). Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J. Virol.* **14**, 8-19.

Honess, R.W., Bodemer, W., Cameron, K.R., Niller, N.-H. and Fleckenstein, B. (1986). The A+T-rich genome of herpesvirus saimiri contains a highly conserved gene for thymidylate synthase. *Proc. Natl. Acad. Sci. USA.* **83**, 3604-3608.

Hope-Simpson, R.E. (1965). The nature of herpes zoster: a long term study and a new hypothesis. *Proc. Royal Soc. Med.,* **58**, 9-20.

Horvat, R.T., Wood, C. and Balachandran, N. (1989). Transactivation of human immunodeficiency virus promoter by human herpesvirus 6. *J. Virol.* **63**, 970-973.

Horwitz, C.A., Henle, W., Henle, G., Snover, D., Rudnick, H., Balfour, H.H. Jr., Mazur, M.H., Watson, R., Schwartz, B. and Muller, N. (1986). Clinical and laboratory evaluation of cytomegalovirus-induced mononucleosis in previously healthy individuals. Report of 82 cases. *Medicine* **65**, 124-134.

Huang, E.S. (1975). Human cytomegalovirus. IV. Specific inhibition of virus-induced DNA polymerase activity and viral DNA replication by phosphonoacetic acid. *J. Virol.* **16**, 1560-1565.

Huang, L.M., Lee, C.Y., Chen, J.Y., Wang, J.D. and Hsu, C.Y. (1992). Primary human herpesvirus 6 infections in children: a prospective serological study. *J. Infect. Dis.* **165**, 1163-1164.

Huang, L.M., Lee, C.Y., Lee, P.I., Chen, J.M. and Wang, P.J. (1991). Meningitis caused by human herpesvirus 6. *Arch. Dis. Child.* **66**, 1443-1444.

Irmiere, A. and Gibson, W. (1983). Isolation of human cytomegalovirus intranuclear capsids, characterization of their protein constituents, and demonstration that the B-capsid assembly protein is also abundant in noninfectious enveloped particles. *J. Virol.* **56,** 277-283.

Irmiere, A. and Gibson, W. (1985). Isolation and characterisation of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. *Virology,* **130**, 118-133.

Irmiere, A. and Gibson, W. (1985). Isolation of human cytomegalovirus intranuclear capsids, characterization of their protein constituents, and demonstration that the B-capsid assembly protein is also abundant in noninfectious enveloped particles. *J. Virol.* **56**, 277-283.

Irving, W.L., Chang, J., Raymond, D.R., Dunstan, R., Grattan-Smith, P. and Cunningham, A.L. (1990). Roseola infantum and other syndromes associated with acute HHV-6 infection. *Arch. Dis. Child.* **65**, 1297-1300.

Jabs, D.A., Enger, C. and Bartlett, J.G. (1989). Cytomegalovirus retinitis and acquired immunodeficiency syndrome. *Archives of Opthamology* **107**, 75-80.

Jacob, R.J., Morse, L.S. and Roizman, B. (1979). Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. *J. Virol.* **29,** 442-457.

Jacobson, M.A. and Mills J. (1988). Serious cytomegalovirus disease in the acquired immunodeficiency syndrome (AIDS). *Annals of Internal Medicine* **108**, 585-594.

Jarrett, R.F., Clark, D.A., Josephs, S.F. and Onions, D.E. (1990). Detection of human herpesvirus 6 DNA in peripheral blood and saliva. *J. Med. Virol.* **32**, 73-76.

Jarrett, R.F., Gledhill, S., Qureshi, F., Crae, S.H., Madhok, R., Brown, I., Evans, I., Krajewski, A., O'Brien, C.J., Cartwright, R.A., Venables, P. and Onions, D.E. (1988). Identification of human herpesvirus 6-specific DNA sequences in two patients with non-Hodgkin's lymphoma. *Leukemia* **2**, 496-502.

Johnston, R.T. (1982). *Virus infections of the nervous system.* Raven Press, New York.

Jones, M.D. and Teo, I.A. (1992). Identification and analysis of the transport/capsid assembly protein (tp/cap) gene of human herpesvirus 6 (HHV-6). *Virology* **197**, 449-454.

Josephs, S.F., Buchbinder, A., Streicher, H.Z., Ablashi, D.V., Salahuddin, S.Z., Guo, Y.L., Wong-Staal, F., Cossman, J., Krueger, G.R.F., Fox,R.I. and Gallo, R.C. (1988). Detection of human B-lymphotropic virus (human herpesvirus 6) sequences in B cell lymphoma tissue of three patients. *Leukemia* **2**, 132-135.

Joyce, C.M. and Grindley, N.D.F. (1983). Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of Escherichia coli. *Proc. Natl. Acad. Sci. USA*, **80,** 1830-1834.

Karlin, S., Mocarski, E.S. and Schachtel, G.A. (1994). Molecular evolution of herpesviruses: genomic and protein sequence comparisons. *J. Virol.* **68**, 1886-1902.

Katsadanas, G.C., Schirmer, E.C., Wyatt, L.S. and Frenkel, N. (1996). *In vitro* activation of human herpesviruses 6 and 7 from latency. *Proc. Natl. Acad. Sci. USA.* **93**, 9788-9792.

Katsafanos, G.C., Shirmer, E.C., Wyatt, L.S. and Frenkel, N. (1996). *In vitro* activation of human herpesviruses 6 and 7 from latency. *Proc. Natl. Acad. Sci. USA.* **93**, 9788-9792.

Kawa Ha, K., Tanaka, K., Inoue, M., Sakata, N., Okada, S., Kurata, T., Mukai, T. and Yamanishi, K. (1993). Isolation of human herpesvirus 7 form a child with symptoms mimicking chronic Epstein-Barr virus infection. *Brit. J. Haematol.* **84**, 545-548.

Kaye, J., Browne, H., Stoffel, M. and Minson, T. (1992). The UL16 gene of human cytomegalovirus encodes a glycoprotein that is dispensable for growth *in vitro*. *J. Virol.* **66**, 6606-6615.

Keay, S. and Baldwin, B. (1991). Anti-idiotype antibodies that mimic gp86 of human cytomegalovirus inhibit viral fusion but not attachment. *J. Virol.* **65**, 5124-5128.

Kemble, G.W., McCormick, A.L., Pereira, L. and Mocarski, E.S. (1987). A cytomegalovirus protein with properties of herpes simplex virus ICP8: partial purification of the polypeptide and map position of the gene. *J. Virol.* **61**, 3143-3151.

Kennedy, P.G.E. (1987). Neurological complications of varicella zoster virus. In, *Infections of the nervous system.* pp. 177-208.Ed. by P.G.E. Kennedy and R.T. Johnson,. Butterworth and Co.

Kido, S., Kondo, K., Kondo, T., Morishima, T., Takahashi, M. and Yamanishi, K. (1990). Detection of human herpesvirus 6 DNA in throat swabs by polymerase chain reaction. *J. Med. Virol.* **32**, 139-142.

Kieff, E. and Liebowitz, D. (1990). Epstein-Barr virus and its replication. In: *Virology* (2nd edn), pp 1889-1921. B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman (eds), Raven press New York.

Kikuta, H., Nakane, A., Lu, H., Taguchi, Y., Minagawa, T. and Matsumoto, S. (1990). Interferon induction by human herpesvirus 6 in mononuclear cells. *J. Infect. Dis.* **162**, 35.

Kishi, M., Harada, H., Takahashi, M., Tanaka, A., Hayashi, M., Nonoyama, M., Josephs, S.F., Buchbinder, A., Schacter, F., Ablashi, D.V., Wong-Staal, F., Salahuddin, S.Z. and Gallo, R.C. (1988). A repeat sequence GGGTTA is shared by DNA of human herpesvirus 6 and Marek's disease virus. *J.Virol.*, **62**, 4824-4827.

Klages, S., Ruger, B. and Jahn, G. (1989). Multiplicity dependent expression of the predominant phosphoprotein pp65 of human cytomegalovirus. *Virus Res.* **12**, 159-168.

Knox, K.K. and Carrigan, D.R. (1994). Disseminated and active infections by human herpesvirus 6 in patients with AIDS. *Lancet* **343**, 577-578.

Kondo, K., Kondo, T., Okuno, T., Takahashi, M. and Yamanishi, K. (1991). Latent human herpesvirus 6 infection of human monocytes/macrophages. *J. Gen. Virol.*, **72**, 1401-1408.

Kondo, K., Nagafuji, H., Hata, A., Tomomori, C. and Yamanishi, K. (1993). Association of human herpesvirus 6 infection of the central nervous system with recurrence of febrile convulsions. *J. Infect. Dis.* **167**, 1197-1200.

Kouzarides, T., Bankier, A.T., Satchwell, S.C., Preddy, E. and Barrell, B.G. (1988). An immediate early gene of human cytomegalovirus encodes a potential membrane glycoprotein. *Virology* **165**, 151-164.

Krainer A.R. and Maniatis, T. (1988). RNA splicing. In: *Transcription and splicing: Frontiers in molecular biology*, 131. Hames, B.D. and Glover, D.M. (eds). IRL Press, Oxford.

Kramarsky, B. and Sander, C.H. (1992). Electron microscopy of human herpesvirus 6 (HHV-6). In: *Human herpesvirus 6: epidemiology, molecular biology, and clinical pathology*, pp 59-58. Ablashi, D.V., Krueger, G.R.F., Salahuddin, S.Z., (eds). Amsterdam: Elsevier.

Krech, U. (1973). Complement-fixing antibodies against cytomegalovirus in different parts of the world. *Bull. WHO* **49**, 103-106.

Krueger, G.R.F., Wassernam, K., deClerk, L.S., Steven, W.J., Bougeois, N., Ablashi, D.V., Josephs, S.F. and Balachandran, N. (1990). Latent human herpesvirus 6 in salivary and bronchial glands. *Lancet,* **336**, 1255-1256.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Lalezari, J.P., Drew, W.L., Glutzer, D.E., James, C., Miner, D., Flaherty, J., Fisher, P.E., Cundy, K., Hannigan, J. and Martin, J.C. (1995). (S)-1-[hydroxo-2-(phosphonylmathoxy)propyl]cytosine (Cidovir): results of a phase I/II study of a novel antiviral nucleotide analogue. *J. Infect. Dis.* **171**, 788-796.

Landini, M.P., Severi, B., Furlini, G. and Badiali, D.G.L. (1987). Human cytomegalovirus structural components: intracellular and intraviral localisation of p28 and pp65-69 by immunoelectron microscopy. *Virus Res.* **8**, 15-23.

Lawrence, G.L., Chee, M., Craxton, M.A., Gompels, U.A., Honess, R.W. and Barrell, B.G. (1990). Human herpesvirus 6 is closely related to human cytomegalovirus. *J. Virol.* **64**, 287-299.

Leach, C.T., Newton, E.R., McParlin, S. and Jenson, H.B., (1994). Human herpesvirus 6 infection of the female genital tract. *J. Infect. Dis.* **169**, 1281-1283.

Leahy, M.A., Krejci, S.M., Friednash, M., Stockert, S.S., Wilson, H., Huff, J.C., Weston, W.L. and Brice, S.L. (1993). Human herpesvirus 6 is present in lesions of Langerhans cell histocytosis. *J. Invest. Dermatol.* **101**, 642-645.

Levine, P.H., Jahan, N., Murari, P., Manak, M. and Jaffe, E.S. (1992). Detection of human herpesvirus 6 in tissues involved by sinus histocytosis with massive lymphadenopathy (Rosai-Dorfmann disease). *J. Infect. Dis.* **166**, 291-295.

Levy, J.A., Ferro, F., Greenspan, D. and Lennette, E.T. (1990). Frequent isolation of HHV-6 from saliva and high seroprevalence of the virus in the population. *Lancet,* **335**, 1047-1050.

Lindahl, T., Adams, A., Bjursell, G., Bornkamm, G.W., Kaschka-Dierich, C. and Jehn, U. (1976). Covalently closed circular duplex DNA of Epstein-Barr virus in a human lymphoid cell line. *J. Mol. Biol.* **102**, 511-530.

Lindahl, T., Klein, G., Reedman, B.N., Johansson, B. and Singh, S. (1974). Relationship between Epstein-Barr virus (EBV) DNA and the EBV-determined nuclear antigen (EBNA) in Burkitt lymphoma biopsies and other lymphoproliferative malignancies. *International J. Cancer* **13**, 764.

Linde, A., Dahl, H., Wahren, B., Friddel, E., Salahuddin, Z., Biberfeld, P. and Wahren, B. (1988). IgG antibodies to human herpesvirus 6 in children and adults and in primary Epstein-Barr virus infections and cytomegalovirus infections. *J. Virol. Methods,* **21**, 117-123.

Linde, A., Fridell, E., Dahl, H., Anderson, J., Biberfeld, P. and Wahren, B. (1990). Effect of primary Epstein-Barr virus infection on human herpesvirus 6, cytomegalovirus, and measles virus immunoglobulin G titers. *J. Clin. Microbiol.* **28**, 211-215.

Lindquister, G.J. and Pellett, P.E. (1991). Properties of the human herpesvirus 6 strain Z29 genome; G+C content, length, and presence of variable-length directly repeated terminal sequence elements. *Virology,* **182**, 102-110.

Littler, E., Stuart, A.D. and Chee, M.S. (1992). Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir. *Nature* **358**, 160-162.

Liu, B. and Stinski, M.F. (1992). Human cytomegalovirus contains a tegument protein that enhances trancription from promoters with upstream ATF and AP-1 cis-acting elements. *J. Virol.* **66**, 4434-4444.

Lopez, C. and Honess, R. (1990). Human herpesvirus 6. In, *Virology,* 2nd edn, Vol. 2, pp2055-2062 B. N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman, (eds). Raven Press, New York.

Lopez, C., Pellett, P.E., Stewart, J., Goldsmith, C., Sanderlin, K., Black, J., Warfield, D. and Feorino, P. (1988). Characteristics of human herpesvirus 6. *J. Infect. Dis.,* **157**, 1271-1273.

Luppi, M., Barozzi, P., Maiorana, A., Marasca, R. and Torelli, G. (1994). Human herpesvirus 6 infection in normal human brain tissue. *J. Infect. Dis.* **169**, 943-944.

Luppi, M., Marasca, R. and Barozzi, P., (1993a). Frequent detection of human herpesvirus 6 sequences by polymerase chain reaction in paraffin-embedded lymph nodes from patients with angioimmunoblastic lymphadenopathy-like lymphoma. *Leuk. Res.* **17**, 1003-1011.

Luppi, M., Marasca, R., Barozzi, P., Ferrari, S., Ceccherini-Nelli, L., Batoni, G., Merelli, E. and Torelli, G. (1993b). Three cases of human herpesvirus 6 latent infection: integration of viral genome in peripheral blood mononclear cell DNA. *J. Med. Virol.* **40**, 44-52.

Lusso, P., DeMaria, A., Malnati, M. and Lori, F. (1991). Introduction of CD4 and susceptibility to HIV-1 infection in human $CD8^+$ T lymphocytes by human herpesvirus 6. *Nature* **349**, 533-535.

Lusso, P., Ensoli, B., Markham, P.D., Ablashi, D.V., Salahuddin, S.Z., Tschachler, E., Wong-Staal, F. and Gallo, R.C. (1989). Productive dual infection of human $CD4^+$ T lymphocytes by HIV-1 and HHV-6. *Nature* **337**, 370-373.

Lusso, P., Markham, P.D., Tschachler, E., DiMarzo Veronese, F., Salahuddin, S.Z., Ablashi, D.V., Pahwa, S. and Gallo, R.C. (1988). *In vitro* cellular tropism of human B-lymphotropic virus (human herpesvirus 6). *J. Exp. Med.* **167**, 1659-1670.

Lusso, P., Secchiero, P., Crowley, R.W., Gerzino-Demo, A., Berneman, Z.N. and Gallo, R.C. (1994). CD4+ is a critical component of the receptor for human herpesvirus 7: interference with human immunodeficiency virus. *Proc. Natl Acad. Sci. USA.*, **91**, 3872-3876.

Marshall, G.S., Gesser, R.M., Yamanishi, K. and Starr, S.E. (1991). Chronic fatigue in children: clinical features, Epstein-Barr virus and human herpesvirus 6 serology and long term follow-up. *Pediatr. Infect. Dis.* **10**, 287-290.

Martin, M.E.D., Thomson, B.J., Honess, R.W., Craxton, M.A., Gompels, U.A., Liu, M.Y., Littler, E., Arrand, J.R., Teo, I. and Jones, M.D. (1991). The genome of human herpesvirus-6 - maps of unit-length and concatemeric genomes for 9 restriction endonucleases. *J. Gen. Virol.* **72**, 157-168.

Martinez, J., Lahijani, R.S. and St. Jeor, S.C. (1989). Analysis of a region of the human cytomegalovirus (AD169) genome coding for a 25-kilodalton virion protein. *J. Virol.* **63**, 233-241.

Masse, M.J., Karlin, S., Schachtel, G.A. and Mocarski, E.S. (1992). Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region. *Proc. Natl. Acad. Sci. USA* **89**, 5246-5250.

McGeoch, D,J. and Davison, A.J. (1995). Origins of DNA viruses. In: *Molecular evolution of viruses.* A.J. Gibbs, F. Garcia-Arenal, C. Callisher eds. Cambridge University press. p67-75.

McGeoch, D.J. (1989). The genomes of the human herpesviruses: contents, relationships, and evolution. *Annu. Rev. Microbiol.* **43**, 235-265.

McGeoch, D.J. (1990). Evolutionary relationships of virion glycoprotein genes in the S region of alphaherpesvirus genomes. *J. Gen. Virol.*, **71**, 2361-2367.

McGeoch, D.J. (1992). Molecular evolution of large DNA viruses of eukaryotes. *Seminars Virol.* **3**, 399-408.

McGeoch, D.J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphate family. *Nucleic Acids Res.* **18**, 4105-4110.

McGeoch, D.J. and Barnett, B.C. (1991). Neurovirulence factor. *Nature* **353**, 609.

McGeoch, D.J. and Cook, S. (1994). Molecular phylogeny of the Alphaherpesvirinae subfamily and a proposed evolutionary timescale. *J. Mol. Biol.* **238**, 9-22.

McGeoch, D.J. and Schaffer, P.A. (1993). In *Genetic Maps. Locus maps of complex genomes. Sixth Edition,* 1.147-1.156. ed. by S. O'Brien. Cold Spring Harbour Press, New York.

McGeoch, D.J., Barnett, B.C. and MacLean, C.A. (1993). Emerging functions of alphaherpesvirus genes. *Seminars in Virology* **4**, 124-133.

McGeoch, D.J., Cook, S., Dolan, A., Jamieson, F.E. and Telford, E.A. (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J. Mol. Biol.* **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. and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* **69**, 1531-1574.

Mellerick, D.M. and Fraser, N.W. (1987). Physical state of the latent herpes simplex virus genome in mouse model system. Evidence suggesting an episomal state. *Virology* **158**, 265-275.

Messerle, M., Keil, G.M. and Koszinowski, U.H. (1991). Structure and expression of murine cytomegalovirus immediate-early gene 2. *J. Virol.* **65**, 1638-1643.

Meyer, H., Bankier, A.T. and Landini, M.P. (1988). Identification and procaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. *J. Virol.* **62**, 2243-2250.

Meyers, J.D., Flournoy, N. and Thomas, E.D. (1986). Risk factors for cytomegalovirus infections after human marrow transplant. *J. Infect. Dis.* **153**, 478-488.

Miller, G. (1990). Epstein-Barr virus: Biology, pathogenesis and medical aspects. In: *Virology,* 2nd edn. pp 1921-1958. B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman (eds), Raven Press, New York.

Mocarski, E.S. (1993). Cytomegalovirus biology and replication. In: Roizman, B., Whitley, R., Lopez, C., (eds). *The human herpesviruses.* New York: Raven; 173-226.

Mocarski, E.S. (1996). Cytomegalovirus and their replication. In, *Virology* 2nd edn, ed. by B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman, Vol. 3, pp. 2447-2492. Raven Press, New York.

Mocarski, E.S. and Roizman, B. (1982). Structure and role of the herpes simplex virus DNA termini on inversion, circularization and generation of virion DNA. *Cell* **31**,89-97.

Moore, P., Gao, S.J., Dominguez, G., Cesarman, E., Lungo, O., Knowles, D.M., Garber, R., Pellett, P.E., McGeoch, D.J. and Chang, Y. (1996). Primary characterization of a herpesvirus agent associated with Kaposis-sarcoma. *J. Virol.* **70**, 549-558.

Morgan, C., Rose, H.M., Holden, M. and Jones, E.P. (1959). Electron microscopic observations on the herpes simplex virus. *J. Exp. Med.,* **110**, 643-656.

Morrison, J.M. (1991). *Virus induced enzymes.* Wiley, Chichester.

Mukai, T., Yamamoto, T., Kondo, T., Okuno, T., Kosuge H. and Yamanishi, K. (1994). Molecular epidemiological studies of human herpesvirus 6 in families. *J. Med. Virol.* **42**, 224-227.

Mullaney, J., Moss, H.W.McL. and McGeoch, D.J. (1989). Gene UL2 of herpes simplex virus type 1 encodes a uracil-DNA glycosylase. *J. Gen. Virol.* **70**, 449-454.

Nahmias, A.J., Keyserling, H. and Lee, F.K. (1991). Herpes simplex viruses 1 and 2: Epidemiology and control. In: *Viral infections of humans*, 3rd edition. Chapter 14, pp 393-418. Ed A. Evans. Plenum press, New York and London.

Navarro, D., Paz, P., Tugizov, S., Topp, K., La Vail, J. and Pereira, L. (1993). Glycoprotein B of human cytomegalovirus promotes virion penetration into cells, transmission of infection from cell to cell, and fusion of infected cells. *Virology* **197**, 143-158.

Neiderman, J., Miller, G., Pearson, H. and Pagano, J. (1976). Patterns of excretion of Epstein-Barr virus in saliva and other oropharyngeal sites during infectious mononucleosis. *N. Eng. J. Med.,* **294**, 1355-1359.

Neipel, F., Ellinger, K. and Fleckenstein, B. (1991). The unique region of the human herpesvirus 6 genome is essentially collinear with the UL segment of human cytomegalovirus. *J. Gen. Virol.*, **72**, 2293-2297.

Nicholas, J. (1994). Nucleotide sequence analysis of a 21-kbp region of the genome of human herpesvirus 6 containing homologues of human cytomegalovirus major immediate-early and replication genes. *Virology* **204**, 738-750.

Nicholas, J. (1996). Determination and analysis of the complete nucleotide sequence of human herpesvirus 7. *J. Virol.* **70**, 5975-5989.

Nicholas, J. and Martin, M.E.D. (1994). Nucleotide sequence of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. *J. Virol.*, **68**, 597-610.

Okada, K., Ueda, K. and Kusuhara, K., (1993). Exanthem subitum and human herpesvirus 6 infection: clinical observations in fifty-seven cases. *Pediatr. Infect. Dis. J.* **12**, 204-208.

Okuno, T., Higashi, K., Shiraki, K., Yamanishi, K., Takahashi, M., Kokado, Y., Ishibashi, M., Takahara, S., Sonoda, T. and Tanaka, K. (1990). Human herpesvirus 6 infection in renal transplantation. *Transplantation* **49**, 519-522.

Okuno, T., Mukai, T., Baba, K., Ohsumi, Y., Takahashi, M. and Yamanishi, K. (1991). Outbreak of exanthem subitum in an orphanage. *J. Pediatr* **119**, 759-761.

Okuno, T., Oishi, H., Hayashi, K., Nonogaki, M., Tanaka K. and Yamanishi, K. (1995). Human herpesvirus 6 and 7 in cervixes of pregnant women. *J. Clin. Microbiol.* **33**, 1968-1970.

Okuno, T., Takahashi, K., Balachandra, K., Shiraki, K., Yamanishi, K., Takahashi, M., and Baba, K. (1989). Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J. Clin. Microbiol.* **27**, 651-653.

Oppenheim, J.J., Zachariae, C.O.C., Mukaida, N. and Matsushima, K. (1991). Properties of the novel proinflammatory supergene "intercrine" cytokine family. *Annu. Rev. Immunol.* **9**, 617-648.

Parker, C.A. and Weber, J.M. (1993). An enzyme-linked immunosorbent assay for the detection of IgG and IgM antibodies to human herpesvirus type 6. *J. Virol. Methods* **41**, 265-276.

Pellett P.E. and Black, J.B. (1996). Human herpesvirus 6. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 3rd edition, pp. 2587-2608. Lippincott-Raven, Philadelphia.

Pereira, L., Hoffman, M., Tatsuno, M. and Dondero, D. (1984). Polymorphism of human cytomegalovirus glycoproteins characterised by monoclonal antibodies. *Virology* **139**, 73-86.

Pfeiffer, B., Thomson, B., and Chandran, B. (1995). Identification and characterization of a cDNA derived from multiple splicing that encodes envelope glycoprotein gp105 of human herpesvirus 6. *J. Virol.* **69**, 3490-3500.

Pietroboni, G.R., Harnett, G.B. and Bucens, M.R. (1989). Centrifugal enchancement of human immunodeficiency virus (HIV) and human herpesvirus type 6 (HHV-6) infection *in vitro*. *J. Virol. Methods* **24**, 85-90.

Pietroboni, G.R., Harnett, G.B., Bucens, M.R. and Honess, R.W. (1988). Antibody to human herpesvirus 6 in saliva. *Lancet,* **1**, 1059.

Plummer, G., Bowling, C.P. and Goodheart, C.R. (1969). Comparisons of four horse herpesviruses. *J. Virol.,* **4**, 738-741.

Portolani, M., Cermelli, C., Mirandola, P. and Luca, D.D. (1995). Isolation of human herpes virus 7 from an infant with febrile syndrome. *J. Med. Virol.* **45**, 282-283.

Pruksananonda, P., Hall, C.B., Insel, R.A., McIntyre, K., Pellett, P.E., Long, C.E., Schnabel, K.C., Pincus, P.H., Stamey, F.R. and Dambaugh, T.R. (1992). Primary human herpesvirus 6 infection in young children. *N. Engl. J. Med.* **326**, 1445-1450.

Rasmussen, L. (1990). Immune response to human cytomegalovirus infection. *Curr. Top. Microbiol. Immunol.* **154**, 221-254.

Rawlinson, W.D., Farrell, H.E., and Barrell, B.G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *J. Virol.* **70**, 8833-8849.

Razzaque, A. (1990). Oncogenic potential of human herpesvirus 6 DNA. *Oncogene* **5**, 1365-1370.

Razzaque, A., Williams, O., Wang, J. and Rhim, J.S. (1993). Neoplastic transformation of immortalised human epidermal keratinocytes by two HHV-6 DNA clones. *Virology* **195**, 113-120.

Read, R., Larson, E., Harvey, J., Edwards, A., Thomson, B., Briggs, M. and Fox, J. (1990). Clinical and laboratory findings in the Paul-Brunnell negative glandular fever-fatigue syndrome. *J. Infect.* **21**, 157-165.

Reeves, W.C., Pellett, P.E. and Gary, H.Jr. (1992). The chronic fatigue syndrome controversy. *Ann. Int. Med.* **117**, 343.

Rixon, F.J. (1993). Structure and assembly of herpesviruses. *Seminars in Virology* **4**, 135-144.

Rixon, F.J. and McGeoch, D.J. (1984). Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. *Nucleic Acids Res.* **13**, 953-973.

Rock, D.L. (1993). The molecular basis of latent infections by alphaherpesviruses. *Seminars in Virology* **4**, 157-165.

Rock, D.L. and Fraser, N.W. (1983). Detection of HSV-1 genome in the central nervous system of latently infected mice. *Nature* **302**, 523-525.

Rock, D.L. and Fraser, N.W. (1985). Latent herpes simplex virus type 1 DNA contains two copies of the virion DNA joint region. *J. Virol.* **55**, 849-852.

Roffman, E. and Frenkel, N. (1990). Interleukin-2 inhibits the replication of human herpesvirus 6 in mature thymocytes. *Virology* **175**, 591-594.

Roffman, E., Albert, J., Goff, J. and Frenkel, N. (1990). Putative site for the acquisition of human herpesvirus 6 virion tegument. *J. Virol.,* **84, 617-618.**

Roizman, B. (1979). The structure and isomerisation of herpes simplex virus genomes. *Cell* **16**, 481-494.

Roizman, B. and Furlong, D. (1974). The replication of herpesvirus. In, *Comprehensive virology,* ed. H. Fraenkel-Conrat and R.R. Wagner, **vol. 3**, pp. 229-403. Plenum, New York.

Roizman, B. and Sears, A.E. (1990). Herpes simplex viruses and their replication. In: *Virology* 2nd Ed. Chr 65, pp 1795-1841. B.N. Fields, D.M. Knipe et al., (eds). Raven, New York.

Roizman, B. and Sears, A.E. (1993). Herpes simplex viruses and their replication. In, *The human herpesviruses,* pp11-68. Edited by Roizman *et al.* Raven Press Ltd., New York.

Roizman, B., Carmichael, L.E., Deinhardt, F., de-The, G., Nahmias, A.J., Plowright, W., Rapp, F., Sheldrick, P., Takahashi, M. and Wolf, K. (1981). Herpesviridae. Definition, provisional nomenclature, and taxonomy. *Intervirology* **16**, 201-217.

Roizman, B., Desrosiers, R.C., Fleckenstein, B., Minson, A.C. and Studdert, M.J. (1992). The family herpesviridae: an update. *Arch. Virol.* **123**, 425-449.

Rowe, W.P., Hartley, J.W., Waterman, S., Turner, H.C. and Hueber, R.J. (1956), Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. *J. Virol.* **64**, 5167-5172.

Russler, S.K., Tapper, M.A. and Carrigan, D.R. (1989). Susceptibility of human herpesvirus 6 to acyclovir and ganciclovir. *Lancet* **2**, 382.

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. and Moore, P.S. (1996). Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV-8). *Proc. Natl. Acad. Sci. USA.* **93**, 14862-14867.

Ruvulo, V.R., Berneman, Z., Secchiero, P., and Nicholas, J. (1996). Cloning, restriction endonuclease mapping and partial sequence analysis of the genome of human herpesvirus 7 strain JI. *J. Gen. Virol.* **77**, 1901-1912.

Sada, E., Yasukawa, M., Ito, C., Takeda, A., Shiosaka, T., Tanioka, H. and Fujita S. (1996). Detection of human herpesvirus 6 and human herpesvirus 7 in the submandibular gland, parotid gland, and lip salivary gland by PCR. *J. Clin. Micro.* **34**, 2320-2321.

Salahuddin, S.Z., Ablashi, D.V., Markham, P.D., Josephs, S.F., Sturzenegger, S., Kaplan, M., Halligan, G., Biberfeld, P., Wong-Staal, F., Karmarsky, B. and Gallo, R.C. (1986). Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. *Science,* **234**, 596-601.

Sample, J.T. (1994). Epstein-Barr virus: molecular biology. In: *Encyclopedia of virology*, vol 1. pp 410-416. Edited by R.G. Webster and A. Granoff. Academic Press.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.

Sarov, I. and Abady, I. (1975). The morphogenesis of human cytomegalovirus. Isolation and polypeptide characterisation of cytomegalovirus and dense bodies. *Virology* **66**, 464-473.

Saxinger, C., Polesky, H., Eby, N., Grufferman, S., Murphy, R., Tegtmeir, G., Parekh, V., Memon, S. and Hung, C. (1988). Antibody reactivity with HBLV (HHV-6) in U.S. populations. *J. Virol. Methods,* **21**, 199-208.

Schalling, M., Ekman, M., Kaaya, E.E., Linde, A. and Bieberfeld, P. (1995). A role for a new herpesvirus (KSHV) in different forms of Kaposi's sarcoma. *Natl. Med.* **1**, 707-708.

Schiewe, U., Neipel, F., Schreiner, D. and Fleckenstein, B. (1994). Structure and transcription of an immediate-early region in the human herpesvirus 6 genome. *J. Virol.* **68**, 2978-2985.

Schrimer, E.C., Wyatt, L.S., Yamanishi, K., Rodriguez, W.J. and Frenkel, N (1991). Differentiation between two distinct classes of viruses now classified as human herpesvirus 6. *Proc. Natl. Acad. Sci.USA.* **88**, 5922-5926,

Schrimpff, S., Serpick, A., Stoler, B., Rumack, B., Mellin, H. and Joseph, M. (1972). Varicella zoster infection in patients with cancer. *Annals of Internal Med.,* **76**, 241-254.

Schwyzer, M., Styger, D., Vogt, B., Lowery, D.E., Simard, C., LaBoissiere, S., Misra, V., Vleck, C. and Paces, V. (1996). Gene contents in a 31-kb segment at the left genome end of bovine herpesvirus-1. *Vet. Microbiol.* **53**

Secchiero, P., Nicholas, J., Deng, H, Xiaopeng, T., van Loon, N., Ruvulo, V.R., Berneman, Z.N., Reitz, M.S. Jr. and Dewhurst, S. (1995). Identification of human telomeric repeat motifs at the genome termini of human herpesvirus 7: structural analysis and heterogeneity. *J. Virol.* **69**, 8041-8045.

Sieczkowski, L., Chandran, B. and Wood, C. (1995). The human immunodeficiency virus *tat* gene enchances replication of human herpesvirus 6. *Virology* **211**, 544-553.

Singer, O. and Frenkel, N. (1997). Human herpesvirus 7 (HHV-7) DNA: analyses of clones spanning the entire genome. *Arch. Virol.* **142**, 287-303.

Sixbey, J.W., Nedrud, J.G., Raab-Taub, N., Hanes, R.A. and Pagano, J.S. (1984). Epstein-Barr virus replication in oropharyngeal cells. *N. Engl. J. Med.*, **310**, 1225-1230.

Sixbey, J.W., Vesterinen, E.H., Nedrud, J.G., Raab-Traub, N., Walton, L.A. and Pagano, J.S. (1983). Replication of Epstein-Barr virus in human epithelial cells infected *in vitro*. *Nature,* **306**, 480-483.

Smiley, M.L., Wlodaver, C.G., Grossman, R.A., Barker, C.F., Perloff, L.J., Tustin, M.B., Starr, S.E., Plotkin, S.A. and Freidman, H.M. (1985). The role of pretransplant immunity in protection from cytomegalovirus disease following renal transplantation. *Transplantation* **40**, 157-161.

Smith, J.D. and de Harven, E (1974). Herpes simplex virus and human cytomegalovirus replication in WI-38 cells. II. An ultrastructural study of viral penetration. *J. Virol.* **14**, 945-956.

Smith, J.D. and de Harven, E (1978). Herpes simplex virus and human cytomegalovirus replication in WI-38 cells. III. Cytochemical localisation of lysosomal enzymes in infected cells. *J. Virol.* **26**, 102-109.

Smith, M. G. (1956). Propagation in tissue culture of a cytopathogenic virus from human salivary gland virus disease. *Proc. Soc. Exp. Biol. Med.* **92**, 424-430.

Sobue, R., Miyazaki, H. and Okamoto, M., (1991). Fulminant hepatitis in primary human herpesvirus 6 infection.*N. Engl. J. Med.* **324**, 1290.

Sola, P., Merelli, E., Marasca, R., Poggi, M., Luppi, M., Montorsi, M. and Torelli, G. (1993). Human herpesvirus 6 and multiple sclerosis: survey of anti-HHV-6 antibodies by immunofluorescence analysis and of viral sequences by polymerase chain reaction. *J. Neurol. Neurosurg. Psychiatry* **56**, 917-919.

Spaete, R.R. and Frenkel, N. (1985). The herpes simplex virus amplicon: analyses of cis-acting replication functions. *Proc. Natl. Acad. Sci. USA* **82**, 694-698.

Spaete, R.R., Perot, K., Scott, P.I., Nelson, J.A., Stinski, M.F. and Pachl, C. (1993). Coexpression of truncated human cytomegalovirus gH with the UL115 gene product or the truncated human fibroblast growth factor receptor results in transport of gH to the cell surface. *Virology* **193**, 853-861.

Spear, P.G. (1985). Glycoproteins specified by herpes simplex viruses. In: *The Herpesviruses* 3, pp 315-356. Roizman, B., (ed). Plenum Press, New York and London.

Spear, P.G. and Roizman, B. (1972). Proteins specified by herpes simplex virus. V. Purification and structural proteins of the herpes virion. *J. Virol.* **9**, 431-439.

Speck, S.T. and Strominger, J.L. (1989). In: *Advances in viral oncology* **8**, *Tumorigenic DNA viruses,* p. 133. Klein G (ed). Raven Press, New York.

Spector, S.A. Weigeist, T., Pollard, R.B., Dieterich, D.T., Samo, T., Benson, C.A., Busch, D.F., Freeman, W.R., Monatgue, P. and Kaplan, H.J. (1993). A randomised, controlled study of intravenous ganciclovir therapy for cytomegalovirus peripheral retinitis in patients with AIDS. AIDS clinical trial group and cytomegalovirus cooperative study group. *J. Infect. Dis.* **168**, 557-563.

Spector, S.A., Hirata, K.K. and Newman, T.R. (1984). Identification of multiple cytomegalovirus strains in homosexual men with acquired immunodeficiency syndrome. *J. Infect. Dis.* **150**, 953-956.

Spira, T.J., Bozeman, L.H., Sanderlin, K.C., Warfield, D.T., Feorino, P.M., Holman, R.C., Kaplan, J.E., Fishbein, D.B. and Lopez, C. (1990). Lack of correlation between human herpesvirus 6 infection and the course of human immunodeficiency virus infection. *J. Infect. Dis.* **161**, 567-570.

Spivack, J.G. and Fraser, N.W. (1987). Detection of herpes simplex virus type 1 transcripts during latent infection in mice. *J. Virol.* **61**, 3841.

Staden, R. (1987). Computer handling of DNA sequencing projects. In: *Nucleic Acid and protein sequence analysis: a practical approach,* pp173-217. Bishop, M.J. and Rawlings, C.J. (eds). IRL Press, Oxford.

Stagno, S. and Cloud, G.A. (1994). Working parents: the impact of day care and breast-feeding on cytomegalovirus infections in offspring. *Proc. Natl. Acad. Sci.* **91**, 2384-2389.

Stagno, S., Pass, R.F., Dworsky, M.E. and Alford, C.A. (1983). Congenital and perinatal cytomegaloviral infections. *Seminars in Perinatology* **7**, 31-42.

Stannard, L.M., Fuller, A.O. and Spear, P.G. (1987). Herpes simplex virus glycoproteins associated with different morphological entities projecting from the virion envelope. *J. Gen. Virol.*, **68**, 715-725.

Stasiak, P.C. and Mocarski, E.S. (1992). Transactivation of the cytomegalovirus ICP36 gene promoter requires the α gene product TRS1 in addition to IE1 and IE2. *J. Virol.* **66**, 1050-1058.

Steeper,T.A., Horwitz, C.A., Ablashi, D.V., Salahuddin, S.Z., Saxinger, C., Saltzman, R. and Schwartz, B. (1988). The spectrum of clinical and laboratory findings resulting from human herpesvirus 6 (HHV-6) in patients with mononucleosis-like illnesses not resulting from Epstein-Barr virus or cytomegalovirus. *Am. J. Clin. Pathol.* **93**

Steiner, I., Spivack, J.G., Lirette, R.P., Brown, S.M., MacLean, A.R., Subak-Sharpe, J.H. and Fraser, N.W. (1989). Herpes simplex virus type 1 latency-associated transcripts are evidently not essential for latent infection. *EMBO J.* **8**, 505-511.

Stenberg, M.F. and Kerry, J.A. (1995) Cytomegalovirus genes: their structure and function. *Scandinavian Journal of Infectious Diseases (Supplement)* **99**, 3-6.

Stenberg, R.M., Depto, A.S., Fortney, J. and Nelson, J.A. (1989). Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate-early gene region. *J. Virol.* **63**, 2699-2708.

Stenberg, R.M., Thomsen, D.R. and Stinski, M.F. (1984). Structural analysis of the major immediate early gene of human cytomegalovirus. *J. Virol.* **49**, 190-199.

Stenberg, R.M., Witte, P.R. and Stinski, M.F. (1985). Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1. *J. Virol.* **56**, 665-675.

Stinski, M.F., Malone, C.L., Herminston, T.W. and Liu, B. (1991). Regulation of human cytomegalovirus transcription. In Wagner EK (ed.) *Herpesvirus transcription and its control,* pp 245-260. Boca Raton: CRC Press.

Stinski, M.F., Thomsen, D.R., Stenberg, R.M. and Goldstein, L.C. (1983). Organization and expression of the immediate early genes of human cytomegalovirus. *J. Virol.* **46**, 1-14.

Streicher, H.L., Hung, C.L., Ablashi, D.V., Hellman, K., Saxinger, C., Fullen, J. and Salahuddin, S.Z. (1988). *In vitro* inhibition of human herpesvirus 6 by phosphonoformate. *J. Virol. Methods* **21**, 301-304.

Suga, S., Yoshikawa, T., Asano, Y., Kozawa, T., Nakashima, T., Kobayashi, T., Yazaki, T., Yamamoto, H., Kajita, Y. and Ozaki, T. (1993). Clinical and virological analyses of 21 infants with exanthem subitum (roseola infantum) and central nervous system complications. *Ann. Neurol.* **33**, 597-603.

Suga, S., Yoshikawa, T., Asano, Y., Nakashima, T., Yazaki, T., Fukuda, M., Kojima, S., Matsuyama, T. and Ono, Y. (1992). IgM neutralising antibody responses to human herpesvirus 6 in patients with exanthem subitum or organ transplantation. *Microbiol. Immunol.* **36**, 495-506.

Suga, S., Yoshikawa, T., Asano, Y., Yazaki, T. and Hirata, S. (1989). Human herpesvirus 6 infection (exanthem subitum) without rash. *Pediatrics* **83**, 1003-1006.

Sullivan, V., Biron, K.K. and Talarico, C. (1993). A point mutation in the human cytomegalovirus DNA polymerase gene confers resistance to ganciclovir and phosphonylmethoxyalkyl derivatives. *Antimicrob. Agents Chemother.* **37**, 19-25.

Sullivan, V., Talarico, C.L., Stanat, S.C., Davis, M., Coen, D.M. and Biron, K.K. (1992). A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus-infected cells. *Nature* **358**, 162-164.

Szilagyi, J.F. and Berriman, J. (1994). Herpes simplex virus L particles contain spherical membrane-enclosed inclusion vesicles. *J. Gen. Virol.*, **75**, 1749-1753.

Takahashi, K., Sonoda, S., Higashi, K., Kondo, T., Takahashi, H., Takahashi, M. and Yamanishi, K. (1989). Predominant CD4 T-lymphocyte tropism of human herpesvirus 6-related virus. *J. Virol.* **63**, 3161-3163.

Takahashi, K., Sonoda, S., Kawaquami, K., Milata, K., Oki, T., Nagata, T., Okuno, T. and Yamanishi, K. (1988). Human herpesvirus 6 and exanthem subitum. *Lancet,* **1**, 1463.

Tanaka, K., Kondo, T., Torigoe, S., Okada, S., Mukai, T., and Yamanishi, K. (1994). Human herpesvirus 7: another causal agent for roseola (exanthem subitum). *J. Pediatr.* **125**, 1-5.

Taylor, P. (1986). A computer program for translating DNA sequences into protein. *Nucleic Acids Res.* **14**, 437-441.

Taylor-Wiedeman, J., Sissons, J.G., Borysiewicz, L.K. and Sinclair, J.H. (1991). Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. *J. Gen. Virol.* **72**, 2059-2064.

Tedder, T.S., Briggs, M., Cameron, C.H., Honess, R.W., Robertson, D. and White, H. (1987). A novel lymphotropic herpesvirus. *Lancet* **2**, 390-391.

Telford, E.A.R., Watson, M.S., Aird, H.C., Perry, J. and Davison, A.J. (1995). The DNA sequence of equine herpes virus-2. *J. Mol. Biol.* **249**, 520-528.

Telford, E.A.R., Watson, M.S., McBride, K. and Davison, A.J. (1992). The DNA sequence of equine herpesvirus 1. *Virology,* **189**, 9-14.

Thompson, J., Choudhury, S., Kashanchi, F., Doniger, J., Berneman, Z., Frenkel, N. and Rosenthal, L.J. (1994). A transforming fragment within the direct repeat region of human herpesvirus type 6 that transactivates HIV-1. *Oncogene* **9**, 1167-1175.

Thompson, R., Honess, R.W., Taylor, L., Morran, J. and Davison, A.J. (1987). Varicella-zoster virus specifies a thymidylate synthase. *J. Gen. Virol.* **68**, 1449-1455.

Thomson, B.J., Dewhurst, S. and Gray, D. (1994). Structure and heterogeneity of the sequences of human herpesvirus 6 strain variants U1102 and Z29 and identification of human telomeric repeat sequences at the genomic termini. *J. Virol.* **68**, 3007-3014.

Thomson, B.J., Efstathiou, S. and Honess R.W. (1991). Acquisition of the human adeno-associated virus type-2 *rep* gene by human herpesvirus 6. *Nature,* **351**, 78-80.

Timbury, M.C. (1991). Herpes simplex virus infections. In: *Notes on medical virology* pp 84-98. Churchhill Livingstone.

Torelli, G., Marasca, R., Montorsi, M., Luppi, M., Barozzi, P., Ceccherini, L., Batoni. G., Bendinelli, M. and Muyombano, A. (1992). Human herpesvirus 6 in non-AIDS related Hodgkin's and non-Hodgkin's lymphomas. *Leukemia* **6**, 46S-48S.

Ueda, K., Kusuhara, K., Okada, K., Miyazaki, C., Hidaka, Y., Tokugawa, K. and Yamanishi, K. (1994). Primary human herpesvirus 7 infection and exanthem subitum. *Ped. Infect. Dis. J.* **13**, 167-168.

Van der Meer, J.T.M., Drew, W.L., Bowden, R.A., Galasso, G.J., Griffiths, P.D., Douglas, A.J., Katlama, C., Spector, S.A. and Whitley, R.J. (1996). Summary of the international consensus symposium on advances in the diagnosis, treatment, and prophylaxis of cytomegalovirus infection. *Antiviral Res.* **32**, 119-140.

Van Loon, N., Dykes, C., Deng, H., Dominguez, G., Nicholas, J., and Dewhurst, S. (1997). Identification and analysis of a lytic-phase origin of DNA replication in human herpesvirus 7. *J. Virol.* **71**, 3279-3284.

Van Zijl, M., Quint, W., Briaire, J., de Rover, T., Gielkens, A. and Berns, A. (1988). Regeneration of herpesviruses from molecularly cloned subgenomic fragments. *J. Virol.* **62**, 2191-2195.

Virgin, H.W. 4th., Latreille, P., Wamsley, P., Hallsworth, K., Weck, K.E., Dal Canto, A.J. and Speck, S.H. (1997). Complete sequence and genomic analysis of murine gammaherpesvirus 68. *J. Virol.* **71**, 5894-5904.

Vlazny, D.A., Kwong, A. and Frenkel, N. (1982). Site-specific cleavage/packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full-length viral DNA. *Proc. Natl. Acad. Sci. USA.* **79**, 1423-1427.

Wagner, E.K. (1985). Individual HSV transcripts. In: *The Herpesviruses,* vol 3 pp 45-104. B. Roizman (ed). Plenum Publishing Corp. New York.

Wakefield, D., Lloyd, A., Dwyer, J., Salahuddin, S.Z. and Ablashi, D.V. (1988). Human herpesvirus 6 and myalgic encephalomyelitis. *Lancet* **1**, 1059.

Ward, K.N. and Grey, J.J. (1994). Primary human herpesvirus 6 infection is frequently overlooked as a cause of febrile fits in young children. *J. Med. Virol.* **42**, 119-123.

Weller, T.H., Macaulay, J.C., Craig, J.M. and Wirth, P. (1957). Isolation of intranuclear inclusion producing agents from infants with illnesses resembling cytomegalic inclusion disease. *Proc. Soc. Exp. Biol. Med.* **94,** 4-12.

Weston, K. (1988). An enhancer element in the short unique region of human cytomegalovirus regulates the production of a group of abundant immediate early transcripts. *Virology* **162,** 406-416.

Weston, K. and Barrell, B.G. (1986). Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus. *J. Mol. Biol.* **192,** 177-208.

Whitley, R.J. (1985). Epidemiology of herpes simplex viruses. In: *The Herpesviruses* vol 3. B. Roizman (ed). Plenum Press, New York and London.

Whitley, R.J. (1990). Herpes simplex viruses. In: *Virology* (2nd edn), pp 1844-1877. B.N. Fields, D.M. Knipe, R.M. Chanock, M.S. Hirsch, J.L. Melnick, T.P. Monath and B. Roizman (eds), Raven Press New York.

Wilborn, F., Schmidt, C.A., Brinkmann, V., Jendroska, K., Oettle, H. and Siegert, W. (1994). A potential role for human herpesvirus type 6 in nervous system disease. *J. Neuroimmunology* **49**, 213-214.

Wildy, P., Russell, W.C. and Horne, R.W. (1960). The morphology of herpes virus. *Virol.* **12,** 204-222.

Williams, A.F. and Barclay, A.N. (1988). The immunoglobin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381-405.

Williamson, W.D., Demmler, G.J., Percy, A.K. and Catlin, F.I. (1992). Progressive hearing loss in infants with asymptomatic congenital cytomegalovirus infection. *Pediatrics* **90**, 862-866.

Wingard, J.R., Chen, D.Y., Burns, W.H., Fuller, D.J., Braine, H.G., Yeager, A.M., Kaiser, H., Burke, P.J., Graham, M.L. and Santos, G.W. (1988). Cytomegalovirus infection after autologous bone marrow transplantation with comparison to infection after allogeneic bone marrow transplantation. *Blood* **71**, 1432-1437.

Winston, D.J., Ho, W.G. and Champlin, R.E. (1990). Cytomegalovirus infection after allogenic bone marrow transplantation. *Rev. Infect. Dis.* **12**, 5776-5792.

Wright, D.A., Staprans, S.I. and Spector, D.H. (1988). Four phosphoproteins with common amino termini are encoded by human cytomegalovirus AD169. *J. Virol.* **62**, 331-340.

Wyatt, L.S. and Frenkel, N. (1992). Human herpesvirus 7 is a constitutive inhabitant of adult human saliva. *J. Virol.*, **66**, 3206-3209.

Wyatt, L.S., Balachandran, N. and Frenkel, N. (1990). Variations in the replication and antigenic properties of human herpesvirus 6 strains. *J. Infect. Dis.*, **162**, 852-857.

Wyatt, L.S., Rodrigez, W.J., Balachandran, N. and Frenkel, N. (1991). Human herpesvirus 7: antigenic properties and prevalence in children and adults. *J. Virol.*, **65**, 6260-6265.

Yadav, M., Chandrashekran, A., Vasudevan, D.M. and Ablashi, D.V. (1994). Frequent detection of human herpesvirus 6 in oral carcinoma. *J. Nat. Cancer Inst.* **86**, 1792-1794.

Yamanishi, K., Okuno, T., Shiraki, K., Takahashi, M., Kondo, T., Asano, Y. and Kurata, T. (1988). Identification of human herpesvirus-6 as a causative agent for exanthem-subitum. *Lancet* **1**, 1065-1067.

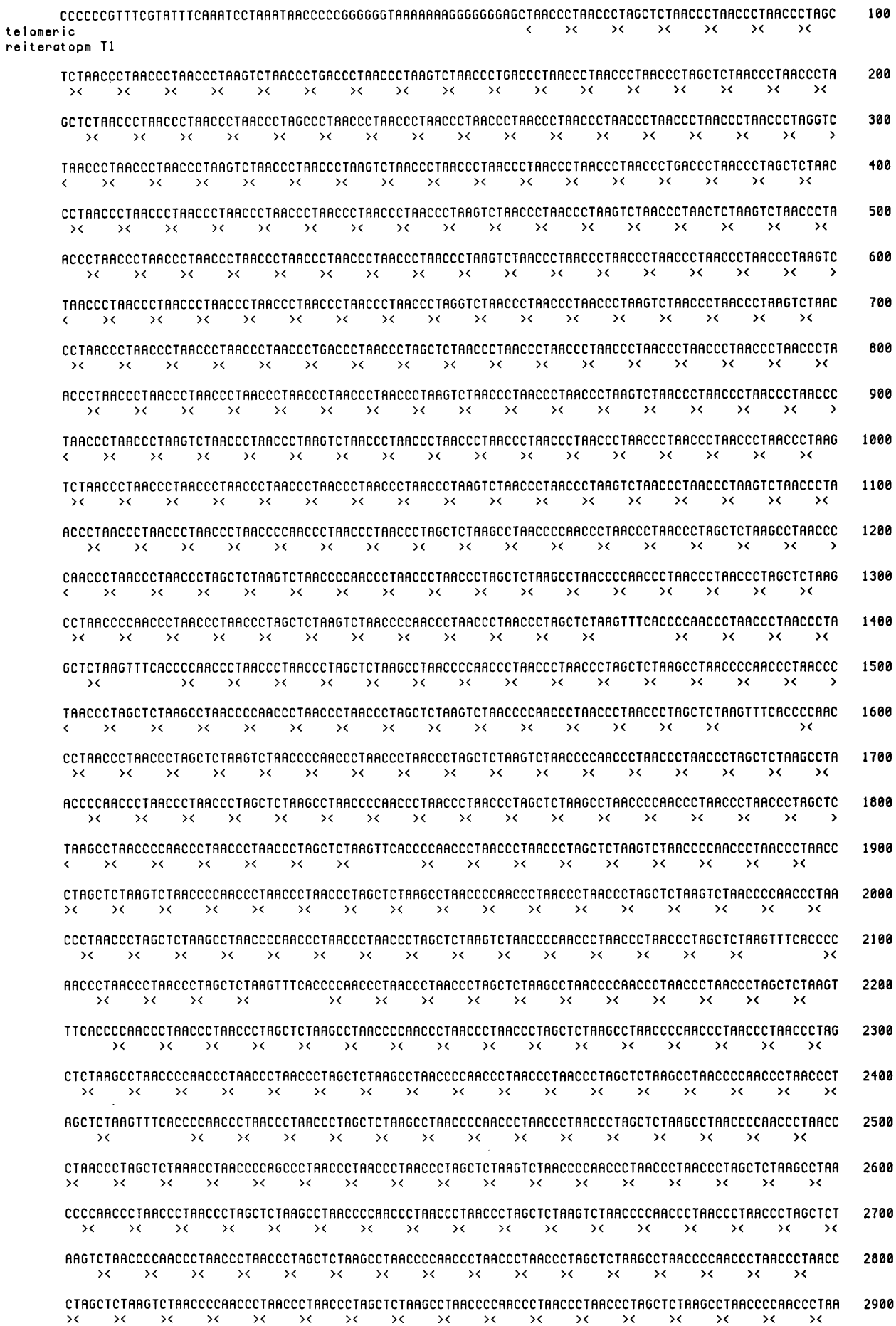
Yoshida, M., Uno, F., Bai, Z.L., Yamada, M., Nii, S., Sata, T., Kurata, T., Yamanishi, K. and Takahashi, M. (1989). Electron microscopic study of a herpesvirus-like virus isolated from an infant with exanthem subitum. *Microbiol. Immunol.*, **33**, 147-154.

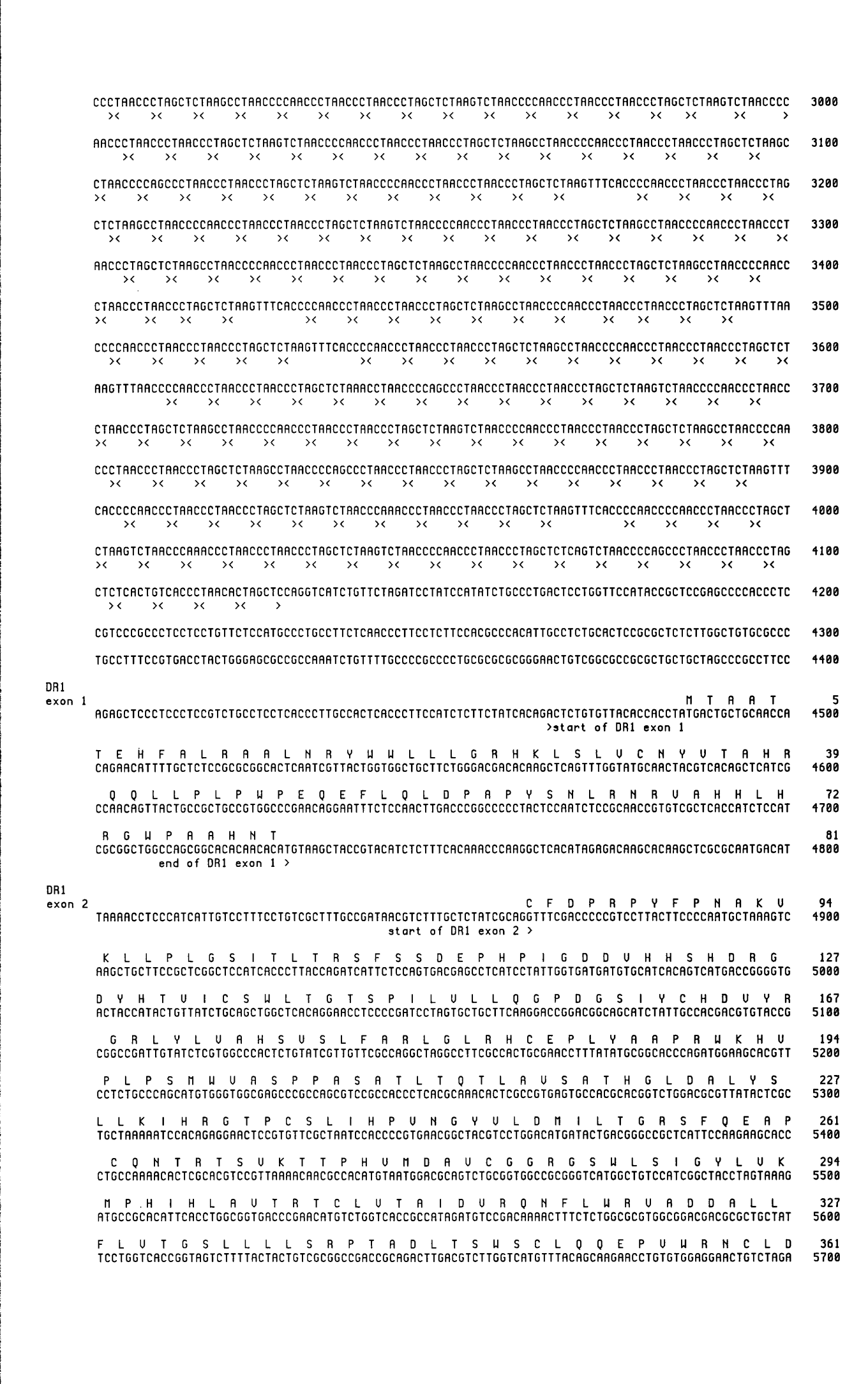
Yoshikawa, T., Asano, Y., Kobayashi, I., Nakashima, T., Yazaki, T., Suga, S., Ozaki, T., Wyatt, S. and Frenkel, N. (1993). Seroepidemiology of human herpesvirus 7 in healthy children and adults in Japan. *J. Med. Virol.*, **41**, 319-323.

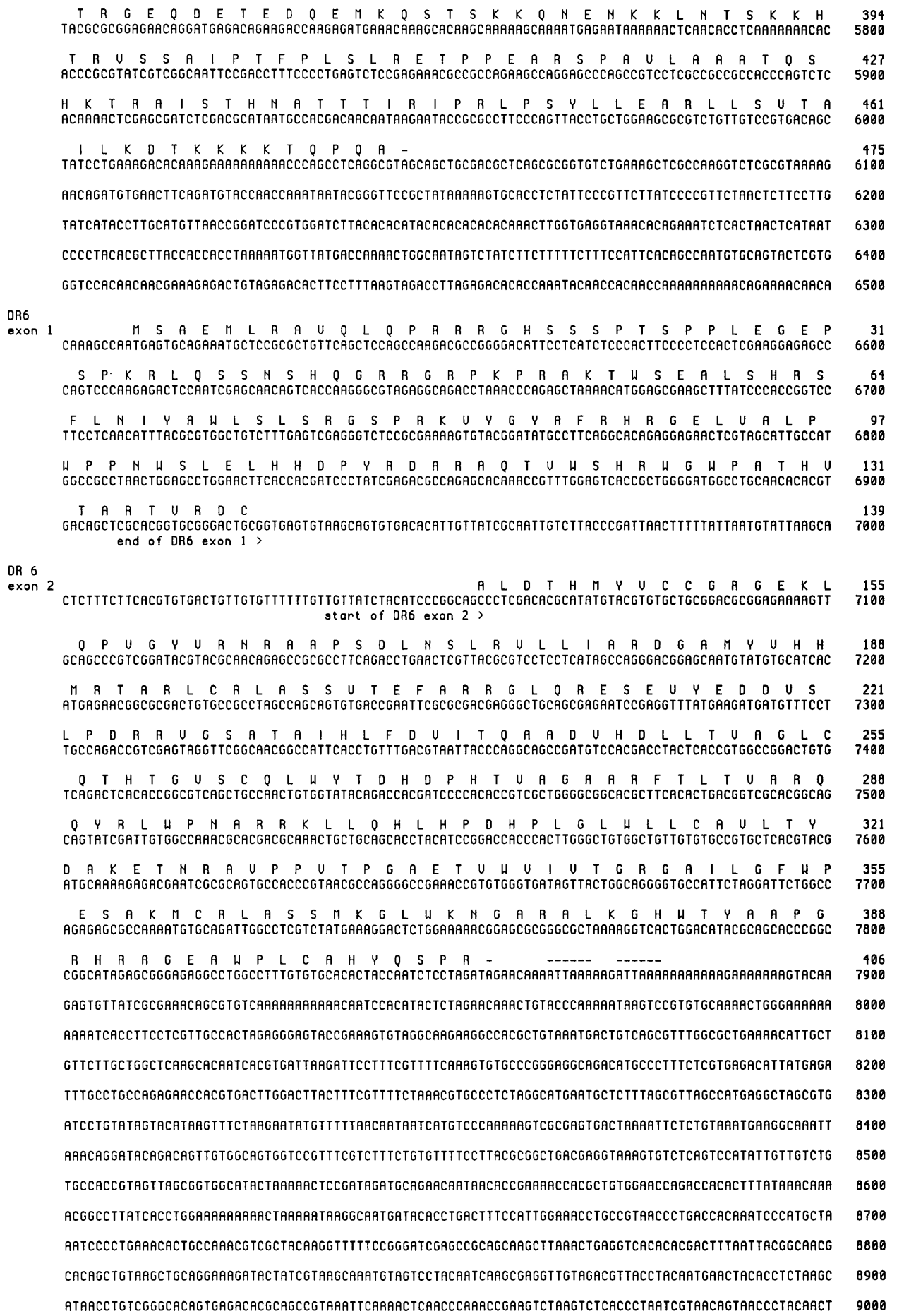
Yoshikawa, T., Suga, S., Asano, Y., Yazaki, T., Kodama, H. and Ozaki, T. (1989). Distribution of antibodies to a causative agent of exanthem subitum (human herpesvirus 6) in healthy individuals. *Pediatrics* **84**, 675-677.

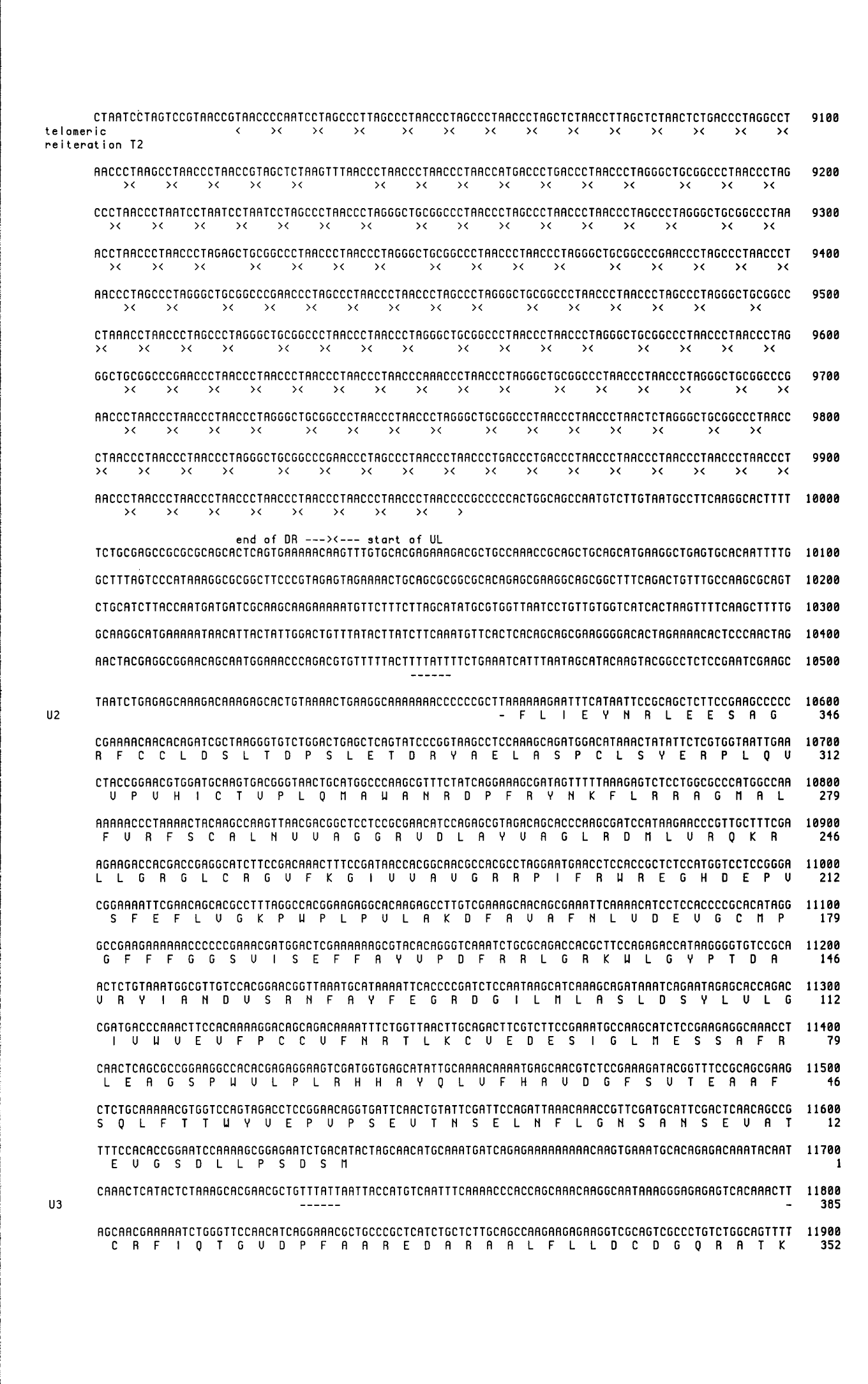
Zur Hausen, H. and Schultz-Hausen, H. (1970). Presence of EB virus nucleic acid homology in a "virus free" line of Burkitt tumour cells. *Nature* **227**, 245-248.

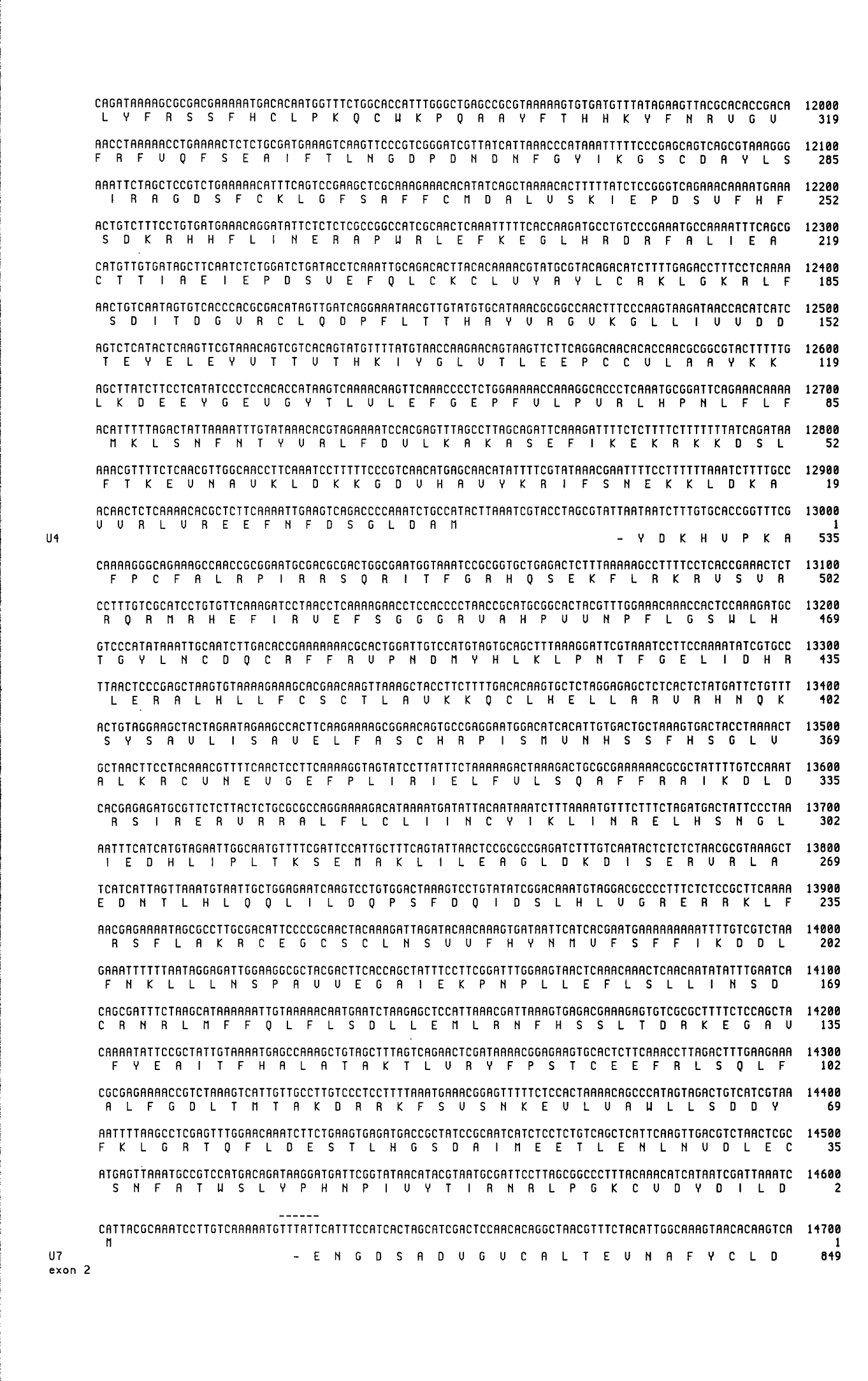
Appendix 1


telomeric
reiteratopm T1


DR1
exon 1
start of DR1 exon 1
end of DR1 exon 1 >
DR1
exon 2
start of DR1 exon 2 >


DR6
exon 1
end of DR6 exon 1 >
DR 6
exon 2
start of DR6 exon 2 >


telomeric reiteration T2
end of DR ---><--- start of UL
U2
U3

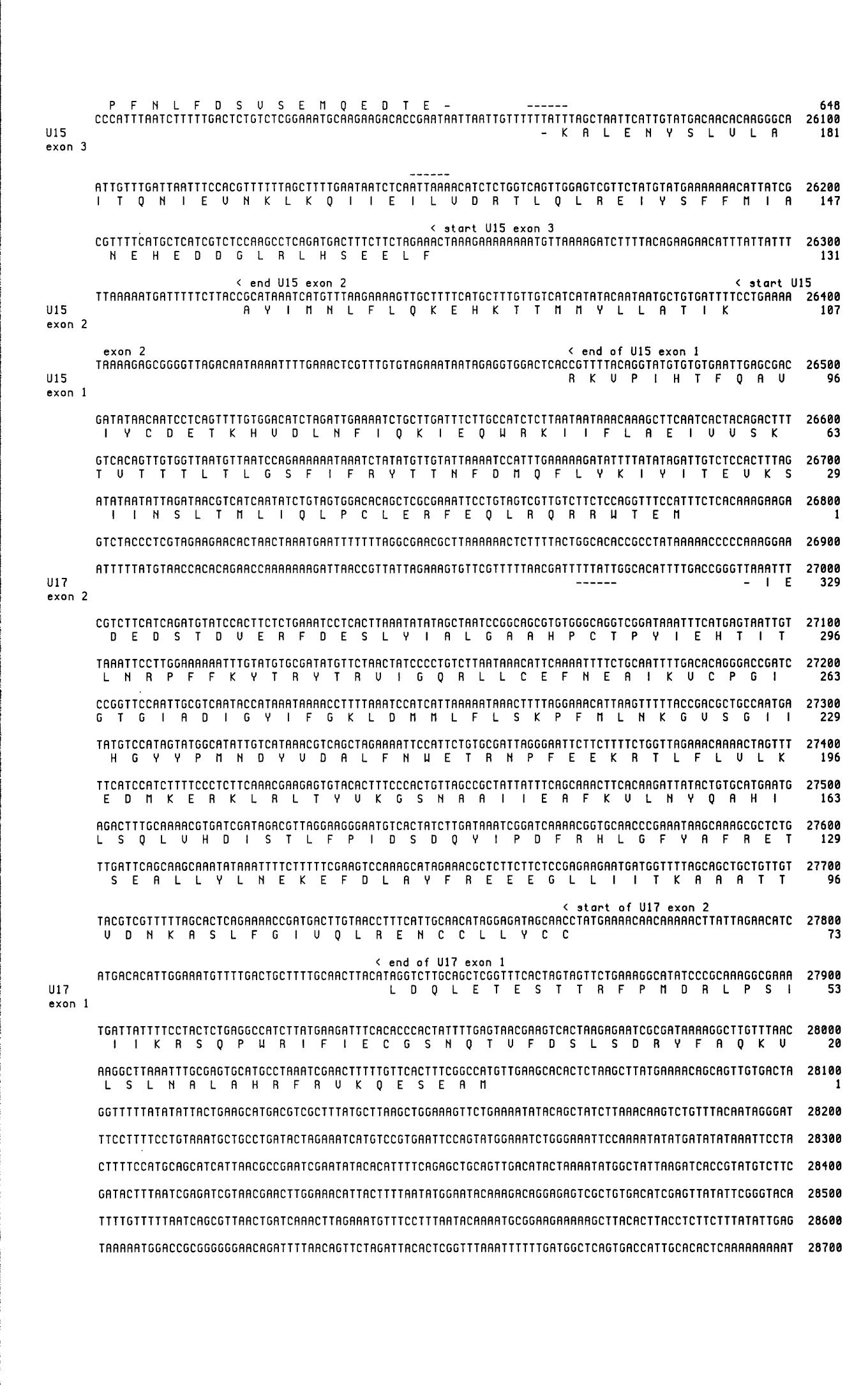

U4
U7
exon 2

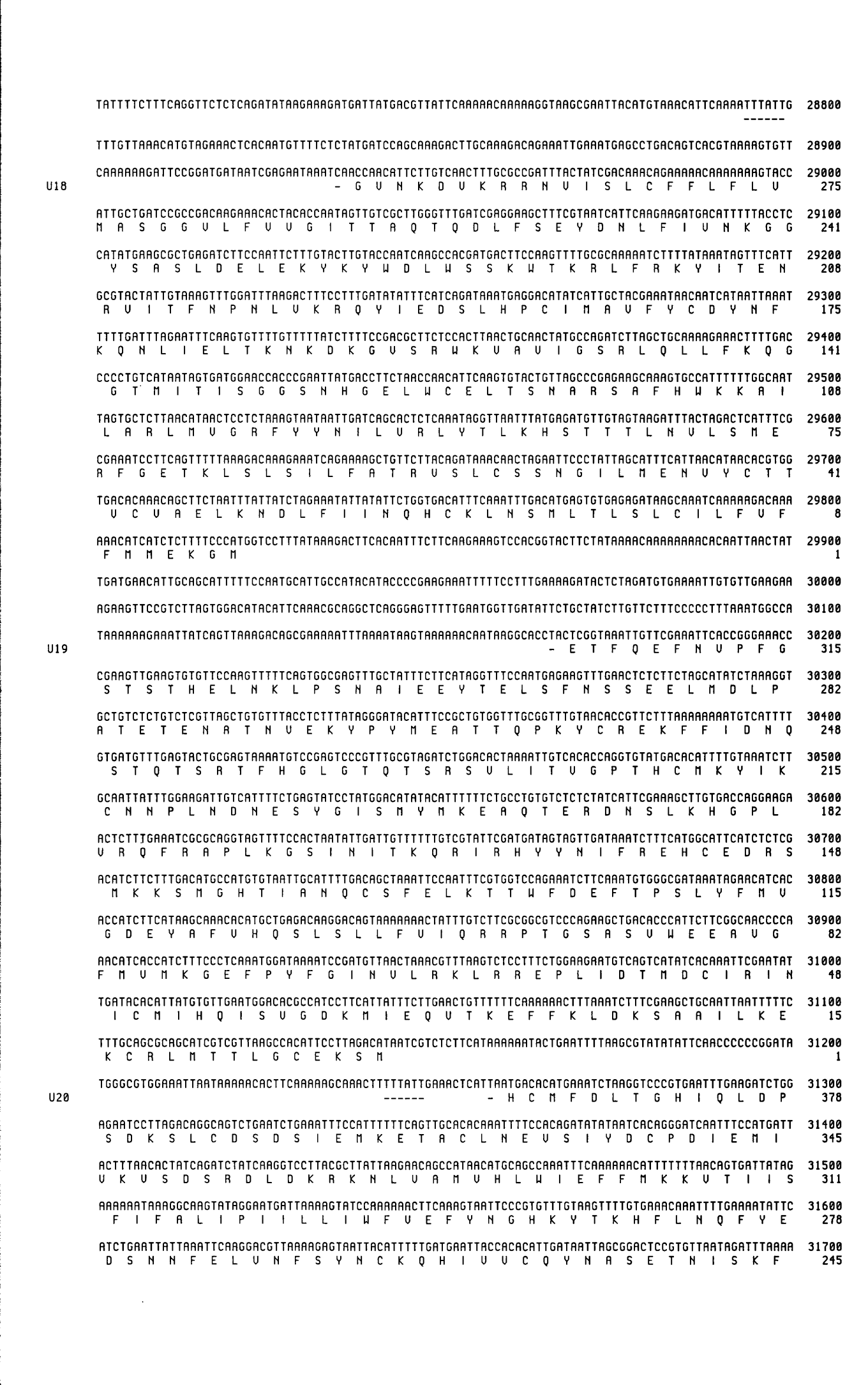
```
TTCTGAATCTTGTCAAGACCCCCCAGCTTATACCCTACAGCTCTTACATCTTTTTCTGGACTGGTCTGGAAATTTCCATTCCCAAAATTTCCAAAATATT 14800
N  Q  I  K  D  L  G  G  L  K  Y  G  V  A  R  V  D  K  E  P  S  T  Q  F  N  G  N  G  F  N  G  F  Y  K   816

TGAGACCACTGAAAAAAAGTGCCATTGCTCCATTAATGTTTTTTGTGTTACTTAAATATTCAGCTTCGATGTTCACGATCGATCTCTCCAGCACATCTAT 14900
  L  G  S  F  F  L  A  M  A  G  N  I  N  K  T  N  S  L  Y  E  A  E  I  N  V  I  S  R  E  L  V  D  I   782

GAAGCTCTTTTCCTTAGCATTCACTATGAAGAAACCTGGTGCTAGATTATTTGCAGCAACCAAGTCCGTGTGCTTCGCAGGCTGCCTTTGGATGTAAAGA 15000
  F  S  K  E  K  A  N  V  I  F  F  G  P  A  L  N  N  A  A  V  L  D  T  H  K  A  P  Q  R  Q  I  Y  L   749

TGATGCAAATAAAAGCTTTCCATGAAGTTTGTTTCATCATAGTTTTTGTAATACGGAATTTCAAAGCTCATCTTTATCACTTCTAAGCAGTTGAAAATAA 15100
H  H  L  Y  F  S  E  M  F  N  T  E  D  Y  N  K  Y  Y  P  I  E  F  S  M  K  I  V  E  L  C  N  F  I  F   682

AGTTTGAGGATGTGCCAGGGTTTCTGACAATATCTTTAACCTGCTTCAACATATTGAAATAAACGATATTGTTCATACAACCCTTACTGCCCAGTTTCAT 15200
  N  S  S  T  G  P  N  R  V  I  D  K  V  Q  K  L  M  N  F  Y  V  I  N  N  M  C  G  K  S  G  L  K  M   649

TTCACATTCCATGCTAGACTCCAAGTTCTTTAATTCATCTAAGTACACATCATAATACATTTTCAGTTCAGTAAACAAAACACTGTTAAGAGTTTGAATT 15300
  E  C  E  M  S  S  E  L  N  K  L  E  D  L  Y  V  D  Y  Y  M  K  L  E  T  F  L  V  S  N  L  T  Q  I   616

ATGAAAAGCAAAAACACAATGTTACTAAGAATATTCAACAGATTTCTTTTGTCCTGACCCATATCCACAGTAACCGAGATAGGAAACCTGGAATGTTCAA 15400
I  F  L  L  F  V  I  N  S  L  I  N  L  L  N  R  K  D  Q  G  M  D  V  T  V  S  I  P  F  R  S  H  E  I   582

TCTGCCCCTCAATAAAACTTAAACACAAATCTTTGGCTACAAACATAGAAAGATTAGAATATTTCTCAGATAAATGTGGAAAAATAGGTAAATCGATCTT 15500
  Q  G  E  I  F  S  L  C  L  D  K  A  V  F  M  S  L  N  S  Y  K  E  S  L  H  P  F  I  P  L  D  I  K   549

ACCCAAAGCTGGAAAGAAATGTAATTTGACAGAGCCAGGATTCTTTTTCATTGTGCCTACATTTTTAAACCTGAAAAAATTGTATCGACATTCTTCGCAA 15600
  G  L  A  P  F  F  H  L  K  V  S  G  P  N  K  K  M  T  G  V  N  K  F  R  F  F  N  Y  R  C  E  E  C   516

CCACAAAATAGTAAAAAGTGCTCAAGTAAGTACAACTGAAAACGATCAAAAAAATCGCGTCTGAAGAATGATTCTGTTTTGTTTTGCATTTCTGTCTTAA 15700
G  C  F  L  L  F  H  E  L  L  Y  L  Q  F  R  D  F  F  D  R  R  F  F  S  E  T  K  N  Q  M  E  T  K  F   482

AAGATTCGCTCCAAAATGCACAGCTAAAGCCCCCTAACTTCTGCTTAGATTCTTTACTCAATTCCTTCAATTTTCCATCTAGTTTAAAAAATAACATAGA 15800
  S  E  S  W  F  A  C  S  F  G  G  L  K  Q  K  S  E  K  S  L  E  K  L  K  G  D  L  K  F  F  L  M  S   449

AAGGTACTCACATAGGTAGCGACACCAGTATCCAATTCCAGGCTTAGCACCCAAATTAAACAATGATGCCCCAGAATAGAACCCTACAGCTGCAAGTTTA 15900
  L  Y  E  C  L  Y  R  C  W  Y  G  I  G  P  K  A  G  L  N  F  L  S  A  G  S  Y  F  G  V  A  A  L  K   416

GTAATGCACTGCCATTTGGGGTCAAAATCTTCCATAAATTTTAAATCGCGCAACTGCAAAAGAAATTTATCAGCCCTTGTGAAAGTTGATGGATACTTGA 16000
T  I  C  Q  W  K  P  D  F  D  E  M  F  K  L  D  R  L  Q  L  L  F  K  D  A  R  T  F  T  S  P  Y  K  F   382

ACATTTCGTACTCATAGTCTATCAACCTATTCCAAATAACACCATCTGAACATAATCTTAGCCCTTTTATTCCAAAGCGTCCAGCAGCTGATGCTTGAAG 16100
  M  E  Y  E  Y  D  I  L  R  N  W  I  V  G  D  S  C  L  R  L  G  K  I  G  F  R  G  A  A  S  A  Q  L   349

GCCGTTGATTAGTCTGGATTCTGTCACCGGAACAATGTACTTCAGTTCCGTGTCTCCGTTTGGAATGATGATTGGATGTAGAAACTTTGTAACACTTTGA 16200
 G  N  I  L  R  S  E  T  V  P  V  I  Y  K  L  E  T  D  G  N  P  I  I  I  P  H  L  F  K  T  V  S  Q   316

                                                                    < start of U7 exon 2
ACAAGAGGATTACTGATTTCAGATGAGTGCACTGCAAACAGAATACTTGGCATGTTCTTCAATAATCTGAAAAAGAACATCCATGGAACCATCAGTAAGA 16300
V  L  P  N  S  I  E  S  S  H  V  A  F  L  I  S  P  M  N  K  L  L                                   294

U7
exon 1                                                < end of U7 exon 1
GTATTGTGTAGAGAATACCGAAAAGAAGAACTTTGATCTTTACATACCTATCACACCTTGGATACATATGCCGTGGTTTCTTCCGAAGAATCGGAGACTC 16400
                                                      R  D  C  R  P  Y  M  H  R  P  K  K  R  L  I  P  S  E   276

TCCATTTAACATTCTGTTCAAATTTTCCGCCTCACTTTCACTCACATAAATCAAACTGTCATCATGCCAGTTCAGGGCATATCTTTCTTTGTATAAGAGT 16500
 G  N  L  M  R  N  L  N  E  A  E  S  E  S  V  Y  I  L  S  D  D  H  W  N  L  A  Y  R  E  K  Y  L  L   243

TCTTCAAACCCTCGCCTGATAAATTCTCTAATAGACCTTGCAACTTTGTTGATCTTGTTATCACAATACACATAAATGTGAAAATTGGGACCTAGTAGGA 16600
E  E  F  G  R  R  I  F  E  R  I  S  R  A  V  K  N  I  K  N  D  C  Y  V  Y  I  H  F  N  P  G  L  L  I   209

TGACAGCTCTACAATTTGGCTCTACCTGACATCTCGCGCAAGTTCCTATGACAACATACTTTCTTCTTAGTATTTTTTCGAGAATACCAGAACCTGTGAT 16700
  V  A  R  C  N  P  E  V  Q  C  R  A  C  T  G  I  V  V  Y  K  R  R  L  I  K  E  L  I  G  S  G  T  I   176

GTAATGGATGTCATTCGAAGAACAGAAAGACAAATCAAAATTTTGAAAAGCATCGATACGAATTTTTTGACCACATTGGCGATTGACAAAATTCTTTAAG 16800
 Y  H  I  D  N  S  S  C  F  S  L  D  F  N  Q  F  A  D  I  R  I  K  Q  G  C  Q  R  N  V  F  N  K  L   143

GCTGGAATATTTTTTTGTAAGAGTACGAACTCCTCACCGCTGCATTTGGTTTCTCCGTACCACAGTCTGTCTAGCTCAATATCTAATTCAATCCTTGCAA 16900
A  P  I  N  K  Q  L  L  V  F  E  E  G  S  C  K  T  E  G  Y  W  L  R  D  L  E  I  D  L  E  I  R  A  F   109

AAGGTGGGAAATTTCTCAGCCCGATACTGCAAAAATGTGCAGGAGATTCCGCTATGATGTACAATTCATCTGTGACCGTATCTAATGCAAGCATCTTATT 17000
  P  P  F  N  R  L  G  I  S  C  F  H  A  P  S  E  A  I  I  Y  L  E  D  T  V  T  D  L  A  L  M  K  N   76

TGATGAGCAGTACATCACAAATGGATCAAAGTCCGGGTCCGGTTCTCGGTAATTTGGATGTGCATAACCAACAGCGCAAAGAAATTCGTCTGCTCCTACC 17100
 S  S  C  Y  M  V  F  P  D  F  D  P  D  P  E  R  Y  N  P  H  A  Y  G  V  A  C  L  F  E  D  A  G  V   43

AAATTACTCCAATCATAAGTGATGACACTACGTCTACATTTGAAATGATAAAATTCGCCAAAAATAAGCACATAGTTTGGCGGCCAAGCAAGACAGCACT 17200
L  N  S  W  D  Y  T  I  V  S  R  R  C  K  F  H  Y  F  E  G  F  I  L  V  Y  N  P  P  W  A  L  C  C  K   9

TTATACCATTGCGAACTATCTCCATGCCTTCTCCATATGAAGACAGCAGCAACTCAACGGCAAAATCAAAAGACGGTAAATCCGGAGTCGGTTCAATGCA 17300
  I  G  N  R  V  I  E  M                                                                              1

AACATCATTGCACACAGGAACAATTTCTGCAAAATTAACTTGTGATTAGTACTTCCGTCACTTTCAGAAAAATTAAAAACCATAAAAAGTACTGTTACTT 17400
                        < start of U7 exon 1
U8                                                                                                - K   361
```

```
ACTTAGATACCAGGAGTACTGTTGAACCGGTGCACTTCCAAGAGTTGATAATCTTCCAAATCTGGCAGTGACTTTTCCAATTTTATCATAACCAAAATCT 17500
 S  L  Y  W  S  Y  Q  Q  V  P  A  S  G  L  T  S  L  R  G  F  R  A  T  V  K  G  I  K  D  Y  G  F  D    329

CCACAATATCTCGCGAAGCCGATCCTTGCAAAAGTCAGTAAACAATCAGCAATCCGCGTCAACACTTTATCAATCCAATCATAATAATAAATTGCTCCTG 17600
 G  C  Y  R  A  F  G  I  R  A  F  T  L  L  C  D  A  I  R  T  L  V  K  D  I  W  D  Y  Y  Y  I  A  G  T    295

TTTCTCCAATGCATAGAATTGGTCTCAAACCATAGTCTTCGGCTGAAATATTTAAAAAACCTAAAACAACTACTTTTTCAAACAGTGCCTCTTGTTCTTC 17700
   E  G  I  C  L  I  P  R  L  G  Y  D  E  A  S  I  N  L  F  G  L  V  V  V  K  E  F  L  A  E  Q  E  E    262

AAATCTTTTTTGGTCTGAGGTGCCACAGTGTATGTTTTTCCTGTCTTGCAGGATTAACGCTGTAAAGATTGGAGTTTCCAAACGAAGTTTTAAACCAGTG 17800
 F  R  K  Q  D  S  T  G  C  H  I  N  K  R  D  Q  L  I  L  A  T  F  I  P  T  E  L  R  L  K  L  G  T    229

TTTATAGAAACGAAGTGTTCAATCTCTGCTAACGTTTTACACCTTGCCAATTCATCTAATTCGGAGAGCATTGAAATATAATTTCTCCTCAAAGCTTTTA 17900
 N  I  S  V  F  H  E  I  E  A  L  T  K  C  R  A  L  E  D  L  E  S  L  M  S  I  Y  N  R  R  L  A  K  L    195

GTGCCAAAGCTGTTTTCCTGTTAAATCTTAGACTAGAATTCAAATTAAGTGGTTCCTGTAATAATCCTAAGTCTTTAAGAGTCTCCAATTTTTCAAACTT 18000
   A  L  A  T  K  R  N  F  R  L  S  S  N  L  N  L  P  E  Q  L  L  G  L  D  K  L  T  E  L  K  E  F  K    162

TGGACCATCATAATAAAAGCGAGTATAAACAGGTTCTGAGCAAGTGATTACATTTGTACCTAGAATACCAAATCTAACAAATTTGTCAAAGTCTTCAGCT 18100
 P  G  D  Y  Y  F  R  T  Y  V  P  E  S  C  T  I  V  N  T  G  L  I  G  F  R  V  F  K  D  F  D  E  A    129

ATGTAATACAAAACATCATCCTCCAAATCGTGACAAAAAATTCGACTCGATTTTCCCACAAAAATTGGCACGTCTTTCCTTCCAATTTGAATCGCTCCAA 18200
 I  Y  Y  L  V  D  D  E  L  D  H  C  F  I  R  S  S  K  G  V  F  I  P  V  D  K  R  G  I  Q  I  A  G  I    95

TCAAGCACAAATTTTCCTTAAATCCTTGGTAATGGTCATTCAGTTCTGCCAATTGTTGTTCTGAGTAACCAGCAACGTCAACTGGCTGTGCAAGCCTTAA 18300
   L  C  L  N  E  K  F  G  Q  Y  H  D  N  L  E  A  L  Q  Q  E  S  Y  G  A  V  D  V  P  Q  A  L  R  L    62

CCACAAATTCTTTGGCCAAATCAAAGACAAGCATTCATTTCGATGTCGTTCAGTAAGAAGGCGAATTCCAGGGAAATCAGTAAAACGACATAAATTCTTC 18400
 W  L  N  K  P  W  I  L  S  L  C  E  N  R  H  R  E  T  L  L  R  I  G  P  F  D  T  F  R  C  L  N  K    29

AATTCCCTTAACAATCCATCATTTAGTTGTCCAACACAAGGCAAAGTTTTATATTGTTCTGCCATCGTCTCAAAATTTTTCATTGAAGTGAATGTGGTAA 18500
 L  E  R  L  L  G  D  N  L  Q  G  V  C  P  L  T  K  Y  Q  E  A  M  T  E  F  N  K  M                    1

GAAGTTTGTATTCTCATCCATTTCTCAAAAACTCAACCAGATAAATATAAGAGAAAAAAAATCACATGACAGGAAATAAATTTTAAAGCTTTGCAGATTTT 18600

TTATTTCCCATTACAGGAAACATTAGAGTGATTAAAACCATAAGCCTGGTGTTATCGGATTAAAAACCTCTTGTCGAAAGTAGGAATCGGATTTTTACCT 18700

TTTTTAAAGATCGATTTTTAGAACATTCACGGTAACAATACAAAACTAACAGTGTCATTAATATTACAAACTCAGAATCTGCCACTGAAAGGAAAATCCC 18800

U10                                M  A  I  A  E  S  G  A  F  E  I  N  V  N  L  G  K  S  I  A  S  I  Q  K    24
TTTTCGCCCTAAGTAACTACACTTAAAAATGGCTATAGCAGAAAGCGGAGCGTTCGAAATAAATGTCAATCTTGGAAAATCCATCGCTTCCATACAAAAA 18900

 N  P  I  I  Y  M  R  R  H  L  S  F  Y  V  E  L  L  K  F  I  I  H  Q  Y  E  Q  C  F  L  P  P  K  G    57
AATCCTATAATCTACATGCGTAGACATCTCTCTTTTTATGTGGAGTTATTAAAATTTATCATTCATCAATATGAGCAATGTTTTTTACCACCGAAAGGAA 19000

 T  I  L  Y  H  N  G  L  I  E  L  N  T  L  I  I  D  L  N  Q  Q  I  T  S  K  Q  Q  I  Y  S  W  T  S  I    91
CGATACTTTACCACAATGGCTTAATCGAACTAAATACTTTAATTATCGATCTGAATCAACAAATTACATCGAAACAGCAGATCTATAGTTGGACGAGCAT 19100

   T  L  P  K  I  F  S  T  K  E  L  Y  F  I  V  A  S  P  E  S  E  N  I  T  L  N  P  A  V  T  K  G  G    124
TACTTTGCCAAAAATATTTTCGACTAAAGAATTATATTTTATTGTCGCGTCACCAGAATCGGAAAACATAACTCTCAATCCAGCTGTTACTAAAGGAGGA 19200

 W  L  S  G  S  F  S  F  P  L  S  L  S  C  A  Y  A  L  T  G  V  S  S  T  I  Y  M  L  P  F  I  P  Y    157
TGGTTATCTGGGAGTTTTAGCTTTCCATTGAGTTTAAGTTGTGCATACGCCCTCACTGGCGTATCTTCAACAATCTATATGTTACCATTTATTCCATATA 19300

 K  F  P  M  T  Y  V  D  F  S  T  L  R  T  Y  E  V  T  S  E  Y  G  S  I  Q  I  I  K  Q  R  N  F  L  F    191
AATTCCCAATGACTTACGTAGACTTTTCGACACTTCGAACATACGAGGTAACAAGTGAATATGGATCAATCCAAATTATAAAACAGCGGAATTTTTTATT 19400

   L  G  I  I  R  D  L  S  W  K  S  Q  R  D  N  K  N  F  I  L  K  A  M  F  V  G  N  W  L  G  I  Q  I    224
TTTGGGAATTATACGGGATCTGTCATGGAAAAGTCAGAGGGATAACAAGAATTTTATCCTGAAAGCAATGTTCGTTGGAAACTGGTTAGGGATACAAATT 19500

 P  E  A  F  A  L  R  L  F  N  N  T  R  F  S  I  Q  D  F  E  F  S  I  N  I  Q  N  I  N  L  T  R  D    257
CCAGAAGCTTTTGCATTGAGACTTTTTAATAACACACGGTTTTCTATTCAGGATTTTGAATTCTCCATTAACATACAAAATATAAACCTTACTAGGGACA 19600

 N  K  I  L  G  S  L  S  T  V  S  C  D  Q  M  P  P  N  L  S  P  E  N  L  P  N  Y  L  V  I  Q  F  E  L    291
ATAAAATTTTGGGTTCTCTTTCCACGGTTTCTTGTGATCAGATGCCACCAAACTTGTCTCCAGAGAATCTCCCAAACTATCTAGTTATTCAGTTCGAATT 19700

   V  S  T  L  A  N  P  D  H  L  L  F  S  C  N  P  K  L  F  F  T  G  D  I  L  N  S  A  I  N  L  Q  H    324
GGTTTCAACTCTCGCAAATCCTGATCATCTACTGTTTTCTTGCAATCCCAAATTGTTTTTCACAGGAGATATTCTGAACAGCGCTATAAATTTACAACAT 19800

 S  P  N  H  Y  E  L  T  V  Y  A  P  H  N  L  H  F  Y  P  S  C  F  H  I  V  T  L  P  I  Q  F  S  S    357
AGTCCTAATCATTATGAGCTTACAGTGTACGCACCACATAATTTACATTTCTATCCCAGCTGTTTTCATATAGTAACATTACCAATTCAGTTTTCATCCA 19900

 R  N  D  R  Q  M  L  V  S  S  Y  P  N  E  G  Y  F  E  V  Q  M  C  P  W  V  Q  N  S  P  L  Q  I  V  I    391
GAAATGATAGACAGATGTTGGTGTCAAGCTATCCTAATGAAGGCTACTTTGAAGTACAAATGTGCCCATGGGTACAGAATTCTCCTCTTCAAATTGTTAT 20000

   K  S  F  S  K  N  L  V  L  P  Q  G  T  P  I  A  I  L  L  Y  M  E  K  M  T  T  G  K  N  S  L  R  D    424
TAAATCTTTTTCAAAAAATCTAGTGCTACCGCAAGGCACCCCTATTGCAATTCTCTTATATATGGAAAAAATGACAACAGGAAAAAATTCTTTACGGGAT 20100

 Q  E  L  K  I  N  K  D  I  T  R  I  G  N  V  N  L  P  K  E  N  F  L  H  Y  N  S  -                    451
CAAGAATTAAAAATTAATAAAGACATTACCCGAATAGGAAACGTAAACCTTCCAAAAGAAAATTTCTTACATTATAATAGCTAATCTTTATTCAACTGTG 20200
                                                                                  ------
                                                                                        ------
TTTTAGACGACGTCTTCTTCTGAGTGTGTTAAATGCGATACTAGTTTATTAACTATATCATCATCTTTTTGAACAGAATTTTCAAGATCGATTAAATCTT 20300
U11    -  V  V  D  E  E  S  H  T  L  H  S  V  L  K  N  V  I  D  D  D  K  Q  V  S  N  E  L  D  I  L  D  K   724

TATCGCCATCTTCTTGAGCCTTGCTAGCACCGAAAATGGTATTTTGCGTCTCTTTGCCTGTTAAAAGTGAACCTGAATCAAAATCCTGGATTTCAGCATT 20400
  D  G  D  E  Q  A  K  S  A  G  F  I  T  N  Q  T  E  K  G  T  L  L  S  G  S  D  F  D  Q  I  E  A  N    691
```

```
                                                    ------
TGGTAAAAAAAAATGGAGACCGCACAGGCTGCATCTTTTCAGTGAAGCTAAAATTTTGAAATAAAGGATCTCTCTTTCTGGAGTTCAGTAAATTCTTCAGG 20500
 P  L  F  F  P  S  R  V  P  Q  M  K  E  T  F  S  F  N  Q  F  L  P  D  R  K  R  S  N  L  L  N  K  L    658

TTGTTGTCATTGAAAGTCGCAATGTTAGAAACATCTGTTTTTTTCGGGGGTTCGAGTACGTCTTTTAAGCTTCGCTCTTCTAAACTAATTGTATCCGTAA 20600
N  N  D  N  F  T  A  I  N  S  V  D  T  K  K  P  P  E  L  V  D  K  L  S  R  E  E  L  S  I  T  D  T  F    624

AATCCGTTAATAATTTCTTAACTTTTTCACCTAGTCCTGATTCAAAAAGTTCTTTTTGAAGAAGATCGTTCCTTCTACTAAGATCCAGTAATGAAGAACT 20700
  D  T  L  L  K  K  V  K  E  G  L  G  S  E  F  L  E  K  Q  L  L  D  N  R  R  S  L  D  L  L  S  S  S    591

CGGAGTCCACTTCACATCTTTGACTAATTGATTTTTTTTAGCTGAAGAAGCACTCTCTTTTTCCTTCATTGAAGCGTTTACGTCATTAGTACCTATCATA 20800
 P  T  W  K  V  D  K  V  L  Q  N  K  K  A  S  S  A  S  E  K  E  K  M  S  A  N  V  D  N  T  G  I  M    558

TTTAAAGTCTTGGTGGAAGAATTTGCTTTATTTCCATTTTTTTGCTTAATGCTTTTTTGATCTTGTGGAGGCAATGACATTTCATTATGCTCGTCATCAA 20900
N  L  T  K  T  S  S  N  A  K  N  G  N  K  Q  K  I  S  K  Q  D  Q  P  P  L  S  M  E  N  H  E  D  D  F    524

AACTTTGTAAGTTCTCGTGGACAATTTCTTTTCTGCGTTTTTCTGAAGGAAGTTTAAAAATGTCTTCATTATTTTGTTTAATTGTTTTGATTCGTTCTCG 21000
  S  Q  L  N  E  H  V  I  E  K  R  R  K  E  S  P  L  K  F  I  D  E  N  N  Q  K  I  T  K  I  R  E  R    491

TAGCTCCAAAATTTTCTGCAAGTCATTGGTCGAGTTTTTAGACAGTTGATCGCGTGATTGCACTCCGTTTTGTCTGTTTACTTCGTTTTCTGGCAGAAAA 21100
 L  E  L  I  K  Q  L  D  N  T  S  N  K  S  L  Q  D  R  S  Q  V  G  N  Q  R  N  V  E  N  E  P  L  F    458

AAGGTAGAGTGATAATCAATTTTATTTTTGGGTAGCTGAACTACGTTAACTTTTGGTGAGCTAACAATGTCTATTACTTTGTTTGAGTTGTGTAGATCTA 21200
F  T  S  H  Y  D  I  K  N  K  P  L  Q  V  V  N  V  K  P  S  S  V  I  D  I  V  K  N  S  N  H  L  D  L    424

GTTTGCTCAAGGTATCTATAACTTGGTTAATACCGTCTGGTGCCACAGCAATTTCATTGAATTGGTTTACTTTGTCCTCTGGTCTGTGCATAGTTTGATC 21300
  K  S  L  T  D  I  V  Q  N  I  G  D  P  A  V  A  I  E  N  F  Q  N  V  K  D  E  P  R  H  M  T  Q  D    391

AGTTAAATCAAGTATTTCAGTCTGTAGTATTCTGTGCATTTTTTTAGAATTGTTTATAGACGTGGAATTTCTTGGCTCGTCCAATATCATGGAATCCTTT 21400
 T  L  D  L  I  E  T  Q  L  I  R  H  M  K  K  S  N  N  I  S  T  S  N  R  P  E  D  L  I  M  S  D  K    358

TTTACCGAGGATAACAACCCATCTTCAGAGCCTACAAATAAATCTTTGGATAGCGTGTTGTTACTCCGTTGTTCAGTTTTTATAAATGGTTCTGGATTAA 21500
K  V  S  S  L  L  G  D  E  S  G  V  F  L  D  K  S  L  T  N  N  S  R  Q  E  T  K  I  F  P  E  P  N  F    324

ATGATGAACCATACGAAGTTTTTTTAGAACTACTCTCAAATTGATTGCTTAAAGACGATTTCGACAAATCTGTTAAAACTTTGAATGTCTGAATCGGTTC 21600
  S  S  G  Y  S  T  K  K  S  S  S  E  F  Q  N  S  L  S  S  K  S  L  D  T  L  V  K  F  T  Q  I  P  E    291

AGGAGTAAATTTTACACGATCATCCCATGTACTCTTTACGGGAAGAGCATGTAACTCAAGACTTTTTGTTATAACATCAAAGTCTTCGGTTAAAATTTTA 21700
 P  T  F  K  V  R  D  D  W  T  S  K  V  P  L  A  H  L  E  L  S  K  T  I  V  D  F  D  E  T  L  I  K    258

AAAAAAGAAGTTACTCTATGACTAGTTAATGGCAAACTTAAAATTTGATAAGAATATATCGAAACAAAATTTTTGTTGTTTTCAATCGCCTTTAAAATTG 21800
F  F  S  T  V  R  H  S  T  L  P  L  S  L  I  Q  Y  S  Y  I  S  V  F  N  K  N  N  E  I  A  K  L  I  A    224

CATTTTTCTTTTCATTAATGGTATTTAATGCGAGTTTCTCCATGTTCATCCAGGATCCATGTAATGTAATTAATAAATTCCCAAGATAAAGCAATCGATT 21900
  N  K  K  E  N  I  T  N  L  A  L  K  E  M  N  M  W  S  G  H  L  T  I  L  L  N  G  L  Y  L  L  R  N    191

GATGTTAGTAACATAGTAGCCATTTTCATTCCACGGATCAACTATTTGGTAAATTGTGAAAGAAAGCATATTGTTTTTTGCCTTTTGTAAATTCTCGCCA 22000
 I  N  T  V  Y  Y  G  N  E  N  W  P  D  V  I  Q  Y  I  T  F  S  L  M  N  N  K  A  K  Q  L  N  E  G    158

ATTTCCGCTCTTTCTTCTGTGGTAAGATTGACGTAATCGGCTGCGTGTGTCACAGTGCCCCAACGAAGAAAACTGCCCAGTTTAAGCAAATCTTTTGCTA 22100
I  E  A  R  E  E  T  T  L  N  V  Y  D  A  A  H  T  V  T  G  W  R  L  F  S  G  L  K  L  L  D  K  A  V    124

CGTTTTTAAATTCTGTTTCAAAGCCAACTTTGTTAAAAGATGTTAAACTAGAATAATCAGTAATCATCTGTCTTTGCTCTAAGTAATCTTTCAGAATTTT 22200
  N  K  F  E  T  E  F  G  V  K  N  F  S  T  L  S  S  Y  D  T  I  M  Q  R  Q  E  L  Y  D  K  L  I  K     91

TTGCACCTTTAGTATTTCTTGCCAAACTTCCTCATAGTCTGGTTTCTTTTTACACAAGGTTTGATGAAAATAAATCCAAAGAATCAAATTGTTGTAAAGT 22300
 Q  V  K  L  I  E  Q  W  V  E  E  Y  D  P  K  K  K  C  L  T  Q  H  F  Y  I  W  L  I  L  N  N  Y  L     58

GTTTTTACATTGTTAATGGAATTTCCAGTTTTAACGATACATTGACTCAAAATCCATGGATTTTCGCGAATATCAACAACTGGTAAAGACGATATGTTTT 22400
T  K  V  N  N  I  S  N  G  T  K  V  I  C  Q  S  L  I  W  P  N  E  R  I  D  V  V  P  L  S  S  I  N  E    24

CAAAAAATCGAGATAAAAAACACTTTGCTTCATCCGAAATCCAAGCAAAGGGTAAATGTGACATTTTCATGACTCGTTTTTTCTGTTGAGCATTATTGAA 22500
  F  F  R  S  L  F  C  K  A  E  D  S  I  W  A  F  P  L  H  S  M  K  M                                1

U12
exon 1                                                                                  M  D  T  L  I  D  F  Q  K  I  L     11
CACTTCTTAAAAAATTAGTTTAATAGTCTCTATTTTCTCTTTGATTAGATACATGCGCTGCCGACATGGACACTCTAATTGATTTCCAAAAAATCCTGGT 22600
                                                                                        end of U12 exon 1 >

U12
exon 2                                                                                  H  D  E  E     15
ATGATAATCTAATGTTATGATTTGCTACAGCTTTGCTAAAAATGTTACCTTTGCATTTTTAATCATCTTGCAAAACTTTTTTTCACAGCATGATGAAGAG 22700
                                                                                     start of U12 exon 2 >

 Y  K  Y  N  Y  T  C  I  T  P  T  V  R  K  A  Q  R  L  E  S  V  I  N  G  I  M  L  T  L  I  L  P  V     48
TACAAGTACAATTATACGTGTATTACGCCAACAGTACGGAAAGCCCAAAGACTTGAAAGCGTAATTAACGGAATTATGCTAACGCTGATACTTCCTGTTA 22800

S  T  V  V  I  C  T  L  L  I  Y  Y  K  W  T  K  Q  T  I  T  S  P  Y  L  I  T  L  F  I  S  D  S  L  H     82
GTACTGTTGTCATATGCACTCTGCTAATCTACTACAAATGGACAAAACAGACAATTACTTCTCCATATCTTATCACACTCTTTATTAGTGATTCTTTACA 22900

  S  L  T  V  L  L  L  T  L  N  R  E  A  L  T  N  L  N  Q  A  L  C  Q  C  V  L  F  V  Y  S  A  S  C    115
TTCATTGACTGTGTTACTTCTCACATTGAACCGAGAAGCTCTCACAAACCTTAATCAGGCTTTGTGTCAATGTGTGCTTTTTGTATACAGTGCGTCCTGC 23000

 T  Y  S  L  C  M  L  A  V  I  S  T  I  R  Y  R  T  L  Q  R  R  T  L  N  D  K  N  N  N  H  I  K  R    148
ACATACAGTCTGTGCATGCTAGCAGTAATATCCACAATACGCTATCGAACCCTGCAAAGAAGGACATTAAACGACAAAAACAATAATCATATTAAAAGGA 23100
```

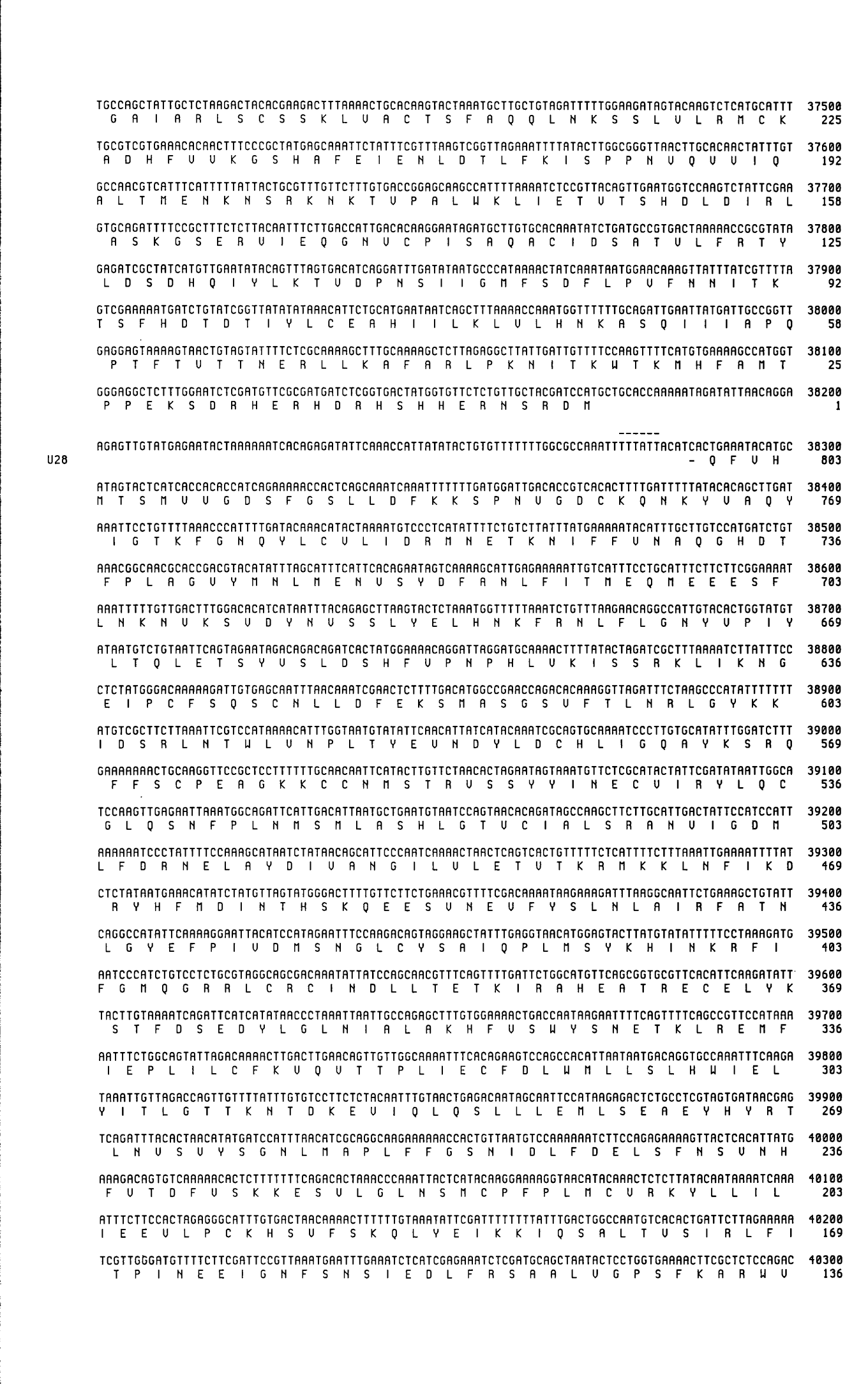
```
     N  V  G  I  L  F  L  S  S  A  M  C  A  I  P  A  V  L  Y  V  Q  V  E  K  K  K  G  N  Y  G  K  C  N  I    182
     ACGTTGGAATTTTATTTCTGTCTTCTGCCATGTGTGCCATTCCAGCAGTATTATATGTTCAAGTGGAAAAGAAAAAAGGCAATTATGGAAAATGTAATAT  23200

       H  I  S  T  Q  K  A  Y  D  L  F  I  G  I  K  I  V  Y  C  F  L  W  G  I  F  P  T  V  I  F  S  Y  F    215
     ACACATCTCAACGCAAAAAGCATATGACTTGTTTATAGGAATTAAAATTGTCTATTGTTTTCTCTGGGGAATTTTTCCAACTGTCATTTTCAGCTATTTT  23300

      Y  V  I  F  G  K  T  L  R  A  L  T  Q  S  K  H  N  K  T  L  S  F  I  S  L  L  I  L  S  F  L  C  I    248
     TATGTGATTTTTGGTAAGACCTTGCGTGCCTTGACCCAAAGTAAACATAACAAAACTCTGTCATTCATTAGCTTACTGATACTATCCTTTTTATGTATTC  23400

     Q  I  P  N  L  L  V  M  S  V  E  I  F  F  L  Y  I  A  N  T  S  C  L  G  T  I  Q  R  E  I  V  Q  I  I    282
     AAATACCAAATCTCCTAGTAATGTCTGTGGAAATTTTTTTTTTGTATATAGCAAATACTTCCTGCTTAGGCACCATACAAAGAGAAATTGTGCAAATAAT  23500

       S  R  L  M  P  E  I  H  C  L  S  N  P  L  V  Y  A  F  T  R  T  D  F  R  L  R  F  Y  D  F  I  K  C    315
     ATCTAGATTAATGCCAGAAATACACTGCTTGTCTAATCCGCTAGTATATGCATTCACTAGAACAGATTTCCGATTACGATTTTACGATTTCATTAAATGT  23600

      N  L  C  N  S  S  L  K  R  K  R  N  P  L  T  I  K  N  -                                              333
     AATTTGTGTAATTCATCTTTAAAGAGAAAGAGAAATCCTCTGACAATAAAAAATTGAACAGTGAAGCTTGCTTAAAACTTTGAAAATTTTTTGTGATTTT  23700

U13  ------                                  M  T  L  N  Q  Q  P  S  N  C  R  L  I  T  A  N  D  P  V  L    20
     AAATAAAATTTAAAATCAATTTATCCTAAGCTTTTCAGGCTATGACTTTGAATCAACAGCCATCTAATTGCAGACTAATTACAGCTAACGATCCAGTACT  23800

       A  S  N  F  T  M  Q  P  T  F  K  I  A  D  K  K  V  V  L  R  D  H  N  Y  I  A  V  K  D  F  V  L  S    53
     TGCATCGAATTTTACAATGCAGCCTACTTTTAAAATAGCCGATAAGAAAGTTGTACTCAGAGATCATAATTACATTGCAGTCAAAGATTTTGTTCTATCA  23900

      R  S  F  M  H  C  I  R  C  Q  E  K  I  E  K  K  T  S  D  S  I  R  A  Y  S  I  K  P  D  F  M  F  Y    86
     AGGTCTTTCATGCATTGCATTCGATGTCAGGAAAAAATTGAAAAGAAGACATCCGATAGCATTAGAGCATATTCCATTAAGCCAGATTTCATGTTTTATG  24000

     D  S  D  D  E  V  R  S  S  P  F  L  -                                                                98
     ACTCAGACGACGAAGTTCGTTCGTCACCATTTTTATAACAATATTTAAAGTAGCTGAATTGTACGTTCTGAAACAGTCTGAACGGATCCGCTGAACGCAG  24100

U14       M  E  Q  Q  K  G  F  S  I  P  F  F  V  T  D  E  N  C  N  F  V  P  E  I  L  P  R  I  H  T  K  F    32
     ACACAATGGAACAGCAAAAAGGATTTTCGATTCCATTTTTTGTCACTGACGAAAACTGCAACTTTGTGCCTGAAATATTACCTCGTATACATACTAAATT  24200

       L  K  D  V  L  I  A  D  S  Y  N  S  V  S  W  A  N  S  F  I  P  M  P  I  Q  T  L  E  Q  I  M  V  L    65
     TCTTAAAGATGTCTTAATTGCCGATTCCTACAACTCTGTTAGTTGGGCAAACAGTTTTATTCCCATGCCTATTCAAACGCTTGAACAAATTATGGTTCTT  24300

      I  T  K  F  K  F  S  R  S  R  D  F  L  F  P  V  I  R  L  A  V  H  I  N  R  F  H  T  G  K  K  Q  L    98
     ATAACAAAGTTCAAGTTTTCTCGCTCGCGTGATTTTTTATTCCCAGTAATTCGATTAGCTGTTCATATCAATAGGTTCCACACAGGGAAGAAACAGCTGA  24400

     K  T  M  I  E  I  M  K  S  L  F  N  T  E  E  A  M  R  R  F  D  E  A  L  M  I  L  F  S  N  E  Q  T  N   132
     AAACCATGATAGAAATTATGAAAAGCTTGTTTAACACCGAAGAGGCTATGCGACGATTCGATGAAGCATTGATGATTCTATTTTCTAATGAGCAAACCAA  24500

       T  Y  M  T  N  I  A  L  S  M  H  E  N  G  L  P  D  S  K  F  M  N  A  L  K  M  I  Y  R  A  G  N  S   165
     TACTTACATGACAAACATAGCTTTATCGATGCATGAGAATGGTCTTCCAGATTCAAAATTTATGAATGCTCTAAAAATGATTTACAGAGCTGGAAATTCT  24600

      F  D  N  Q  P  D  N  D  I  E  S  Y  N  E  K  L  K  I  Y  N  Y  L  I  K  I  P  K  Y  T  L  K  A  G    198
     TTTGATAATCAACCAGACAATGATATAGAAAGCTACAACGAGAAATTAAAAATCTACAACTACCTAATTAAAATACCTAAGTACACACTAAAAGCTGGAG  24700

     V  D  L  Y  N  E  N  I  K  D  L  S  I  G  I  Q  R  Q  P  T  L  L  F  T  S  R  N  D  F  S  L  K  A  I   232
     TTGATTTATATAATGAAAATATAAAAGATCTTTCGATTGGAATCCAAAGACAACCTACTTTATTATTCACATCTCGTAATGATTTTTCATTGAAAGCTAT  24800

       Y  N  D  V  L  F  L  V  S  A  W  N  M  I  I  N  Y  K  K  E  Q  R  R  L  F  S  W  I  T  F  E  I  N   265
     TTATAATGACGTGTTATTTCTAGTTTCTGCATGGAATATGATCATCAACTATAAAAAAGAACAAAGAAGACTCTTTAGTTGGATAACATTTGAAATAAAC  24900

      S  L  M  E  N  V  V  L  A  A  F  Q  L  P  D  L  K  E  M  T  L  D  L  S  A  L  I  A  N  M  M  N  L  L   298
     TCTTTAATGGAGAATGTTGTGCTTGCAGCCTTTCAGTTACCAGATTTAAAAGAAATGACACTTGATTTAAGCGCTTTGATTGCCAACATGAATCTTTTAA  25000

     K  P  N  D  D  Y  S  P  H  F  K  L  I  I  N  K  F  F  E  I  G  I  F  V  T  K  S  Y  I  C  I  L  P  S   332
     AACCAAATGATGATTACAGCCCCCATTTTAAACTAATTATCAACAAATTTTTTGAAATCGGAATTTTTGTCACGAAATCATATATTTGTATTTTGCCTTC  25100

       F  V  K  S  Q  L  I  S  F  E  N  V  L  S  S  N  R  H  A  E  D  V  T  F  I  L  T  S  S  K  E  S  D   365
     TTTTGTCAAAAGCCAACTCATTTCGTTTGAAAACGTTTTAAGTTCAAACAGACATGCGGAAGATGTGACTTTCATACTAACATCTTCCAAGGAATCTGAT  25200

      D  E  Y  D  E  D  K  P  P  R  Q  V  D  P  D  R  V  D  N  I  L  M  E  S  D  F  F  N  V  K  P  E  N   398
     GATGAATATGATGAAGATAAACCTCCACGACAAGTAGATCCAGACAGAGTGGACAACATTTTAATGGAATCAGATTTTTTTAATGTGAAACCGGAAAACG  25300

     A  F  S  E  I  A  L  M  P  I  S  H  D  K  I  I  D  V  N  N  S  N  I  Q  V  L  E  T  E  L  A  H  T  N   432
     CCTTTTCAGAAATCGCATTAATGCCAATTTCACATGACAAAATTATAGATGTCAACAACTCAAACATACAAGTTCTTGAAACTGAATTGGCACATACAAA  25400

       L  F  V  Y  S  A  I  A  Q  K  Y  D  L  P  L  K  E  Y  V  E  R  L  N  V  Y  N  P  D  L  S  S  G  N   465
     TTTATTTGTGTATAGCGCAATAGCTCAAAAATATGATTTGCCTTTAAAAGAATACGTAGAGCGCTTAAACGTCTACAATCCTGATCTATCATCTGGAAAC  25500

      S  T  P  A  R  N  S  N  S  I  R  T  T  P  V  L  N  I  S  R  P  G  S  T  T  P  S  G  N  S  A  R  Y   498
     TCTACACCCGCGCGGAATTCTAATTCTATTCGTACAACTCCAGTTCTAAATATATCCAGACCAGGAAGCACTACACCCTCTGGGAACTCTGCAAGATATG  25600

     G  N  N  T  P  R  S  I  T  P  V  L  E  I  S  R  S  R  S  A  T  P  S  G  N  S  E  I  Y  E  N  R  T  S   532
     GGAATAACACACCTAGAAGTATAACTCCGGTTTTAGAGATTTCAAGATCTAGAAGTGCTACCCCTTCTGGAAATTCTGAAATCTATGAGAATAGAACATC  25700

       P  T  F  R  V  S  R  S  A  T  P  I  E  R  S  S  R  S  A  S  I  I  S  G  E  S  V  P  G  F  F  N  D   565
     CCCAACGTTCCGTGTTTCTAGGAGTGCAACTCCAATAGAAAGAAGTTCGAGATCTGCTAGTATAATTTCTGGAGAATCTGTTCCTGGTTTTTTTAATGAC  25800

      Q  E  R  L  S  T  N  S  P  I  S  I  N  G  N  T  P  R  Q  Q  S  H  G  D  N  E  I  Q  T  I  D  S  T   598
     CAAGAACGTTTGAGCACCAATTCGCCTATTTCCATAAATGGTAATACTCCAAGACAGCAAAGCCATGGGGACAATGAAATACAAACTATAGATTCCACGG  25900

     D  E  D  S  M  N  A  P  Q  S  P  Q  S  I  Y  S  I  S  S  Y  V  S  T  D  D  Q  L  L  H  S  P  T  N  S   632
     ATGAAGACTCAATGAATGCCCCACAATCACCCCAATCCATCTATTCAATTTCTAGCTATGTTTCAACAGACGATCAACTTCTACATTCACCGACAAATAG  26000
```


U15 exon 3
U15 exon 2
U15 exon 1
U17 exon 2
U17 exon 1
< start U15 exon 3
< end U15 exon 2
< start U15
exon 2
< end of U15 exon 1
< start of U17 exon 2
< end of U17 exon 1



```
AAAAAGTAGAATTTTTTCCCTTTTAAGATGTCCGATGCCAAATATTCTTCTGTCAAATCATCAGGAACCAGTTTATTTTCTAGCTTCCATTGAATACTCA 31800
F  F  Y  F  K  K  G  K  L  I  D  S  A  L  Y  E  E  T  L  D  D  P  V  L  K  N  E  L  K  W  Q  I  S  L   211

ATCCCAGTGGACTATAAAGCTGCGCCGAACACGTTATGTTAGATCCACGTTTTGTGATGGACACCACAGGTGGCCGATATCCATCTTGACCGAAGGTCAC 31900
   G  L  P  S  Y  L  Q  A  S  C  T  I  N  S  G  R  K  T  I  S  V  V  P  P  R  Y  G  D  Q  G  F  T  V   178

ATGTGCAGAAAAATCTGCGAGTTGTTTAACACAAAACTGCTGCACATAAAATTTGAGAATAAGTAATAATTCAACATTAGCATACATGTTTTCGTCAAAA 32000
   H  A  S  F  D  A  L  Q  K  V  C  F  Q  Q  V  Y  F  K  L  I  L  L  L  E  V  N  A  Y  M  N  E  D  F     145

CCACTGCTAGAAAAATCCAACAGGTTGACATTCAGTGTCGTGTCATTATGAACTTTTAACAAATAATGACTTTTCAAAACGGTGTCAATTTCACAAATGA 32100
G  S  S  S  F  D  L  L  N  V  N  L  T  T  D  N  H  V  K  L  L  Y  H  S  K  L  V  T  D  I  E  C  I  F   111

ATTCAGCGTTAAAATTGTGGTGATAATGAACACTGCGATTCCTATGTGACCAGACAAAATAAGAGAAAATGCTGTTTAATGTAGATTCGTCCTTGTTGGT 32200
   E  A  N  F  N  H  H  Y  H  V  S  R  N  R  H  S  W  V  F  Y  S  F  I  S  N  L  T  S  E  D  K  N  T     78

AGAAAAGACGTAAGACAAAGCATTAAATTTCGTTGGAAAATCTGTGAGGAAAGACGAAATACTAGTATTGCCGACAGACAGATTCTTATAATGTGCCAGT 32300
   S  F  V  Y  S  L  A  N  F  K  T  P  F  D  T  L  F  S  S  I  S  T  N  G  V  S  L  N  K  Y  H  A  L     45

TCATAGTTATCCATCATCAATTCATATTTCTCAAGATGTTCTTTATTGATTATTGACATTGTCAACTTATGTGCCACATATTTGGTAAAATGTACAATGA 32400
E  Y  N  D  M  M  L  E  Y  K  E  L  H  E  K  N  I  I  S  M  T  L  K  H  A  V  Y  K  T  F  H  V  I  F   11

AAACTAAAAGACATGTTTTTTTCACAAACATTGCTGTTTGAGAGATTCTTTTTTAGGTAACGGTTCTTGTCCAAACATCAAGTTTGTTAGATTGATTATT 32500
   V  L  L  C  T  K  K  V  F  M                                                                         1
U21                        -  L  C  Q  Q  K  L  S  E  K  K  P  L  P  E  Q  G  F  M  L  N  T  L  N  I  I     406

TTCTCTGATACTGGTTCTTTTTGTATGTGCTGTAAAGGCCACGTCAGAACTTTCTGAACACCGTCTAAACAGGAAACTCCTAATAAAACTACCGCTGTAA 32600
K  E  S  V  P  E  K  Q  I  H  Q  L  P  W  T  L  V  K  Q  V  G  D  L  C  S  V  G  L  L  V  V  A  T  V   372

CAAATGCAAGGAATAATATAGCAATCAATGTGTAGCAAAGCATGTCTACGAAAACATTACTAGACGTTTTTTCACTATCTAATAGATCCAAATCCACATC 32700
   F  A  L  F  L  I  A  I  L  T  Y  C  L  M  D  V  F  V  N  S  S  T  K  E  S  D  L  L  D  L  D  V  D   339

AGGTAATGTTACCTGTATTTGTGTTCCTGTGTTCTTATTCGTACAGGTACAAACATATGTGGTCATTTCTACTCTTAGAAAAATTAGACCCCAAAGCACA 32800
   P  L  T  V  Q  I  Q  T  G  T  N  K  N  T  C  T  C  V  Y  T  T  M  E  V  R  L  F  I  L  G  W  L  V   306

TCACTACCATACAAAGGAGTAAAAGCACCCGTAGAGTTGAAACCAACTCTTTTCCACTCTTCTCCAAACGGACTTTTTATCTTTTGCAGAACTTCTAAGG 32900
D  S  G  Y  L  P  T  F  A  G  T  S  N  F  G  V  R  K  W  E  E  G  F  P  S  K  I  K  Q  L  V  E  L  S   272

AGCCGAAACCGCTAACACCCCATCCTGTATATTTACAAACTGCTACCTTTACACCATTCTGTCTAAATTTCGAAAATGAACAATCTATGTTATTAGTTCT 33000
   G  F  G  S  V  G  W  G  T  Y  K  C  V  A  V  K  V  G  N  Q  R  F  K  S  F  S  C  D  I  N  N  T  R   239

ATATCGTTGTTCATTTCTTTGTTCAATCATAGGATTAAAGCCTTGCTTAAATGTAGCTTCGAATGTTTTCGCGAGTGATAGGTACTTTCTTTGCCAAAAA 33100
   Y  R  Q  E  N  R  Q  E  I  M  P  N  F  G  Q  K  F  T  A  E  F  T  K  A  L  S  L  Y  K  R  Q  W  F   206

AAAAGTTGATTGAAGTCATTTTCACAAGAATTGATCAGAAAGTTTTTAGCGTGTGTCAAAGTGTGCCAAATTCGATAAGTGAAGGGATCATCTTCACCTA 33200
F  L  Q  N  F  D  N  E  C  S  N  I  L  F  N  K  A  H  T  L  T  H  W  I  R  Y  T  F  P  D  D  E  G  L   172

AATAGTTTTTAATCGCTACCCACGTGCTGTTTCTTTTATTTTGCAAGTAGAATGTGGTTTCAAGAATTCTAGGAGAAATTCCAGTTTCATTCATACGTCT 33300
   Y  N  K  I  A  V  W  T  S  N  R  K  N  Q  L  Y  F  T  T  E  L  I  R  P  S  I  G  T  E  N  M  R  R   139

CAGACGGATAAAACTTTAAAGGGTGAACAGATAATTTGAATATGAATGTAAGACTCGACATCAAGACAGAGTTTATCATATCGCTTCATCAGTCCTTTT 33400
   L  R  I  F  S  K  F  P  S  C  I  I  Q  I  H  I  Y  S  E  V  D  L  C  L  K  D  Y  R  K  M  L  G  K   106

GTTTTTTAAAGTTCTTAAACAAAAAATGATTTAGTGTAGAAAATACTTGATAACAGTTGAAAGCCAAATATTTGCCAAAATCGGTTTCAATTTGCGTTT 33500
T  K  K  F  N  K  F  L  F  H  N  L  T  S  F  V  Q  Y  C  N  F  A  L  Y  K  G  F  D  T  E  I  Q  T  E   72

CGCTGTTGCCAAAAAAGAAAGCATTGTCTGTCATTGCCGCGACTGCAGAGTGTCTGATTTTATCGTCTGAACATTGTTTGCTCAAAGATATGTCGAACAA 33600
   S  N  G  F  F  F  A  N  D  T  M  A  A  V  A  S  H  R  I  K  D  D  S  C  Q  K  S  L  S  I  D  F  L   39

TAGCAAATCATCCACTGTAGCATTCACATCAAAATCCACAACTGCTAAGGGACAGAAATCCGGATAGAGTTCTCCATAAATCACAGGAACGCAAAACAGC 33700
   L  L  D  D  V  T  A  N  V  D  F  D  V  V  A  L  P  C  F  D  P  Y  L  E  G  Y  I  V  P  V  C  F  L     6

AAGATAGTCCACATGTTGAATTAGAAAAGAGTGATGAGAAAGAAAGCTTTTATTTCTTAATATATATTTGACGACTGAACGTCACAGTTACGCCCATTTT 33800
L  I  T  W  M                                                                                          1

ACGGTGATAAACGTTAGGGGTTCCGCAAAAATGTGTGCATTGATCTTTCTGTGATGACACAGTGGTTAGCAAAGCACATCTAACATAATAAGTCTGTTTT 33900

TCTCAGTCATAGGATTCCACAACAGGATAAGACATTTGTGCTGGTCTGATTCCTAATTACATGAGCTGTATTTACGCCAGAATTTTGTTTATTTCCTTTA 34000
                                                                                                  ------  ----

ATTACCCTCAAAAAATGAAAACATTGATGAAACATACTTCCCAGTGAAACGCTTTCAACGGAGATCATGCGTTGATGGTAGATAAAACACAACAAGGAGA 34100
---

                  ------
ACTTTGCCCAAAAAAACATTTATTAATCCCAGACACCAATAAAGGCTTGGTTATTGTATGCATTTTTCGGTTTCTTAGAATGGTTGCGTTTGGGGTTTTT 34200
U23                        -  D  W  V  G  I  F  A  Q  N  N  Y  A  N  K  P  K  K  S  H  N  R  K  P  N  K    147

AATTAGTAAAACTAACAACGCTAACGCTACAATACCAATAAAACTAACTGCTGCAATTAAAATCACAAGGAACAAAACGAACTTTTTTAAAGGGCTATTC 34300
 I  L  L  V  L  L  A  L  A  V  I  G  I  F  S  V  A  A  I  L  I  V  L  F  L  V  F  K  K  L  P  S  N    114

AAAAGTTGAGGTTTTGTCAGTTCTGGAGTTGAAATTGTCAGTGTGAAAAATTCAGATGCATTCCCACTGTTTTCATGATTAGTAGTACTGCCACTGGAGA 34400
L  L  Q  P  K  T  L  E  P  T  S  I  T  L  T  F  F  E  S  A  N  G  S  N  E  H  N  T  T  S  G  S  S  F   80

ATGTCGAGGTCTTTGTAAAATTTAAGGATGTAGATTCGACATTCTCTTTAGTTGTTGCGATAGTCGTAGGTATACTCATAGACACAGACATTAAATCGCT 34500
  T  S  T  K  T  F  N  L  S  T  S  E  V  N  E  K  T  T  A  I  T  T  P  I  S  M  S  V  S  M  L  D  S    47

CGTACTCTCTGTATCTGATGTGATTGACGTGGAACTAAAATCTGAAGTTCTTGACATTTCCAGAGTGGGCGTACTAGAAACAGTACTCATGTTAACAAAG 34600
 T  S  E  T  D  S  T  I  S  T  S  S  F  D  S  T  R  S  M  E  L  T  P  T  S  S  V  T  S  M  N  V  F     14
```

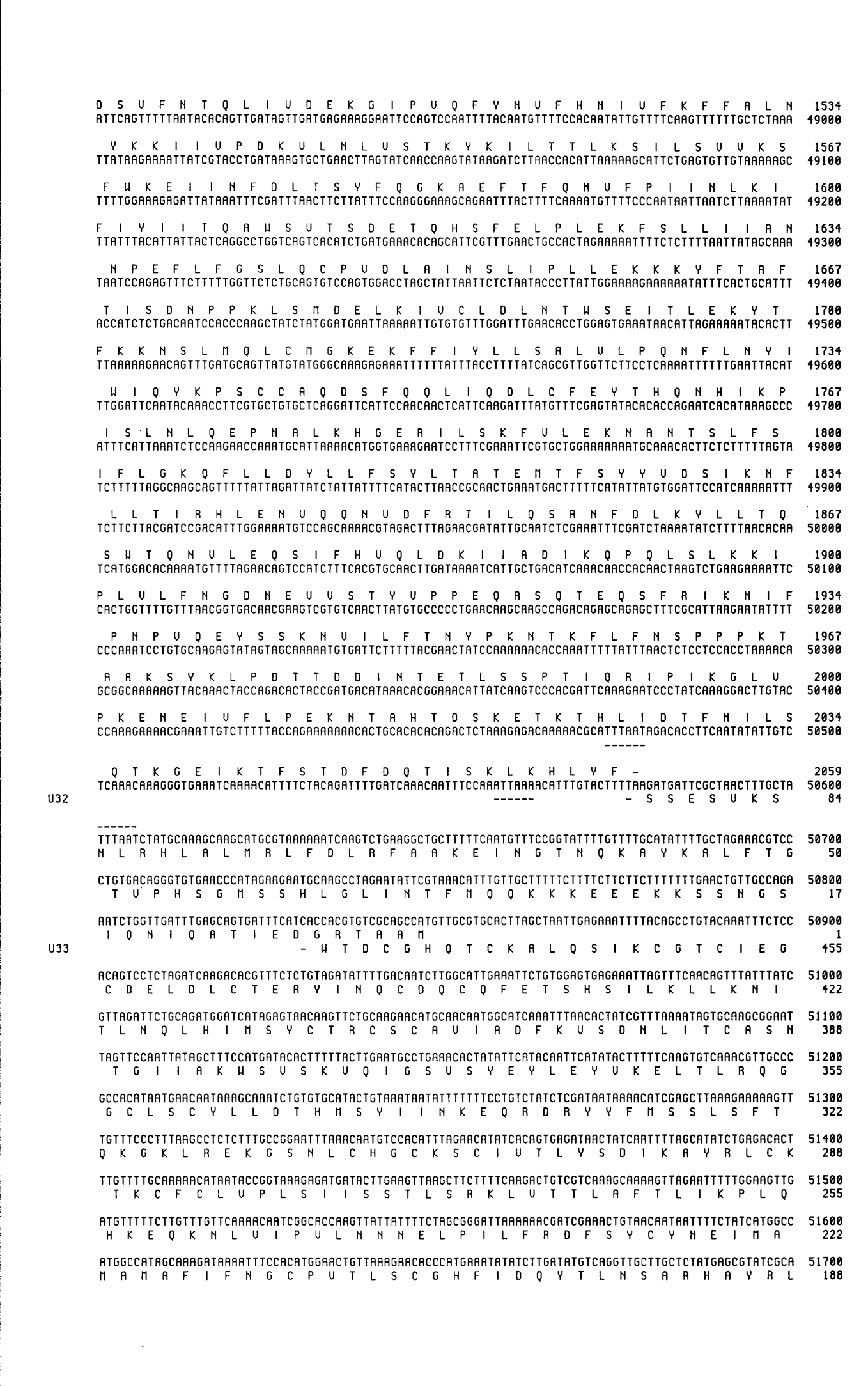
```
      GTAACCATTTCTAAAAAAAATAAAATTCCAAATAACATTTTAAACACTGCGCAATTAGAAGAAGCCTAGTTTAATATTGTACGTGTGCCGACTAAATTAC  34700
      T  V  M  E  L  F  F  L  I  G  F  L  M                                                                  1

      TGTTTTTAAGCAACGCAAAAAAGTGCTGGAATCGGTGTAAAATTCAGGGCTTCGTACCCCCATATCGAACATTGAAGATGTGTAGAATAAGAATCAGACA  34800
U24                                        -  P  K  T  G  G  Y  R  V  N  F  I  H  L  I  L  I  L  C        65

      TAAGATAACCGATATTATGAGACATGCCAGTATTATAAAAGTCCCTCTTCGGCGCACATTTTTGATTTCAGAACTAGAAGTTGAATTTATTGTTCTTGGT  34900
       L  I  V  S  I  I  L  C  A  L  I  I  F  T  G  R  R  R  V  N  K  I  E  S  S  S  T  S  N  I  T  R  P     32

      GATAGATTTTCCTGATGCAAAAAAACGGAATGATCGTGGAACATTTGCAACATTACATCATTATAAGAAGGTGGCGGTGTCTCATGAGTCATCTTCTAGG  35000
      S  L  N  E  Q  H  L  F  V  S  H  D  H  F  M  Q  L  M  V  D  N  Y  S  P  P  P  T  E  H  T  M           1
U24a                                                                                                  -  P     57

      GATATTTAGAGCACCTGATAAATGATAACATATTCTCTGCTATTGAAGGACTTATATAATCATGATATGCAATGCAAGCCAAAATAAACAGAACAAACAA  35100
        Y  K  S  C  R  I  F  S  L  M  N  E  A  I  S  P  S  I  Y  D  H  Y  A  I  C  A  L  I  F  L  V  F  L     23

      ATAATAAATCCCTTTCCAACAATCTTCATAGGTCTCTGGGAAAAAATGGTTGTACTTATAAAACATTATGTGCAGAATTATTTTTTCTACCAACACTTTT  35200
U25                                                                 -  L  V  N  H  A  S  N  N  K  R  G  V  S  K    307
       Y  Y  I  G  K  W  C  D  E  Y  T  E  P  F  F  H  N  Y  K  Y  F  M                                       1

      CCATAGTAATGGTTGTTTAAATATGTTCCTTTTTTTCCTTTTACAAGTATTTCCCCATGGACAGTTGATCTGCCGTGAAAGCTCTTCTTCCACATTTCTA  35300
       W  L  L  P  Q  K  F  I  N  R  K  K  R  K  C  T  N  G  W  P  C  N  I  Q  R  S  L  E  E  E  V  N  R    274

      ACATCGTTTGTTCTTTCAAACCAATACCCATGTAGTCGACGATTTTCATTAAATCGTAAACAACCGATTTTCAAAAATTCAGTTATGTTATGTGCGAGAA  35400
      V  D  N  T  R  E  F  W  Y  G  H  L  R  R  N  E  N  F  R  L  C  G  I  K  L  F  E  T  I  N  H  A  L  F   240

      AAATGATTTGTTCATCATCTACACCAAAAATTTTCCCATTCACATCGAATATTATCACAAGTTCAAACCACTTCCCAAAAACGTTGTGTTTTACACAAAA  35500
        I  I  Q  E  D  D  V  G  F  I  K  G  N  V  D  F  I  I  V  L  E  F  W  K  G  F  V  N  H  K  V  C  F   207

      AAGAACATCTACAATGTTGCAACGAATGACCCTTTTCCAACGGTGAAGGTCTTGATCAGTATAAGGTGTCATGGGAAGTCTGCAAAATCGTGTGTATGTT  35600
       L  V  D  V  I  N  C  R  I  V  R  K  W  R  H  L  D  Q  D  T  Y  P  T  M  P  L  R  C  F  R  T  Y  T   174

      TTATAAGGACTACTTTTCAATATAAAACAAACGTTTATATTTTCATCCCTATAAATGATCATATCAGAGAAAGATTTGCTATGCTCCACGATTTTGCTAC  35700
      K  Y  P  S  S  K  L  I  F  C  V  N  I  N  E  D  R  Y  I  I  M  D  S  F  S  K  S  H  E  V  I  K  S  G   140

      CAGTTTCAAAGTCCACGATAAAATCACACAATTCGTAGATGGGGTAGAAAAACTTCAATCCTACTAAAAAGAAATCTTGTAAGTTTGATGAAACATAGTA  35800
        T  E  F  D  V  I  F  D  C  L  E  Y  I  P  Y  F  F  K  L  G  V  L  F  F  D  Q  L  N  S  S  V  Y  Y   107

      AAGACTTTTGTCAAAAAAACCATTGTAGACATACAAATGCCCTCTGTCACCTTGTAAAAGAATTGGTCCTCTGCAAGTTGATTTTTTAACATGATGTAAC  35900
       L  S  K  D  F  F  G  N  Y  V  Y  L  H  G  R  D  G  Q  L  L  I  P  G  R  C  T  S  K  K  V  H  H  L    74

      ACACCAACAACGATCAACGTTTTTTCACAACACAAGTAAGTGTTTCGAATTTCTTCGATGTCTTCCTTTGAAATTTCTGGGATTGTTTCTAATGAGTGAA  36000
      V  G  V  V  I  L  T  K  E  C  C  L  Y  T  N  R  I  E  E  I  D  E  K  S  I  E  P  I  T  E  L  S  H  F   40

      ATGTAACTCGCCATTCTTCTGGGAAAGGCAATACGAAAGATGTACCTTCCTTACTTAAGATGAATTCTTCGAGAAGAGTAAGTTTATTGACACAGCACAA  36100
        T  V  R  W  E  E  P  F  P  L  V  F  S  T  G  E  K  S  L  I  F  E  E  L  L  T  L  K  N  V  C  C  L    7

      ATTCGCTAAGAAATCCATGCTTCAGTTTTAAAATTACCCAAGAAAAAACAAAGGCATAAATATTTCAATTTTTTTTTATTTAGCACATTTAGAAGCTGCA  36200
       N  A  L  F  D  M                                                          ------                    1

      AATTTTCTTTAGAAAAAAATTACAAAGCAAGCTGAACAATAAAAAAGGTTCAACAAATCGCCTGTAAAAATATCTTTTCTACCCACTGTTTTTGAAATTT  36300
U26          -  F  F  I  V  F  C  A  S  C  Y  F  L  N  L  L  D  G  T  F  I  D  K  R  G  V  T  K  S  I  Q   264

      GAGCAAATAGAGCGCTATAACCTAAGCCAAGAAAGAACTCAAACAACAATGAAAAGAAACAATGTTCTGTTATTACCCCTGCTAACAAACACAGAATACT  36400
        A  F  L  A  S  Y  G  L  G  L  F  F  E  F  L  L  S  F  F  C  H  E  T  I  V  G  A  L  L  C  L  I  S   231

      ACACAGAACAAGTTGATATAGTTTAGTGATAAACAAATTACTATGTAAGACTGCCAAGAAAAATAAATGCTGGCAACTTTTTCTTGGTAAATACAATTTT  36500
       C  L  V  L  Q  Y  L  K  T  I  F  L  N  S  H  L  V  A  L  F  F  L  H  Q  C  S  K  R  P  L  Y  L  K   198

      CCAAACCAAACCAGCAGCGTTAAGGTTAAAAATTGCAGCAAATAAACAATAACATCTTTTATAGCCATTTGGTACGGGCCATTCGGTTTATAGATCGGTT  36600
      G  F  W  V  L  L  T  L  T  L  F  Q  L  L  Y  V  I  V  D  K  I  A  M  Q  Y  P  G  N  P  K  Y  I  P  K   164

      TGTATCGCTGAAGACACATGGCAAACTGATGAGAGTTAACCAGAAACTTATAACTGGTTACATATGGCGTTATCAGAAACACTGTAACAAGCAATCTGTA  36700
        Y  R  Q  L  C  M  A  F  Q  H  S  N  V  L  F  K  Y  S  T  V  Y  P  T  I  L  F  V  T  V  L  L  R  Y   131

      ATGAACCAAATCTGATATTTTTAACACAATGGCAATTAGGGTGCCGAAATAAGATGCCATTGAATGCGTTGGAAATATTAGCATGCATCTACAGACTATG  36800
       H  V  L  D  S  I  K  L  V  I  A  I  L  T  G  F  Y  S  A  M  S  H  T  P  F  I  L  M  C  R  C  V  I    98

      GCTCTCAAGCGGATTTCAGAAAAAAATAAAAACGAGCAGATTACCAAAAAAAATTCATCGGTCATTTGCAACACGAGTAATGATAAAGTTATAATTTTGA  36900
      A  R  L  R  I  E  S  F  F  L  F  S  C  I  V  L  F  F  E  D  T  M  Q  L  V  L  L  S  L  T  I  I  K  L   64

      GATCGTTGAAAGGACAATAGTTAAACAAATTCACCCGACGGTAAGTCAATAGCAAACCTCCGCTAAAAAGGAATATTAGCACTACTTGTTCAAACAGCGT  37000
        D  N  F  P  C  Y  N  F  L  N  V  R  R  Y  T  L  L  L  G  G  S  F  L  F  I  L  V  V  Q  E  F  L  T   31

      TTTCTTTATCATAAGATGGAAACAGCTTTCTCCCGGAATTACATTCCCTAAAACAAGGCCCATAGAAAAACTTATCATTACATTCCACATCTTTGTCTTT  37100
       K  K  I  M  L  H  F  C  S  E  G  P  I  V  N  G  L  V  L  G  M  S  F  S  I  M  V  N  W  M           1
U27                                                                                    -  M  G  C  R  Q  R  K   358

      TCGAATTTGGGGTATAGTCAAAATGTGCAGAATCGTCACTTTCGCTGTCCTCTTTCTCATTAAAAAAATTTGTTCCTGTGCCACCATTTTTAGTGGGCAC  37200
        S  N  P  T  Y  D  F  H  A  S  D  D  S  E  S  D  E  K  E  N  F  F  N  T  G  T  G  G  N  K  T  P  V   325

      CATGTAACTTGTAATTTTGTGTTGCGTTTCCTGTTTACAAAAAAAATTATTTCCTTCCTCCACCTTGTGATTTTTCTCTCCTTTTTCCTCTGTTTTTGCT  37300
       M  Y  S  T  I  K  H  Q  T  E  Q  K  C  F  F  N  N  G  E  E  V  K  H  N  K  E  G  K  E  E  T  K  A    292

      TGACGTTCAAATTTTGGGTCGTCTTTAACCTGCTCTTGACTTAAATAAGCTTCTACGGTCAGAAAAACATTTTTAGCGATTAAAAATAATTTGTATTCTT  37400
      Q  R  E  F  K  P  D  D  K  V  Q  E  Q  S  L  Y  A  E  V  T  L  F  V  N  K  A  I  L  F  L  K  Y  E  K   258
```

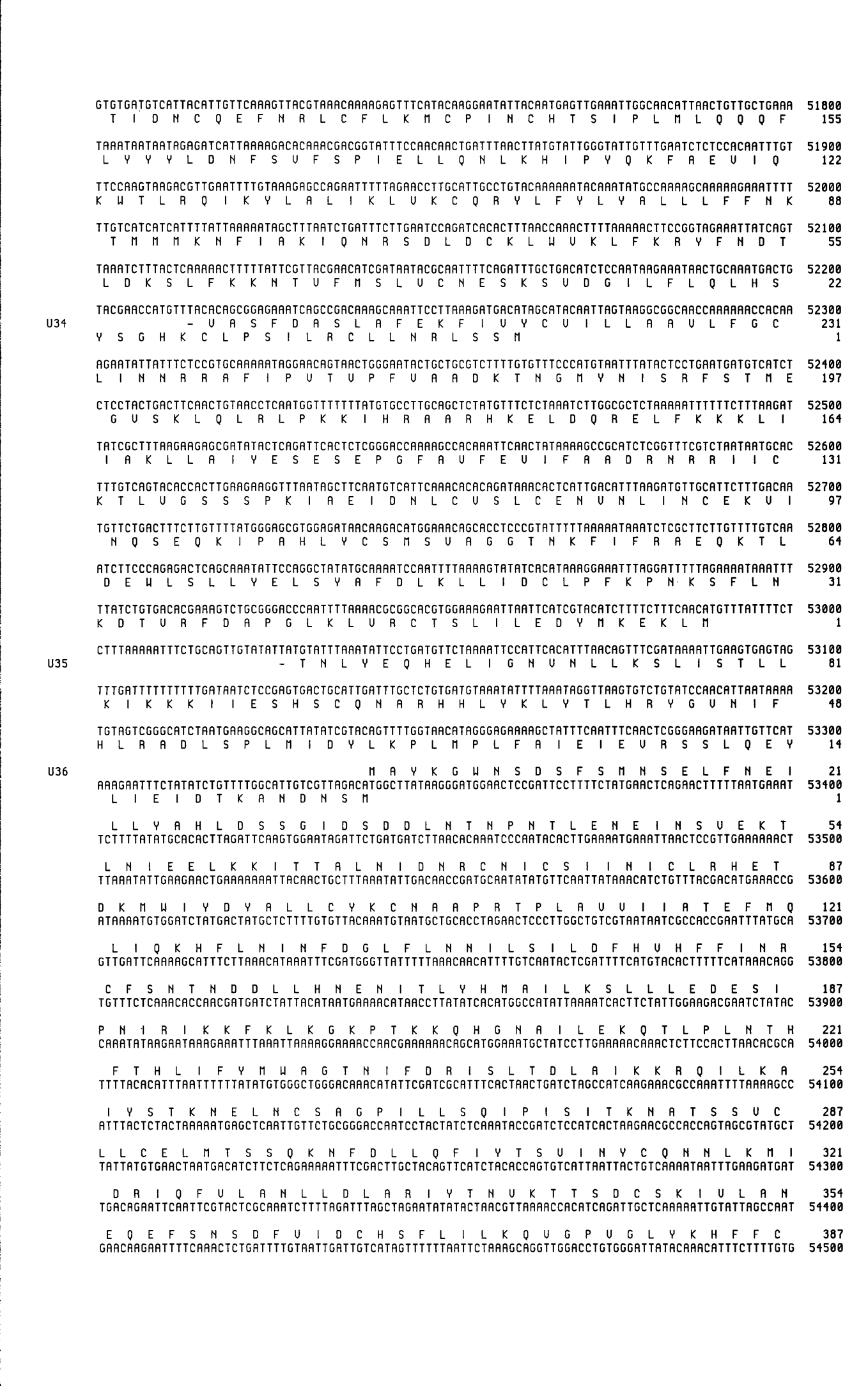


```
ATGGATAGCTGCTTGCAAGATTTTTTCAAGTTCTACTTTGGAGTCACGTACAAATTCAATCAATTCTTTTTGTAAAATATTTTCACTTTCTAAAATAAGT  40400
 H  I  A  A  Q  L  I  K  E  L  E  V  K  S  D  R  V  F  E  I  L  E  K  Q  L  I  N  E  S  E  L  I  L     103

GAGATGGAATCAAATTCCGAAGCAGAGTTTGTTTTGAGATAGAACTGGCCTAACAAACGATCAACGTGCGGTGCATCATGCATTGCTCCCAGTGCTGAAA  40500
 S  I  S  D  F  E  S  A  S  N  T  K  L  Y  F  Q  G  L  L  R  D  V  H  P  A  D  H  M  A  G  L  A  S  I      69

TTATGGCATCACGTAAGCACCACATATCTAAGCCCGAATTTAAGCGCCGGAGAATTTTTGAAATTTCATATAACTGCTTATAAAGAAGAGTTTTGTGATG  40600
   I  A  D  R  L  C  W  M  D  L  G  S  N  L  R  R  L  I  K  S  I  E  Y  L  Q  K  Y  L  L  T  K  H  H      36

GAAGCGTGAATAGCAAAACCGTTTTTTGTTATATTCATCATGACAAAGACATTTCCTCAAGAGACCCATGTTTTTATTTATGGAAACTTTATTGTTAGGA  40700
 F  R  S  Y  C  F  R  K  K  N  Y  E  D  H  C  L  C  K  R  L  L  G  M  N  K  N  I  S  V  K  N  N  P      3

TCCATAATCACAAGAAAATTTTTCTTTCACACAATTCGACAAAGACGTCTTTGAGTCTTATAATGGTTAATGAGATACCCTGTAAGTCTAAAATTCCTGA  40800
 D  M                                                                                                     1
U29        -  L  F  I  K  R  E  C  L  E  V  F  V  D  K  L  R  I  I  T  L  S  I  G  Q  L  D  L  I  G  S    257

TTTTATGTTTTGGGTTTTACAATGACCAATTCGTTGAAGGCTTCCAAAATAGTTTAGATTTCCATCACCGTTGATCAAATAGTTCATAAACATAAATCGT  40900
 K  I  N  Q  T  K  C  H  G  I  R  Q  L  S  G  F  Y  N  L  N  G  D  G  N  I  L  Y  N  M  F  M  F  R     224

TCTTTGGAAAACTTGTTTCTTAATAAATCTATCAATACTTCTTCATGTGTCAGCTCAGATAAAACAATGTAGATCCCGTATTCTATTTTATTAAATGCTT  41000
 E  K  S  F  K  N  R  L  L  D  I  L  V  E  E  H  T  L  E  S  L  V  I  Y  I  G  Y  E  I  K  N  F  A  Q    190

GCCTATAGATATAACAGAGGTAAATGTACTGAATTGCATTTTGTCTTGGTTCCCAGTCAGGTTCCAATAATAAATCTTTCAATATGACCAATCCATATAT  41100
   R  Y  I  Y  C  L  Y  I  Y  Q  I  A  N  Q  R  P  E  W  D  P  E  L  L  L  D  K  L  I  V  L  G  Y  I    157

ATTTTCCTTTCTTACGTCATATACATGACCTTGCATTACACAGGATATTAGACAATTAGTCCCTTGAGAGATATAATTAGTTACTCCAAACAAAGACACC  41200
 N  E  K  R  V  D  Y  V  H  G  Q  M  V  C  S  I  L  C  N  T  G  Q  S  I  Y  N  T  V  G  F  L  S  V     124

TTTTCTTCTAAGTAAGAACAACTATCTAACATTGTGTTGATACTGTCAACTAATGAAGTTTTATACAGATTAGCTAAAGAAGTCCTCGTGCAGAAAGACA  41300
 K  E  E  L  Y  S  C  S  D  L  M  T  N  I  S  D  V  L  S  T  K  Y  L  N  A  L  S  T  R  T  C  F  S  V     90

CATTGAAGGCATTCATGGCTAAAAACAATACACTAATCAACAAAAATTTTGGACTCTCCGCGTCTGTTTCAAGTCTGAACATAAGCATTCCTAAATGGTC  41400
   N  F  A  N  M  A  L  F  L  V  S  I  L  L  F  K  P  S  E  A  D  T  E  L  R  F  M  L  M  G  L  H  D     57

TCGTTCAGGTCCTAAAAGTTCGCAAGGAGTAAAACTTCGCGAAAGCACACACTCTTTGGGAACACCAAAAGCTAGCAAACTTCTGATCTTTTTTTTCTCA  41500
 R  E  P  G  L  L  E  C  P  T  F  S  R  S  L  V  C  E  K  P  V  G  F  A  L  L  S  R  I  K  K  K  E     24

U30                                                                                 M  D  C  T  N  L      6
TTTTCAATGAGATTATTTAAGGAAGTTTCATCAATTGTCTGATGTTTCCTTTTTTGAAGCATTTCCATTGTTCGCAAGCGAAATGGATTGTACAAATTTA  41600
N  E  I  L  N  N  L  S  T  E  D  I  T  Q  H  K  R  K  Q  L  M  E  M                                    1

 Q  F  E  S  I  L  N  L  L  K  T  D  V  S  D  N  T  F  L  T  I  T  A  K  I  E  I  F  S  L  A  A  E     39
CAATTTGAGTCAATATTGAATTTACTTAAAACAGATGTGTCCGATAATACATTCTTAACTATAACAGCGAAAATTGAAATTTTTTCTCTTGCGGCTGAAA  41700

 S  I  T  A  D  K  I  L  S  F  A  K  L  L  P  I  H  N  Y  H  F  N  F  I  K  N  H  M  V  F  Y  I  L  N     73
GTATCACAGCAGATAAGATATTGTCATTTGCGAAATTATTACCAATACACAATTATCATTTCAACTTTATTAAAAATCATATGGTGTTTTATATATTAAA  41800

   Y  S  T  L  G  F  A  K  R  F  S  I  A  E  T  L  C  R  E  L  K  I  F  H  R  T  F  Q  S  T  S  K  L    106
TTATAGTACTCTAGGATTTGCCAAAAGATTTTCAATAGCTGAAACACTATGTAGAGAATTAAAAATTTTCCACAGAACATTCCAATCAACGTCTAAATTA  41900

 Q  N  L  N  N  A  E  V  L  Y  Q  F  E  K  L  I  E  S  I  Q  V  F  K  A  E  L  A  S  P  I  G  K  L    139
CAAAACTTAAATAATGCAGAAGTCCTTTATCAATTTGAAAAACTTATTGAATCCATTCAAGTATTCAAGGCGGAATTGGCATCGCCTATTGGTAAACTTC  42000

 L  R  Q  E  T  M  I  Y  D  K  P  T  E  K  S  V  K  Y  V  I  I  Q  Q  N  L  I  K  I  N  N  L  I  Q  D    173
TACGTCAAGAAACAATGATCTATGACAAACCCACAGAAAAATCAGTTAAATATGTCATTATTCAGCAAAATTTGATAAAAATTAATAACTTAATTCAAGA  42100

   C  E  S  I  K  N  C  R  L  I  A  E  L  I  D  E  L  Y  Q  K  V  Y  R  W  F  M  Q  I  F  T  Y  D  D    206
CTGTGAATCGATAAAAAACTGTAGACTCATAGCGGAGCTGATAGATGAATTATATCAAAAAGTTTATAGGTGGTTTATGCAAATTTTCACATACGACGAT  42200

 I  I  F  P  N  D  N  F  L  D  R  L  L  K  M  D  F  C  Y  T  Y  Y  T  A  S  N  Q  H  L  L  S  L  F    239
ATTATTTTCCCCAATGACAATTTCTTAGATAGACTTTTAAAAATGGATTTCTGCTATACATATTACACGGCATCTAATCAGCATTTATTATCACTTTTTG  42300

 E  Q  T  I  D  N  Q  I  F  I  D  P  S  P  Y  F  E  I  N  P  V  Y  S  P  E  L  Q  F  M  S  T  F  S  L    273
AACAGACAATTGACAATCAAATTTTTATCGACCCTTCTCCATACTTCGAAATAAATCCAGTATATTCTCCCGAATTACAATTTATGAGCACATTTAGCTT  42400

   K  I  F  S  K  N  I  S  A  K  N  E  D  L  Y  I  Y  P  L  L  K  T  N  F  S  I  L  T  F  L  S  L  E    306
AAAGATATTTTCAAAAAATATTTCAGCGAAAAACGAAGATCTATATATTTATCCACTTTTAAAAACAAACTTTTCCATACTTACCTTTCTGAGTTTAGAA  42500

 N  I  F  F  H  H  G  F  I  Y  H  I  L  H  R  T  N  I  T  P  T  E  E  K  K  L  G  K  I  N  G  F  L    339
AATATTTTTTTCCATCATGGATTTATATACCACATCTTGCATAGAACAAATATCACACCAACTGAAGAAAAAAAATTAGGAAAAATAAATGGATTCTTAA  42600

 N  T  V  I  Q  Q  V  L  I  K  K  N  N  L  Q  I  T  F  P  E  L  L  D  K  I  Y  H  L  H  R  I  G  L  N    373
ACACGGTTATTCAGCAAGTCCTAATTAAAAAGAATAACTTACAAATCACTTTCCCGGAGTTGCTTGATAAAATCTATCATTTGCATAGAATAGGCTTGAA  42700

   I  E  T  A  Q  I  F  L  Q  M  L  T  T  Y  K  P  A  T  T  N  N  K  L  Q  T  F  F  T  N  F  S  I  I    406
TATCGAAACGGCACAAATTTTTTTACAAATGTTAACAACATACAAACCAGCGACGACAAATAATAAGTTGCAGACATTTTTCACGAATTTTTCAATAATA  42800

 I  F  S  A  Y  I  F  F  L  C  I  E  L  F  S  P  T  F  I  F  H  N  K  K  K  L  I  L  E  K  Q  K  S    439
ATTTTTTCGGCATATATTTTTTTTCTATGTATTGAGTTATTCAGCCCAACTTTTATTTTCCATAATAAAAAAAAACTTATCCTCGAGAAACAAAAATCCA  42900

 I  I  L  I  L  G  E  Q  F  S  F  I  W  K  E  V  N  E  I  V  D  L  L  F  S  S  T  V  T  E  T  Y  F  K    473
TCATACTAATTCTCGGTGAACAGTTCTCTTTCATTTGGAAAGAGGTAAATGAAATTGTAGATTTGTTATTTAGTAGCACAGTTACAGAAACCTATTTCAA  43000

   F  Y  S  K  G  A  D  D  Y  E  K  D  F  L  Y  K  D  L  M  E  K  W  G  E  L  F  F  P  L  T  Y  S  M    506
ATTTTATTCTAAAGGTGCCGATGACTACGAAAAAGATTTTCTTTATAAAGACTTAATGGAAAAATGGGGTGAATTGTTTTTCCCATTAACTTATTCAATG  43100
```

```
      T  T  P  Q  K  Y  T  D  K  H  V  S  S  T  V  L  K  N  L  C  D  T  A  Y  Q  S  K  M  E  T  A  Y  E     539
ACAACGCCACAGAAATATACCGACAAGCATGTATCAAGTACAGTTTTGAAAAATTTATGTGATACAGCATATCAAAGTAAGATGGAAACAGCGTATGAAA  43200

      S  L  L  P  Y  I  T  H  P  E  F  K  F  I  F  I  T  H  Y  V  R  P  S  L  S  L  I  T  N  L  T  F  E  E     573
GTCTATTACCATATATTACTCACCCAGAGTTCAAATTCATTTTTATTACACATTACGTCAGACCTTCTTTGTCATTAATCACAAACTTAACCTTTGAAGA  43300

        I  K  D  N  R  R  L  L  I  L  I  F  A  C  K  L  L  M  P  S  N  Y  L  L  S  H  Y  L  L  L  L  H  A     606
AATAAAAGATAATCGAAGATTACTAATACTAATATTCGCGTGTAAACTGCTAATGCCATCAAATTATCTTTTATCACATTACTTATTATTGTTACACGCT  43400

      F  T  L  Q  I  F  K  V  D  L  G  H  F  S  I  I  H  A  I  T  Q  K  I  F  D  N  I  N  S  L  T  Q  T     639
TTCACACTACAAATTTTCAAAGTAGATCTTGGCCATTTTTCAATTATACATGCAATTACTCAAAAAATTTTTGACAATATTAATTCTCTGACCCAAACTA  43500

      I  F  I  P  K  T  N  F  L  V  S  L  L  L  T  A  Y  T  V  H  M  Q  T  Y  V  N  P  W  I  Q  K  T  I  S     673
TTTTTATTCCAAAAACAAATTTTTTAGTCAGCCTTTTATTAACAGCATACACTGTGCATATGCAAACTTATGTGAATCCTTGGATACAAAAAACAATCAG  43600

        E  N  I  A  L  L  K  E  Y  I  D  F  T  K  K  C  S  S  T  L  A  T  T  C  Y  L  N  L  E  N  F  A  V     706
TGAAAATATAGCCCTTCTTAAAGAATACATCGATTTCACAAAGAAATGTTCAAGTACTTTGGCAACTACGTGCTATTTGAACTTAGAAAATTTTGCAGTA  43700

      N  M  Y  F  G  K  N  K  V  G  S  T  S  L  S  A  F  Y  R  T  C  S  K  L  I  E  E  S  K  L  F  K  D     739
AATATGTATTTTGGGAAAAACAAGGTTGGAAGTACTTCACTATCAGCTTTTTATAGAACTTGCTCTAAGCTAATTGAGGAATCTAAACTGTTTAAAGATA  43800

      R  L  Q  E  I  K  V  S  K  T  L  F  I  E  M  L  Q  N  V  V  K  N  I  T  K  F  K  D  L  V  S  N  Q  T     773
GGCTTCAAGAAATCAAAGTTTCAAAAACATTGTTTATAGAGATGCTGCAAAATGTTGTGAAAAACATCACAAAATTTAAGGATTTAGTCTCAAATCAAAC  43900

        L  Q  N  F  I  I  I  V  E  R  I  S  S  H  A  N  T  T  Y  Q  D  V  L  N  S  I  D  E  C  H  F  S  N     806
TTTGCAAAATTTTATAATTATTGTTGAAAGAATCTCCTCTCATGCAAATACAACATATCAAGATGTTCTCAATAGCATAGATGAATGCCATTTTTCAAAC  44000

      M  Q  L  I  Q  S  F  K  N  I  V  Y  V  I  D  V  L  N  T  K  N  I  F  N  F  S  L  A  S  Q  L  I  E     839
ATGCAACTTATCCAATCTTTTAAAAACATCGTGTATGTCATTGATGTTCTGAATACTAAAAACATCTTTAACTTTTCTCTCGCATCACAATTAATTGAAG  44100

      A  K  K  L  V  K  K  Q  D  T  Y  N  Q  L  N  V  Q  D  D  F  V  T  V  L  K  S  H  L  N  N  L  F  E  K     873
CGAAAAAACTTGTAAAAAAACAGGACACCTATAATCAATTAAATGTGCAAGATGATTTTGTCACTGTATTAAAGTCACATCTAAATAATCTGTTTGAAAA  44200

        Q  K  P  T  I  N  I  E  R  R  F  M  L  E  G  I  P  D  I  K  Q  I  P  F  L  D  V  F  D  E  R  Y  R     906
GCAGAAGCCTACAATTAATATTGAAAGAAGATTTATGTTAGAAGGAATACCCGACATAAAACAGATTCCATTCTTGGATGTTTTTGATGAAAGATATAGA  44300

      L  I  P  Q  I  E  K  Y  L  H  W  Y  I  A  Y  S  E  A  A  Q  A  D  L  V  E  P  L  L  L  K  L  G  -     938
CTAATACCACAAATTGAAAAGTATCTGCATTGGTACATTGCATATAGTGAAGCTGCGCAGGCTGATTTGGTCGAGCCTTTACTCTTAAAACTTGGTTAGA  44400

U31   M  R  I  I  A  G  S  T  N  Q  N  D  P  K  Y  G  P  R  A  G  K  Q  C  M  S  N  C  F  S  F  L  H  T  V     34
TGAGAATCATAGCAGGAAGTACAAATCAAAACGATCCTAAATACGGACCAAGAGCCGGAAAGCAATGTATGTCAAATTGTTTTTCTTTCTTGCATACAGT  44500

        Y  L  N  G  I  N  N  V  L  N  K  E  S  I  D  I  I  M  E  N  G  A  L  L  D  N  I  S  T  T  T  L  K     67
TTATTTGAACGGAATAAACAATGTGTTAAATAAAGAGTCTATTGACATAATCATGGAAAATGGAGCATTATTGGATAATATCAGTACAACGACATTGAAA  44600

      L  E  T  G  N  I  P  E  Y  R  F  F  T  E  I  P  K  K  I  S  S  N  F  G  E  T  I  H  E  L  S  R  P     100
CTCGAAACTGGCAATATCCCAGAATATCGATTTTTCACAGAAATCCCAAAAAAAATTAGTTCTAATTTTGGCGAAACAATACATGAATTATCTAGACCCT  44700

      F  N  G  T  L  E  S  Q  H  I  D  N  E  V  Y  L  G  L  L  D  F  L  L  Y  G  K  N  K  K  P  A  F  I  V     134
TTAATGGTACCTTAGAATCACAACATATAGATAATGAAGTTTATCTTGGACTGTTAGACTTTCTATTGTATGGGAAAAATAAGAAACCAGCTTTTATTGT  44800

        I  T  I  G  V  M  A  R  A  I  F  I  V  D  E  L  F  Y  L  F  D  S  H  A  S  D  T  E  N  S  A  A  I     167
CATCACTATAGGGGTAATGGCACGAGCTATATTTATAGTTGATGAATTGTTTTACCTTTTTGATTCACATGCATCAGACACAGAAAACTCTGCAGCCATC  44900

      Y  I  C  E  D  I  D  E  L  Y  A  L  L  A  I  E  N  V  A  E  F  Y  Y  D  A  V  F  S  Y  F  I  E  T     200
TATATCTGTGAGGATATTGACGAATTATATGCTCTATTGGCCATAGAGAATGTTGCGGAATTTTACTATGATGCAGTTTTTTCATATTTCATTGAAACGA  45000

      T  D  L  S  L  E  D  G  D  A  T  I  L  I  L  K  T  Y  K  D  P  D  I  A  L  S  L  N  D  F  L  S  M  Y     234
CTGATTTATCTCTTGAAGACGGAGATGCAACAATTTTGATTTTAAAGACTTACAAAGATCCAGATATAGCTCTTAGTTTGAATGATTTTTTATCAATGTA  45100

        S  S  T  S  S  T  K  T  A  E  T  N  T  L  I  S  K  Q  S  P  S  K  R  K  Q  E  K  T  S  L  N  S  N     267
TTCATCTACATCCTCAACAAAGACAGCGGAAACAAACACTTTAATTTCAAAACAATCACCAAGCAAACGCAAACAAGAAAAAACCAGTCTAAATTCAAAT  45200

      S  L  E  K  K  R  K  Q  G  S  S  L  K  Y  Y  N  N  E  V  D  L  V  P  S  F  Y  E  L  R  P  Q  F  N     300
TCTCTAGAAAAAAAAAGGAAGCAGGGCTCATCACTCAAATACTATAACAATGAAGTAGATTTAGTACCAAGTTTTTATGAGCTTAGACCTCAATTTAACA  45300

      N  I  L  F  E  L  S  N  F  P  I  V  K  E  N  V  N  W  T  L  Y  I  Q  K  F  A  T  K  S  T  Q  P  F  T     334
ATATTTTATTTGAGCTTTCTAATTTCCCAATTGTAAAGGAAAATGTAAATTGGACCCTTTACATACAGAAATTTGCAACAAAGTCTACACAGCCATTTAC  45400

        K  P  F  I  W  N  R  V  F  H  L  F  S  Q  V  V  D  A  L  I  M  I  K  N  D  H  W  D  E  T  Q  Q  Q     367
AAAACCTTTTATATGGAATAGGGTATTCCATCTATTTTCTCAAGTGGTTGACGCCTTAATTATGATTAAAAACGATCATTGGGATGAGACACAACAGCAA  45500

      K  Q  F  F  T  H  F  L  P  F  K  E  F  S  E  E  F  E  N  A  L  E  A  C  R  E  N  N  N  L  D  L  I  L     400
AAACAATTCTTCACACATTTCTTGCCGTTCAAAGAATTTTCTGAGGAATTTGAAAACGCTCTAGAGGCTTGCCGAGAAAATAATCTTGATCTAATATTGC  45600

      L  Y  K  N  Y  L  S  K  T  T  A  F  K  N  L  E  R  I  L  L  T  K  F  S  A  I  V  S  P  V  H  E  K  H     434
TTTATAAAAACTATCTTTCGAAAACTACTGCATTCAAAAACCTTGAAAGAATTTTATTAACGAAGTTTAGCGCCATTGTCAGTCCGGTGCATGAAAAACA  45700

        Y  T  L  V  N  T  W  L  T  N  L  I  Q  K  L  V  K  H  P  E  D  T  N  A  F  I  N  D  Y  V  L  K  N     467
CTACACGCTTGTAAACACATGGCTAACTAACTTAATACAAAAACTAGTCAAACATCCCGAGGATACCAATGCTTTCATCAATGACTACGTGTTAAAAAAC  45800

      P  L  N  H  F  I  C  L  N  K  K  E  K  Q  S  I  A  L  L  L  N  K  K  R  M  S  M  L  K  D  V  E  I     500
CCCTTAAATCATTTCATTTGTTTGAATAAGAAGGAAAAGCAGAGCATCGCTCTACTACTGAATAAAAAAAGAATGAGTATGCTAAAAGATGTGGAAATCG  45900

      E  K  N  G  F  V  Q  L  Q  A  F  I  E  N  I  G  E  A  P  A  N  Y  L  D  P  E  N  A  R  K  V  N  V  E     534
AAAAGAATGGTTTTGTTCAACTCCAAGCATTCATCGAGAACATAGGAGAAGCTCCAGCAAATTATTTAGATCCGGAAAATGCACGCAAAGTGAATGTTGA  46000
```

```
  E  V  S  E  K  D  I  P  T  L  S  T  D  K  V  S  I  P  N  E  S  M  F  T  S  N  K  K  H  S  I  E  K    567
AGAAGTCTCAGAAAAAGACATCCCAACATTATCCACAGATAAAGTTTCCATACCCAATGAAAGTATGTTCACATCAAATAAAAAACACAGCATAGAAAAG  46100

  L  I  H  A  K  L  K  A  I  L  S  T  M  G  Q  R  L  T  R  I  I  Q  E  N  Y  N  N  I  A  A  G  F  L    600
TTAATACATGCTAAGCTGAAAGCCATTTTGAGTACAATGGGGCAAAGATTAACTAGAATTATTCAAGAAAATTATAATAACATCGCTGCAGGTTTTCTGC  46200

  P  V  N  D  L  N  N  L  F  A  Y  L  V  K  L  Y  F  D  V  Y  S  I  T  I  N  G  F  V  V  E  N  E  L  I    634
CGGTGAACGATCTTAATAATCTGTTTGCCTATTTGGTCAAACTCTATTTTGATGTCTATAGCATCACCATTAACGGATTTGTGGTGGAAAACGAATTGAT  46300

  K  N  I  E  Q  I  Y  D  N  T  Q  Y  L  R  F  G  L  T  R  F  N  M  Q  N  L  T  P  F  T  I  S  V  R    667
AAAAAATATTGAACAAATTTACGACAATACGCAATATCTGAGATTTGGATTGACACGCTTCAATATGCAAAATTTGACACCGTTTACTATATCTGTCCGC  46400

  K  M  F  L  D  F  F  L  S  Q  K  T  L  I  D  R  A  E  E  I  I  E  N  L  E  F  K  S  V  T  P  E  G    700
AAAATGTTCCTGGATTTTTTTCTATCACAAAAAACTCTGATAGATAGAGCTGAAGAAATTATAGAGAACCTTGAGTTTAAATCAGTTACACCAGAAGGAA  46500

  K  Q  K  L  A  T  K  N  M  L  R  E  Q  L  E  Q  L  N  A  M  D  V  D  D  T  I  N  L  K  T  D  T  L  T    734
AACAAAAACTTGCCACAAAGAATATGCTCAGAGAACAATTAGAACAGTTGAATGCTATGGATGTGGATGATACAATAAATCTGAAAACAGACACATTAAC  46600

  H  Q  V  L  F  S  D  Q  E  L  R  M  I  Q  D  F  I  L  Q  L  S  I  H  N  I  P  S  I  N  F  V  K  S    767
ACATCAAGTATTATTTTCAGACCAAGAATTACGCATGATACAAGACTTCATTTTACAACTCTCCATTCACAATATTCCAAGCATTAACTTTGTGAAATCT  46700

  L  K  L  H  I  I  L  E  K  R  P  D  I  L  L  A  L  Q  E  K  V  Q  N  I  L  Y  F  Y  F  Q  D  L  V    800
TTGAAATTACATATTATTTTAGAAAAAAGACCTGATATACTATTAGCTCTACAAGAAAAAGTCCAGAATATTCTATATTTTTATTTTCAAGATCTAGTTA  46800

  N  E  I  P  A  Q  E  N  V  L  S  T  M  L  F  I  I  E  L  F  P  A  D  S  R  I  H  L  L  E  T  G  Y  I    834
ACGAGATACCTGCTCAAGAAAATGTTTTGTCAACAATGTTATTTATAATAGAGCTTTTTCCAGCCGACAGTAGAATACATCTACTAGAAACCGGATATAT  46900

  S  R  H  I  V  K  K  W  L  N  M  K  S  L  Q  D  A  E  D  L  I  R  F  I  N  I  N  K  E  Q  L  G  K    867
TTCCAGACATATTGTAAAAAAATGGCTAAACATGAAATCATTGCAAGATGCTGAGGATTTAATTCGATTTATAAATATTAATAAAGAACAACTAGGAAAA  47000

  F  E  H  Q  P  F  G  K  E  I  Q  K  L  I  E  K  I  H  L  F  Y  K  Q  K  V  I  E  Y  Q  E  D  V  W    900
TTTGAACATCAGCCATTTGGAAAAGAAATTCAAAAACTAATTGAAAAAATACACTTGTTCTATAAGCAAAAAGTAATCGAGTATCAGGAAGATGTCTGGA  47100

  S  E  M  A  K  N  I  I  L  T  S  P  S  E  L  S  Q  F  L  A  S  A  P  T  Q  R  I  I  Q  K  H  K  N  N    934
GTGAAATGGCTAAAAATATAATTTTAACTTCACCCTCTGAATTATCTCAATTTTTAGCTTCAGCTCCCACTCAACGTATCATACAAAAACACAAAAATAA  47200

  L  D  Q  K  L  L  I  H  M  E  N  Q  A  K  Q  A  M  E  D  D  K  K  R  V  A  C  S  K  I  N  L  E  R    967
TTTAGATCAGAAACTTTTAATACATATGGAAAACCAAGCCAAGCAAGCAATGGAAGATGACAAAAAAAGAGTTGCCTGTTCTAAAATTAATCTGGAACGA  47300

  H  L  N  D  L  L  L  L  L  K  D  R  Q  F  A  S  I  Q  A  S  V  L  I  V  C  E  N  I  F  K  T  I  P   1000
CACCTGAATGATTTACTGCTCTTATTGAAAGACAGACAATTTGCTTCCATACAGGCTTCTGTTTTGATTGTGTGCGAAAATATATTTAAAACGATACCAG  47400

  D  D  N  L  I  I  Q  F  S  H  A  L  L  S  V  L  L  D  I  E  K  D  L  K  S  Y  S  S  E  I  L  E  K  I   1034
ATGATAACCTAATTATTCAATTTTCACATGCTCTGCTTTCAGTTTTACTTGACATTGAAAAGGATTTAAAAAGCTATTCATCAGAAATATTAGAGAAAAT  47500

  L  I  N  R  P  L  E  T  S  R  L  L  V  F  K  D  A  Y  G  N  L  K  E  F  L  N  A  L  K  Q  S  L  F   1067
ACTAATAAATAGGCCCCTCGAAACCAGTAGATTATTAGTGTTTAAAGACGCTTATGGTAATCTGAAAGAGTTTTTAAACGCCTTAAAACAATCACTTTTT  47600

  A  T  A  D  V  Q  N  K  A  D  F  L  I  Q  I  L  D  F  T  Y  K  F  R  H  K  T  N  K  G  K  L  L  H   1100
GCCACAGCGGATGTTCAAAACAAGGCTGATTTTCTTATCCAAATTTTAGATTTTACCTATAAATTTAGACATAAGACAAATAAAGGTAAACTTCTACATT  47700

  S  I  Y  N  E  D  F  K  L  Y  E  E  T  L  T  E  L  R  K  K  A  T  D  A  K  E  S  L  T  K  L  F  K  A   1134
CCATTTATAATGAGGATTTCAAACTATACGAAGAAACATTAACAGAGTTAAGAAAAAAAGCAACAGATGCAAAAGAGTCGTTAACTAAACTTTTTAAAGC  47800

  S  E  Q  K  I  E  L  S  R  T  I  P  L  K  E  I  Y  L  N  I  E  T  V  N  F  Q  G  Y  G  N  V  V  F   1167
ATCCGAACAAAAGATCGAGCTGTCGCGTACGATTCCGTTAAAGGAAATCTACCTGAACATAGAAACTGTTAATTTCCAAGGTTATGGCAACGTAGTTTTT  47900

  R  E  S  A  F  K  R  A  I  E  V  E  I  K  N  Y  E  M  K  L  N  D  L  I  K  H  F  N  S  H  L  K  T   1200
CGAGAAAGTGCTTTTAAACGCGCAATAGAGGTAGAAATAAAAAATTACGAAATGAAATTAAACGATCTAATAAAACACTTTAATTCACACTTAAAAACAA  48000

  K  I  D  H  M  Q  I  L  N  L  S  F  D  N  K  W  K  D  F  V  S  K  S  K  I  S  F  P  P  E  L  T  I  S   1234
AAATTGACCACATGCAGATTCTTAATCTATCTTTTGATAACAAATGGAAAGATTTTGTCTCCAAGTCAAAAATATCTTTCCCACCAGAACTGACAATAAG  48100

  S  Q  E  L  I  K  D  P  I  K  V  I  T  E  T  L  N  K  A  S  N  D  L  A  Y  V  I  S  E  K  I  L  K   1267
TTCACAAGAGTTGATCAAGGATCCCATTAAAGTTATAACTGAAACTCTAAACAAAGCCTCAAACGATTTAGCGTATGTGATTAGTGAAAAAATATTGAAG  48200

  W  L  I  V  F  V  K  E  L  N  T  F  F  V  A  T  M  S  E  F  G  E  V  I  P  F  D  Y  K  H  F  R  A   1300
TGGTTAATAGTTTTTGTCAAAGAACTGAATACCTTTTTTGTAGCTACAATGTCAGAATTTGGAGAAGTTATCCCCTTTGACTATAAACATTTCAGAGCTT  48300

  L  E  Y  E  I  N  S  K  Y  I  E  I  E  N  K  I  I  C  N  E  I  I  E  N  T  D  N  I  E  K  L  S  T  L   1334
TGGAATACGAAATTAATTCTAAGTACATAGAGATTGAAAATAAAATAATCTGCAACGAAATTATCGAAAATACTGACAATATAGAAAAACTCTCAACCTT  48400

  I  K  Q  I  D  P  N  R  I  A  G  G  K  Q  K  F  Q  D  Y  L  S  K  I  L  T  A  E  T  N  Q  Q  Q  T   1367
GATAAAACAAATAGATCCAAATCGTATTGCTGGTGGTAAACAGAAATTTCAGGATTATCTGAGCAAAATTCTAACAGCTGAAACGAACCAGCAACAAACA  48500

  R  Y  K  E  Q  L  K  K  Q  Y  F  D  L  L  D  N  I  A  H  F  R  F  A  F  D  F  N  H  Q  Q  N  L  I   1400
CGCTATAAAGAACAGTTAAAAAAACAGTACTTTGACCTTTTAGATAACATCGCCCATTTCCGGTTCGCATTCGATTTTAACCATCAACAAAATTTAATTT  48600

  L  K  L  K  D  K  F  K  T  L  R  T  D  T  V  F  E  R  F  P  N  L  D  D  T  F  V  S  S  M  N  V  E  N   1434
TAAAACTGAAAGACAAATTCAAAACTCTTAGAACAGACACTGTATTTGAAAGATTTCCAAATTTAGATGATACTTTTGTCAGTTCAATGAATGTCGAGAA  48700

  F  L  Q  A  L  E  A  L  S  H  F  V  Q  A  A  Q  N  F  L  Q  N  V  L  T  E  Q  A  D  L  F  P  Q  T   1467
CTTTTTGCAGGCACTTGAGGCTTTAAGCCATTTCGTGCAGGCAGCACAAAATTTTCTACAAAACGTTTTAACAGAGCAAGCGGATTTATTTCCACAGACG  48800

  N  F  I  P  V  E  L  S  T  V  K  T  I  P  K  S  D  I  N  L  R  M  K  I  H  T  P  Q  T  F  F  Q  V   1500
AATTTCATTCCTGTCGAACTTTCCACCGTCAAAACAATTCCAAAATCAGATATAAATTTACGTATGAAAATACACACACCCCAAACTTTTTTTCAGGTTG  48900
```

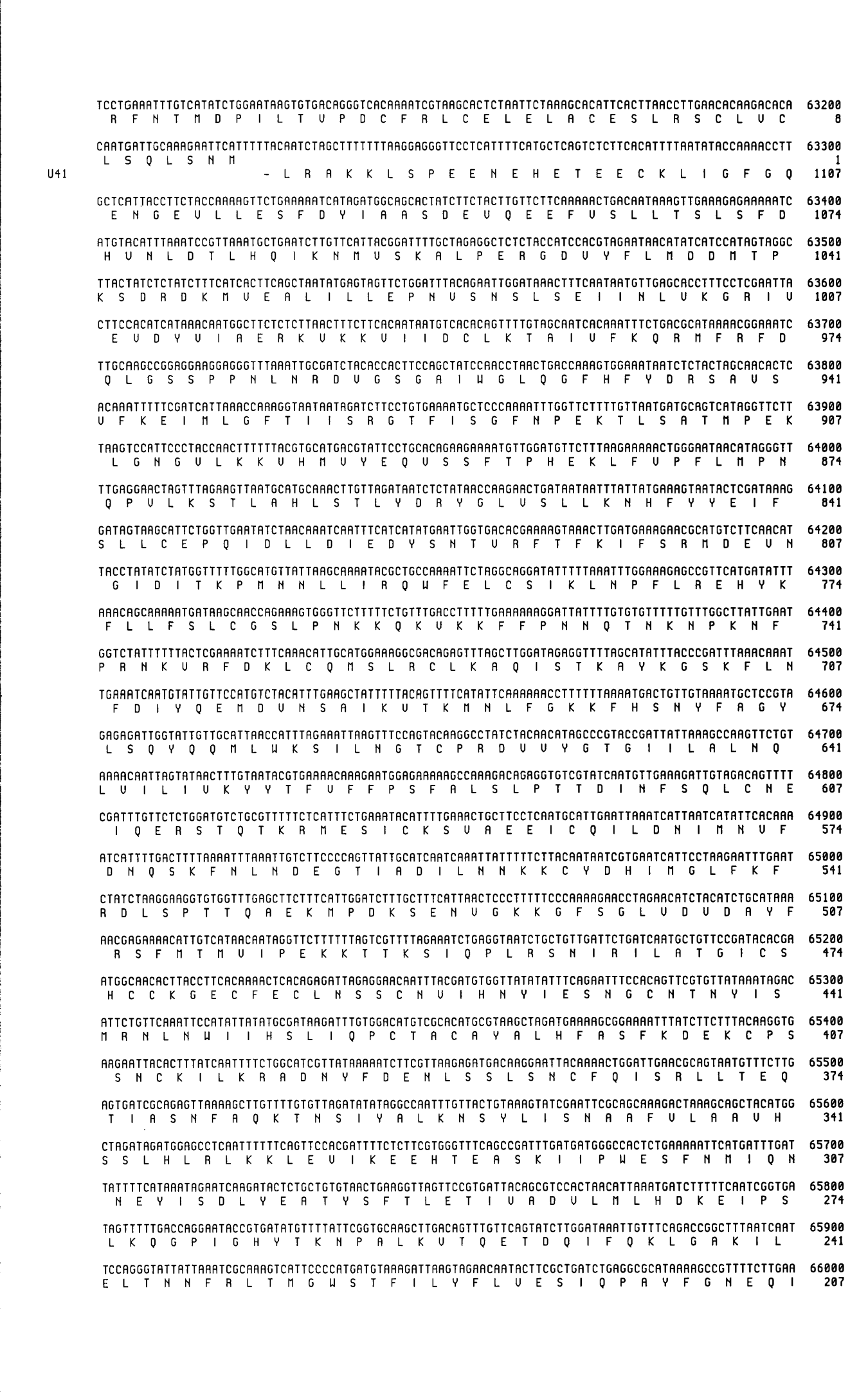


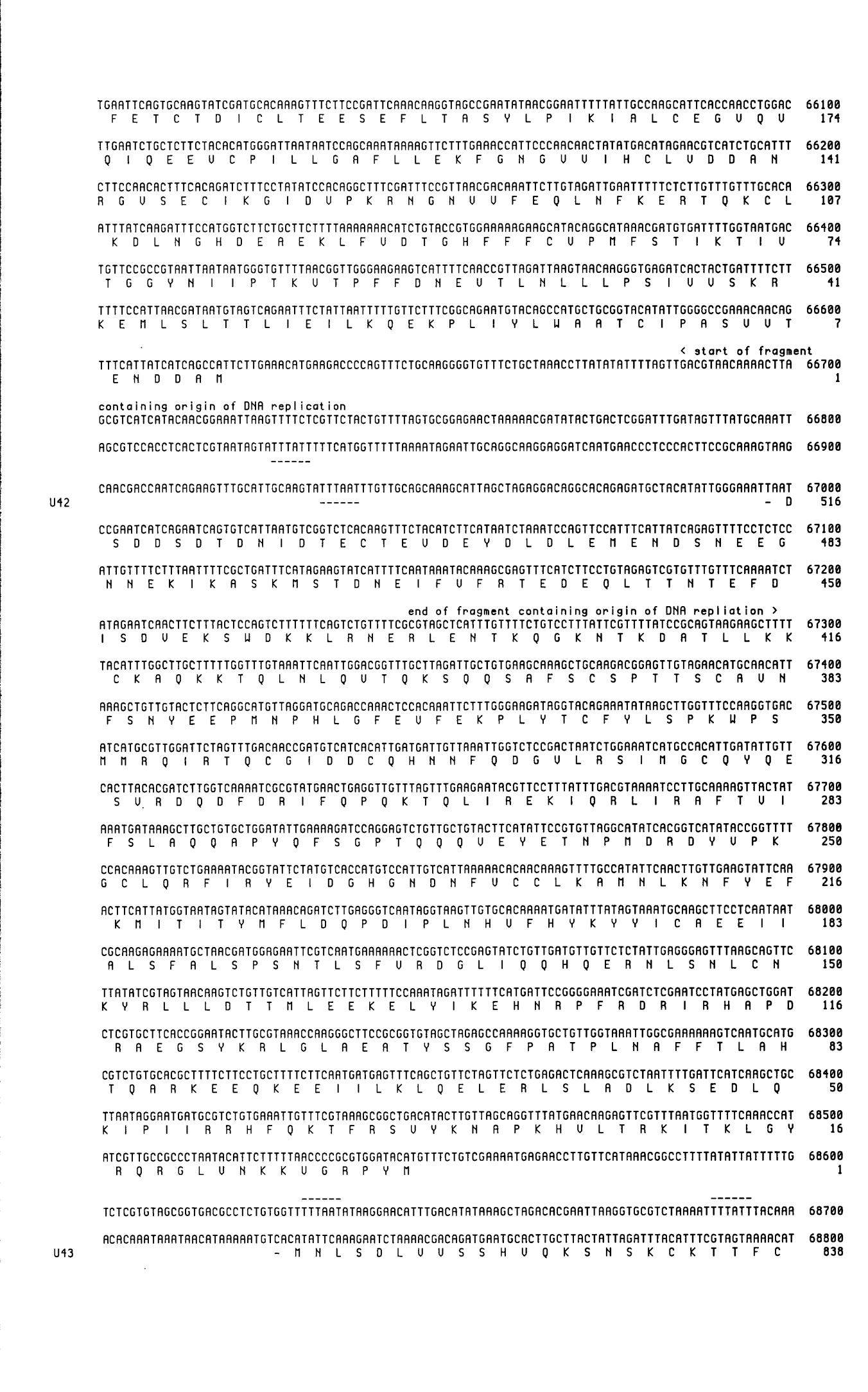


```
      D  P  L  C  I  A  N  I  K  T  I  K  P  H  I  L  F  Y  T  T  E  S  C  I  L  Q  D  F  K  V  A  I  C  Y    421
      ATCCACTGTGCATAGCAAACATTAAGACAATAAAACCCCATATTTTATTCTATACAACAGAAAGCTGTATACTACAAGACTTCAAAGTTGCAATTTGTTA  54600

        Q  N  E  Y  L  N  S  V  E  K  H  V  W  L  A  I  H  F  F  K  A  F  Q  V  S  K  L  N  H  K  N  K  T    454
      TCAGAATGAGTATTTAAACAGTGTTGAAAAACATGTTTGGTTAGCCATTCATTTTTTTAAAGCATTTCAGGTTTCAAAATTAAACCACAAAAATAAAACG  54700

U37                                                                                                   M      1
      L  I  S  D  F  L  K  D  F  T  Q  L  L  A  D  Q  N  F  E  I  V  D  P  T  F  T  I  H  Y  Y  V  -        485
      CTTATATCTGATTTCTTAAAGGATTTCACACAGCTCTTAGCAGATCAAAACTTTGAAATTGTCGATCCTACATTTACCATTCATTATTACGTTTAGCATG  54800

        I  I  Q  S  T  R  R  L  R  R  A  S  S  L  L  K  K  S  K  P  Y  N  K  E  K  T  N  L  S  L  S  L  S     34
      ATTATTCAAAGTACACGTAGACTGAGACGAGCATCTAGCTTGTTAAAGAAAAGCAAACCTTACAACAAAGAAAAAACTAACTTATCTTTATCTTTGTCAC  54900

      L  K  E  L  H  S  V  F  K  L  F  P  E  Y  E  L  K  F  L  N  M  M  K  L  P  I  T  G  K  E  P  I  K  I     68
      TTAAAGAACTCCATTCGGTTTTCAAATTATTTCCAGAGTATGAATTGAAATTTCTAAATATGATGAAACTTCCAATAACCGGTAAAGAACCTATCAAAAT  55000

        P  F  D  L  S  L  H  H  Q  H  T  C  L  D  L  S  P  Y  A  N  E  Q  V  S  K  S  A  C  V  N  C  G  T    101
      TCCATTCGATCTAAGTCTACATCATCAACATACGTGCTTAGACTTATCACCATATGCCAATGAACAAGTTTCAAAAAGTGCATGTGTTAATTGTGGTACA  55100

        T  N  I  P  T  A  S  D  A  M  V  A  Y  M  N  Q  I  S  N  V  M  Q  N  R  L  Y  Y  Y  G  F  Q  K  K    134
      ACAAACATTCCAACAGCTTCAGATGCTATGGTGGCATATATGAATCAAATTTCAAACGTAATGCAAAATAGATTATATTACTACGGTTTTCAGAAAAAGG  55200

      V  E  L  I  R  M  S  A  K  Q  P  T  L  F  Q  I  F  Y  I  L  S  S  I  A  S  N  F  L  P  I  M  F  E  N    168
      TTGAACTGATTCGTATGTCAGCCAAGCAACCAACGCTTTTTCAAATTTTTTATATTCTTTCAAGCATAGCAAGTAATTTTTTTACCAATCATGTTTGAAAA  55300

        N  E  K  L  N  M  Y  V  V  F  Q  T  R  T  L  H  I  P  C  E  C  I  N  Q  I  M  T  V  S  S  G  Y  T    201
      TAACGAGAAATTAAATATGTATGTTGTTTTCCAAACAAGAACTCTTCATATCCCTTGCGAATGTATCAATCAGATTATGACAGTCTCCTCCGGATACACT  55400

      V  L  L  D  I  L  H  D  S  I  V  L  H  V  L  C  K  T  I  E  T  S  N  I  Q  I  D  I  N  V  L  Q  R    234
      GTACTTTTAGACATCTTACACGACAGCATTGTACTACATGTTCTTTGTAAAACTATTGAGACTAGCAATATTCAAATTGATATTAATGTTCTTCAGCGAA  55500

      K  I  E  E  M  D  V  P  D  E  I  G  D  K  F  E  K  L  K  H  I  L  P  F  I  -                         259
      AAATTGAGGAAATGGACGTACCCGATGAAATAGGTGACAAGTTCGAAAAGCTGAAACACATTCTACCGTTTATTTAAAACAGATCACAGAACATCTGTCA  55600
U38                                                                  ------        - L  V  D  T  L  1009

                                                  ------
      GGTCACAAAAGCTTTGATCTACATATACTTTTAAAGGTAGGATTGAAAGTAAAAACTTTTCTTTCTTAATGCCCGTTTTCGGAAATATTGGAGAGATAGC  55700
        D  C  F  S  Q  D  V  Y  V  K  L  P  L  I  S  L  L  F  K  E  K  K  I  G  T  K  P  F  I  P  S  I  A    976

      ATTTGTTACAGCTTTCATTATTTGTTCAAAATATTTATCTGCATTGATTTTTAAGTTGTTCTCAATAGCATATGCTGGATCTTCTGCTAACTCATAATTA  55800
        N  T  V  A  K  M  I  Q  E  F  Y  K  D  A  N  I  K  L  N  N  E  I  A  Y  A  P  D  E  A  L  E  Y  N    943

                                                  ------
      TGAAATGTTTTTTTTACGGTTTCTGTGGGTGCAATTAAAATGTACATAACTCTATCTCCTATGTTTGGCAACTCTTCCTTTCTTTGAGCTAACCTTTTAA  55900
      H  F  T  K  K  V  T  E  T  P  A  I  L  I  Y  M  V  R  D  G  I  N  P  L  E  E  K  R  Q  A  L  R  K  V    909

      CAACAGCCAAATGTGGTAAATTTGCTTGCTTGTAGGCAGAAATGTCTTTTGATAAGACTGAAGACAAAACCAAACTTTAATGTTAACCCTGTTTAAGAA  56000
        V  A  L  H  P  L  N  A  Q  K  Y  A  S  I  D  K  S  L  V  S  S  L  V  L  S  K  I  N  V  R  N  L  F    876

      AAGTGCGTCTCGTGCGTCACACAACTTTTTAATTATTTTTTGGATACCATCGGGAACTCCATTTTCATAGATTTCTTGTATTGTCATATTAGAGAGCTTT  56100
        L  A  D  R  A  D  C  L  K  K  I  I  K  Q  I  G  D  P  V  G  N  E  Y  I  E  Q  I  T  M  N  S  L  K    843

      TCAGCAGACTTTTGTACATCGACATCCCAAAAGAGAAGATCGATGATGTCTTTAACTACCACTTTAACAAAATCGCAAGAAGTCTTGCGAACTAACTCCA  56200
      E  A  S  K  Q  V  D  V  D  W  F  L  L  D  I  I  D  K  V  V  V  K  V  F  D  C  S  T  K  R  V  L  E  V    809

      CACCTTTAAAAACAAGCGTTGCATCATCTAACTTTCCAATGTAACGTTTCTTGCATATTAAAATTAATGGAAATAGGATCTTTTCAAATTCTAATTTGAT  56300
        G  K  F  V  L  T  A  D  D  L  K  G  I  Y  R  K  K  C  I  L  I  L  P  F  L  I  K  E  F  E  L  K  I    776

      AGGATGTTTAAATAACGTGTTTGTTATGTGGCTTGCAATAGAAGGAGCTATTCTTTTCAAAGCTTTGGGACAAACATTTTTGAAAGTAACAAAAAGACTG  56400
        P  H  K  F  L  T  N  T  I  H  S  A  I  S  P  A  I  R  K  L  A  K  P  C  V  N  K  F  T  V  F  L  S    743

      TCAGTATCACCATAAATCACCTCTATTCCAAATGTCCCTGAAAAATCACCGCGTGTCAATCCAAATTTCTCAATGAAAAAAGTGTCCGAATACATGGCGC  56500
      D  T  D  G  Y  I  V  E  I  G  F  T  G  S  F  D  G  R  T  L  G  F  K  E  I  F  F  T  D  S  Y  M  A  S    709

      TGTCAACGTAATCAACAGTTTTACAAAGCATTTCACGCCCAAGACATGTGACAGAAGCTGCTATTGCCACACATGGAAGTAAGCTATGTGTTGCCCCGGT  56600
        D  V  Y  D  V  T  K  C  L  M  E  R  G  L  C  T  V  S  A  A  I  A  V  C  P  L  L  S  H  T  A  G  T    676

      GACTCCGTAGACTGAATTACAAGTAGTTTTTAATGCAAGCTGCTTCTTATCTAAAAGCATTTCCATCACGGGGTTGTTACACATTTTCATCTGCATTTTT  56700
      V  G  Y  V  S  N  C  T  T  K  L  A  L  Q  K  K  D  L  L  M  E  M  V  P  N  N  C  M  K  M  Q  M  K    643

      ACTTCTTTTCTTTTATCCAGCCATTTTTTCAATAAACTCGCTAAAATTGATTCCCTGACTGTCTTTTTAACAAACCTGTGCGTTACAGGTCCAACATGCA  56800
      V  E  K  R  K  D  L  W  K  K  L  L  S  A  L  I  S  E  R  V  T  K  K  V  F  R  H  T  V  P  G  V  H  V    609

      CAGTGAGAATGTCATCTGCATGTAAACCAATTACTGCATTTTCATCAACAACCAGGGTACTATAACACAGATTATGAGCCATCATAATACTTGGATACAA  56900
        T  L  I  D  D  A  H  L  G  I  V  A  N  E  D  V  V  L  T  S  Y  C  L  N  H  A  M  M  I  S  P  Y  L    576

      ACTTTGGAAATCAAAAACTACTGTTGGAGTAGCATAATAACCTATTTTCGGTTCTAAAACTGTAGCACCTTTATAGCCCACATTTTCTTTGCCTTTATTA  57000
      S  Q  F  D  F  V  V  T  P  T  A  Y  Y  G  I  K  P  E  L  V  T  A  G  K  Y  G  V  N  E  K  G  K  N    543

      AAGTTAGTGTTCATACTTGGTAAAATCATATTTAAATTTTTAGCTTCGTGTAAAATGCATGGGAAAATTTTTTTTTGTTGTCCCTCAAATACGGCGCAAC  57100
      F  N  T  N  M  S  P  L  I  M  N  L  N  K  A  E  H  L  I  C  P  F  I  K  K  Q  Q  G  E  F  V  A  C  R    509

      GAATAGTAATATAAGCGAGACTTGCAACCTCAGCCATTTCATAGTGATAATTAATCTTTGTAAAAAGCTGAACAACCAATAGAGAATCTTGTATACAATA  57200
        I  T  I  Y  A  L  S  A  V  E  A  M  E  Y  H  Y  N  I  K  T  F  L  Q  V  V  L  L  S  D  Q  I  C  Y    476

      TCTGCCAACCACAGCTCTACCTTTGGAACCATCAATAAATTTTTTAGGAATTTCCTTGTAAGAGAGATTTTTTTTTTCCTGATTTAAACAAAGTTTGGCT  57300
      R  G  V  V  A  R  G  K  S  G  D  I  F  K  K  P  I  E  K  Y  S  L  N  K  K  E  Q  N  L  C  L  K  A    443
```

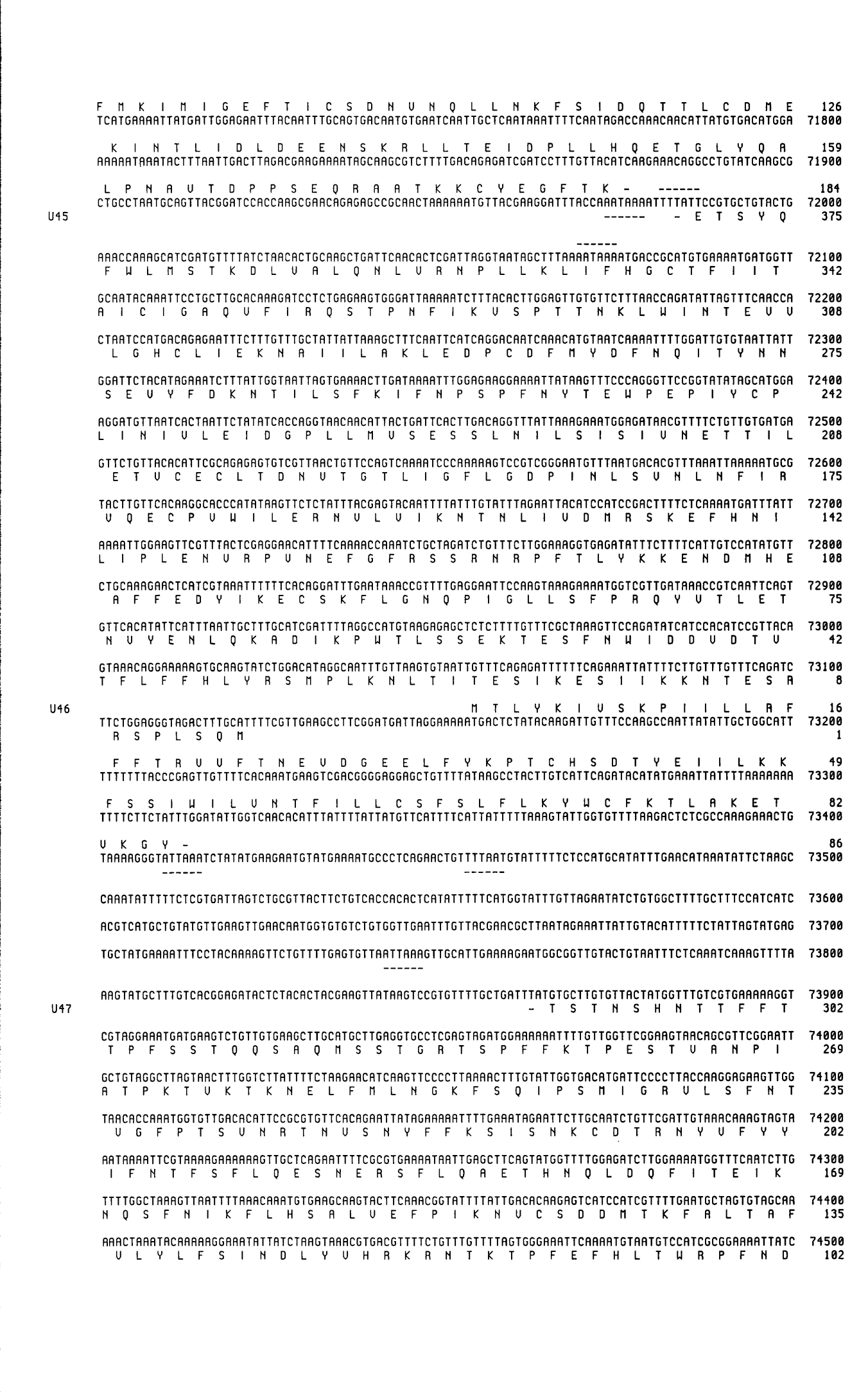
```
ATAGTGTCTAATTTATAATTCTGAGCTGAGATTTTAGATGAATACACATTATACATGTCGAAGCATAAAATCCCAGACATGTTAACTTTAGTTAAAGAAT 57400
I  T  D  L  K  Y  N  Q  A  S  I  K  S  S  Y  V  N  Y  M  D  F  C  L  I  G  S  M  N  V  K  T  L  S  N   409

TCAAAAACTTCTTATGCTGTTCGTAGGGAACGATGACAGAAAACTTTCCACGTTTTAGTTTCGAAAAACCCCCTATCTCAAAATTATATATTTTTTCCAT 57500
   L  F  K  K  H  Q  E  Y  P  V  I  V  S  F  K  G  R  K  L  K  S  F  G  G  I  E  F  N  Y  I  K  E  M   376

TCTCGTACATAAATACTTTAAATCGAAATTTATGATGTTGTATCCAGTAAGGATTTCTGGTGATTTACATTTTAAAAAAAGGAAAAAAGCATATAGAAGT 57600
   R  T  C  L  Y  K  L  D  F  N  I  I  N  Y  G  T  L  I  E  P  S  K  C  K  L  F  L  F  F  A  Y  L  L   343

TCAAATTCTGATGCAAACTCATAGATGAATACCCCTTCGATTTGCTCGCAAGTGCCCAAAGTAAACAAATGTTTACTTTGGTAATTTCCTTCAGAATCAA 57700
E  F  E  S  A  F  E  Y  I  F  V  G  E  I  Q  E  C  T  G  L  T  F  L  H  K  S  Q  Y  N  G  E  S  D  F   309

AATCGATGACCGAAATTTGAATGATGATGTCACCCATTTGCTCAGCATCTGGAAAGTTTCCATTTTGGCTTAGACATTCAATATCAAAAGAACAACAATC 57800
   D  I  V  S  I  Q  I  I  I  D  G  M  Q  E  A  D  P  F  N  G  N  Q  S  L  C  E  I  D  F  S  C  C  D   276

ATATAATGGCCAAGAATCTTCCTTTAACAAAAATAAATCTGACACATGACAATTTAATTCAATTTCAACATTACTGCTTTTTGCAAATTCTTGAATAGAT 57900
   Y  L  P  W  S  D  E  K  L  L  F  L  D  S  V  H  C  N  L  E  I  E  V  N  S  S  K  A  F  E  Q  I  S   243

AGATAATTAATTTGGTACCAACCAAAGCTTTTTAAGTTGTTATCAATGAAGAAACGATTGAGAATCTCAACTTCTGCTTCATAAACAGAAACGCCCTCAT 58000
L  Y  N  I  Q  Y  W  G  F  S  K  L  N  N  D  I  F  F  R  N  L  I  E  V  E  A  E  Y  V  S  V  G  E  N   209

TTAGTAAAATTTTTCCAATTCGATTACTGATATAAAAGTTTGAAAATGATAATTTAAATAAATTTTTGATAGGTTCCGTGTTATAACCGTAGAAATTGTA 58100
   L  L  I  K  G  I  R  N  S  I  Y  F  N  S  F  S  L  K  F  L  N  K  I  P  E  T  N  Y  G  Y  F  N  Y   176

TTTAGTTACAGATTCAATAGAAAATGAGCATGACATTTTTATTTCAGAACTATTCAGAAGTGAACAGATTCTAGAATTCAGCTCTTTTTTACATTGATAT 58200
K  T  V  S  E  I  S  F  S  C  S  M  K  I  E  S  S  N  L  L  S  C  I  R  S  N  L  E  K  K  C  Q  Y   143

TCACAGTAGAAATAACTATTTTGTCCAAATACATTGATACAAACTTTTTTACCACATTCTGTTTTTCCAAAAAGTTTTATGACATTACCAGATGGAATAA 58300
E  C  Y  F  Y  S  N  Q  G  F  V  N  I  C  V  K  K  G  C  E  T  K  G  F  L  K  I  V  N  G  S  P  I  V   109

CGAAATGTCTATATTGGAAAGGGATGTTTTCTATAGAATCAGTAAATAATAAAGATTCGGATGCATCATAGATATGAAATTTTAAAGGCGCACGCAATTC 58400
   F  H  R  Y  Q  F  P  I  N  E  I  S  D  T  F  L  L  S  E  S  A  D  Y  I  H  F  K  L  P  A  R  L  E   76

TTTTTCGGCTACTTGAGAAAGCGATGGCCATGTCATATCATTTTTGAGAATGTATTGCTTATCTTGATAGAACATTCTCGGTTCCGAGTCACAAAGAGTC 58500
K  E  A  V  Q  S  L  S  P  W  T  M  D  N  K  L  I  Y  Q  K  D  Q  Y  F  M  R  P  E  S  D  C  L  T   43

TTCATTAAGCCGGGTGCTCCATCGTGCATTATACCGCGCGGAAAAATCCGGAGAAAAGTAGATTTTGTTTTTTTCTTAGTTCGTACATTCTCTAGATACG 58600
K  M  L  G  P  A  G  D  H  M  I  G  R  P  F  I  R  L  F  T  S  K  T  K  K  K  T  R  V  N  E  L  Y  P   9

GATTAAAGAAGGACACCAGATCCATCACAGTTCTTCTGTTGAAAGTTTCTGATAGCCCCTGTATTGAATCCTTTCAAGTAGGCTTGGTCTCTGTGATTTT 58700
   N  F  F  S  V  L  D  M                                                                              1
U39                     -  L  E  E  T  S  L  K  Q  Y  G  R  Y  Q  I  R  E  L  L  S  P  R  Q  S  K   799

TTTGCCTCAGCGATTTCCGCTTTTTTATATGACTCATCTAATAATTTAATAGCTTTCAACATTTTTAAAGCTTCCACTTGTGAATATTCATTAGTGTGTG 58800
K  A  E  A  I  E  A  K  K  Y  S  E  D  L  L  K  I  A  K  L  M  K  L  A  E  V  Q  S  Y  E  N  T  H  T   765

TTATGTCATCACCTTGCGGTTCTGTGTCTTTGTTAGTTTTTAAATAAGGCGGAGATGAATCTAAAGGTTCTTTCACTTGCGCTTGAACTGATTGCGTTGC 58900
   I  D  D  G  Q  P  E  T  D  K  N  T  K  L  Y  P  P  S  S  D  L  P  E  K  V  Q  A  Q  V  S  Q  T  A   732

TTGAAGTACCGAATTTGTTGATTTTGATGTATAAGGAAATAAAATATCAATCGGTGTTTGTGCTAATCGTTTATGTCTTAAAAATAAGAAAATAACTAAG 59000
Q  L  V  S  N  T  S  K  S  T  Y  P  F  L  I  D  I  P  T  Q  A  L  R  K  H  R  L  F  L  F  I  V  L   699

CCAATCACTCCCAAAGTTAGTAAGATAGTTAGTGCTCCACCAAATGGATTTTTAAGAAAAGAGAAAACTCCATTCACCACATCTCCCAAAGCACCAGCCG 59100
G  I  V  G  L  T  L  L  I  T  L  A  G  G  F  P  N  K  L  F  S  F  V  G  N  V  V  D  G  L  A  G  A  T   665

TTACACCCAGCACAGCCCCCAAACCAAGACCAAGAGCGCCAAGACCTTTGAAAAATGTGTCAATCCCTCCTATATATGTTGGAGTAACAGTTGCAATTTT 59200
   V  G  L  V  A  G  L  G  L  G  L  A  G  L  G  K  F  F  T  D  I  G  G  I  Y  T  P  T  V  T  A  I  K   632

TGTTTCTATATGGTGTAGCGCACTTTTGTAAGAATTGTATTCTCTAAGAATGGTTTCCAAATCAAAAACGTTCGCTTTGCTCAGTTCATCTTTTGTATAC 59300
T  E  I  H  H  L  A  S  K  Y  S  N  Y  E  R  L  I  T  E  L  D  F  V  N  A  K  S  L  E  D  K  T  Y   599

AATTCCAATAATGTAAAATCTGCATTTTCTAAAGGATCTATGTTTAAGTCAACAAACGCATCTAAAGCTTCTATTTCTGTTATCAAACTAGAATTAACGT 59400
L  E  L  L  T  F  D  A  N  E  L  P  D  I  N  L  D  V  F  A  D  L  A  E  I  E  T  I  L  S  S  N  V  Y   565

ACGTATAGTCTTTAAAAATATGAGCGAATTTACCAGACAAAAAAATTTTCGTGTTAGATTGTTCACATTCTTCTGTTCTGTGATTTCCCAGTAATATCTC 59500
   T  Y  D  K  F  I  H  A  F  K  G  S  L  F  I  K  T  N  S  Q  E  C  E  E  T  R  H  N  G  L  L  I  E   532

ATTGTCTAAACCAAGTTGTCCTAAATAAGTTTCCTTCGAAGAATTTGCAAATGAATATGTTAACAATGGCCGAGAATAACATCTTAGAGCATCGTAACTT 59600
N  D  L  G  L  Q  G  L  Y  T  E  K  S  S  N  A  F  S  Y  T  L  L  P  R  S  Y  C  R  L  A  D  Y  S   499

GAATCTTTTGTCAATCTCATACTTTTATGTAGCTGAACAGAAGTCTGGTTGACTTCAATGCACTTTGAGACTGCTAATACATCACCAATTAATTTAGCGG 59700
S  D  K  T  L  R  M  S  K  H  L  Q  V  S  T  Q  N  V  E  I  C  K  S  V  A  L  V  D  G  I  L  K  A  S   465

ACATGGGTTTACCGTACACTGCTGAAATGATTCCAGAAGGACTTATTTTACTAAGCTCGTGCAACACTGTGATAGTGCGTTTTTGATCTAAACACCAGGC 59800
   M  P  K  G  Y  V  A  S  I  I  G  S  P  S  I  K  S  L  E  H  L  V  T  I  T  R  K  Q  D  L  C  W  A   432

TTCAGCAAGCTTACCTAGTGCTGTGTTAATGTAATCTTTCAGAGTGTCGTATAGGTATTGTAGTTGAACATATACGATGTCTTTGTTAGTCTCTAATTCC 59900
E  A  L  K  G  L  A  T  N  I  Y  D  K  L  T  D  Y  L  Y  Q  L  Q  V  Y  V  I  D  K  N  T  E  L  E   399

CTCTTCCTCCTTTTTCTAGAAGCATTAGAAAAATTTTCCAAAACCGTTAATTTCCGTTGTATTAACGGTTGCCAAATCAGAATTAAATCTCCCTCAGTTT 60000
R  K  R  R  K  R  S  A  N  S  F  N  E  L  V  T  L  K  R  Q  I  L  P  Q  W  I  L  I  L  D  G  E  T  K   365

TAAAAATTTCATAACTACCATTTTTGTCATGCGAATTGTTGTATTGCGTTTGGTAAATCTCATCCAAAATTGAAGTATAATTGCTGTTAATACAATTATA 60100
   F  I  E  Y  S  G  N  K  D  H  S  N  N  Y  Q  T  Q  Y  I  E  D  L  I  S  T  Y  N  S  N  I  C  N  Y   332

TTTTGAATCAGAAAGCGTATAATTACTTTTTCCTGTGACAAAAGAAGCAGTTAGTTCCTGCGCAATAAAGTGAAAACTGTTAGCATTTTCTGCCCGTAAA 60200
K  S  D  S  L  T  Y  N  S  K  G  T  V  F  S  A  T  L  E  Q  A  I  F  H  F  S  N  A  N  E  A  R  L   299
```

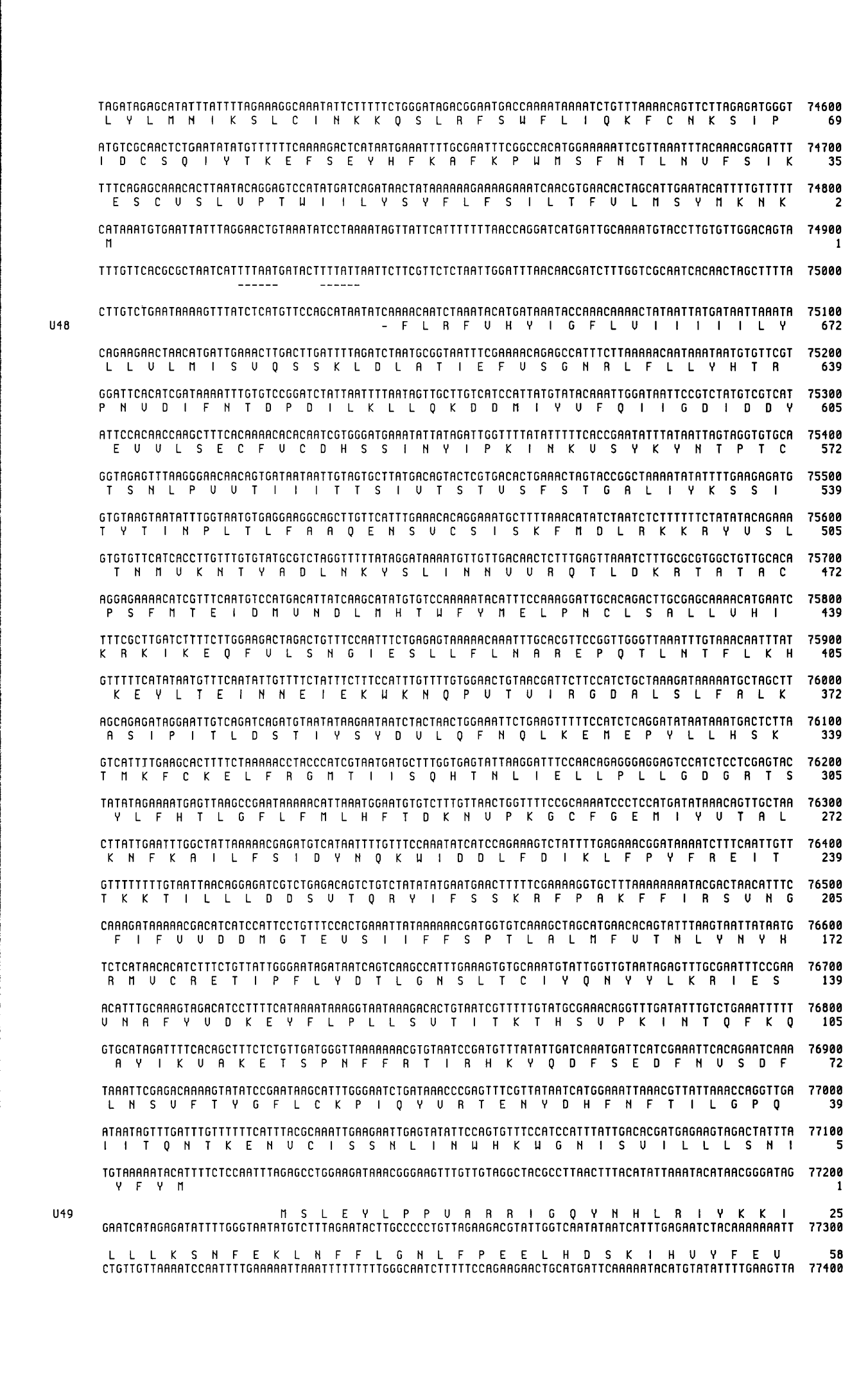
```
GCGTGTGGGATTGTCATCCAATGCTTTAACAAACAAATGGAAGACTCTTCATTCTGCACTTCCCATGAAAAAATAGTATTTCCCTTTTCCAGAAAAGCCA 60300
A  H  P  I  T  M  W  H  K  L  L  C  I  S  S  E  E  N  Q  V  E  W  S  F  I  T  N  G  K  E  L  F  A  M   265

TTTTAGTTAACGTCACAAAATTTCCTTTTGATCCATCATCAAACGTTTTCAGCATAGTATAATTATTTCTAAACAAAATTTTTTCCGTTGGTTCATTAAA 60400
  K  T  L  T  V  F  N  G  K  S  G  D  D  F  T  K  L  M  T  Y  N  N  R  F  L  I  K  E  T  P  E  N  F    232

TGTTTTAGAATTAATACCATTATAAAACGGTGACCCTTCCACGGTTTCCCCTGTTGATAAAGCAAAAAAATCAAAAGGATATTTAGTCTTTGCAATGCAG 60500
 T  K  S  N  I  G  N  Y  F  P  S  G  E  V  T  E  G  T  S  L  A  F  F  D  F  P  Y  K  T  K  A  I  C     199

TCTGTAACTATGCAATTTATCGATGTAGATGTTGAATAGAACCAGAGGGGACCATTTCTTAAAAATGGTTCTTTAGTAGTTATATATCTTCTGACAGTAT 60600
D  T  V  I  C  N  I  S  T  S  T  S  Y  F  W  L  P  G  N  R  L  F  P  E  K  T  T  I  Y  R  R  V  T  D   165

CAGATTTAAAATTTAACGGAATCAAATCCAACGTTTTATTCACATATTCATCTTTATGATATGCAACATACTCCTCTTCCTCCGAACGTTTTACAGAGAT 60700
  S  K  F  N  L  P  I  L  D  L  T  K  N  V  Y  E  D  K  H  Y  A  V  Y  E  E  E  E  S  R  K  V  S  I    132

AGAGGACAAACAACGCGCCTCGGTGTTTACCATGTGCACTTCTTCAATTGGCATTGGCAAAGTTGTAACAGTTCGATCTAAGAAATACACAGTGCCAACA 60800
 S  S  L  C  R  A  E  T  N  V  M  H  V  E  E  I  P  M  P  L  T  T  V  T  R  D  L  F  Y  V  T  G  V     99

TCGCGATATGTTGTTTGAAATGTAAGTTCTTTTTTAAAAGTTCTGACAGAAAACGTGTGTGCTTCAATCTTTGTTTTGTAAATTATCAAAATACCTTCGG 60900
D  R  Y  T  T  Q  F  T  L  E  K  K  F  T  R  V  S  F  T  H  A  E  I  K  T  K  Y  I  I  L  I  G  E  T    65

TAGTTTTAATGTTAGAGCCGTAAGACGCGCAAGAAACCTCCCTGTCAAAACGCACTAAATCTGTCCCGGTGGCAATTGAACAAATTCGAAATGGTAAATG 61000
  T  K  I  N  S  G  Y  S  A  C  S  V  E  R  D  F  R  V  L  D  T  G  T  A  I  S  C  I  R  F  P  L  H    32

CTGATTGTGTCCAGTCATGACAAAGTCAGCTTCTGTTTGTAGAGATAGCTGTAAGCTAAAAGTTATAAAAACACTCAGGAATAGAATTTTCATTTTGTCT 61100
U40                                                    -  F  N  Y  F  C  E  P  I  S  N  E  N  Q  R    708
 Q  N  H  G  T  M  V  F  D  A  E  T  Q  L  S  L  Q  L  S  F  T  I  F  V  S  L  F  L  I  K  M          1

ACATAGATCTAACAGATTTGGAGCCGGTATATGGGAAACTTGCTTAGTTTGTGAAATTTGAACAGGATCGCATTGAGTCTTTAAATCGAAATACAATAAA 61200
 C  L  D  L  L  N  P  A  P  I  H  S  V  Q  K  T  Q  S  I  Q  V  P  D  C  Q  T  K  L  D  F  Y  L  L    675

GAGTACAAATCATTAAATTTTGTCGTGTTCGTTCCACGCTTGAATAGTAAAGGAGAACCACTATCAAAAGTTAAAAGTATTTTTTCGACAGTTTCTGTCT 61300
S  Y  L  D  N  F  K  T  T  N  T  G  R  K  F  L  L  P  S  G  S  D  F  T  L  L  I  K  E  V  T  E  T  E    641

CTTCGCCAAACTTGACGATTGATAATTGTTTTCCAAATACATCATTATATAGTGCAACAGACAAAACTAATTCCCTGATAAAATTCCACATTTGTCTTTG 61400
  E  G  F  K  V  I  S  L  Q  K  G  F  V  D  N  Y  L  A  V  S  L  V  L  E  R  I  F  N  W  M  Q  R  Q    608

CATATCGTTGATATTATGCAAACCATCAAAATTGAAAAATTTATTATATGTACACAGCATCCAATTTTCCGGGTAGATCGTGCCTTCAACACATTTAGCC 61500
 M  D  N  I  N  H  L  G  D  F  N  F  F  K  N  Y  T  C  L  M  W  N  E  P  Y  I  T  G  E  V  C  K  A    575

AGATCTTCTTTCATATGTGGTAAAAAATCTGCAACATCGCAAGCATACGCCATGGATATGTTGCTGGACAATGGAAAAGAAGCGGTGTGATAATTCGACA 61600
L  D  E  K  M  H  P  L  F  D  A  V  D  C  A  Y  A  M  S  I  N  S  S  L  P  F  S  A  T  H  Y  N  S  L    541

ATGGACCAGTCAACAATCGATACATTTCTTGACTTAAATTAGGTAAAAGTTCCTGAGGCAATTTTTTTGAAAGTAGATTGTTTTTAATGTATAAGTGATC 61700
  P  G  T  L  L  R  Y  M  E  Q  S  L  N  P  L  L  E  Q  P  L  K  K  S  L  L  N  N  K  I  Y  L  H  D    508

ATCATAAGTAAATGGATCAGCAGAGAAATTTCCTAGTGACTGCAGTTGTCTATTTGCAAATGCGTTCATCATTTTCGAAAGACTTTCGAAAATGTGATTT 61800
 D  Y  T  F  P  D  A  S  F  N  G  L  S  Q  L  Q  R  N  A  F  A  N  M  M  K  S  L  S  E  F  I  H  N    475

CCTAGGATATTAATATTCATTAACTTATTTATTAATGCTTCTTGCTCTTTGATACACAGCATTACTTTACTATAACCCGATTCGGAAATTTTTTGGTAAT 61900
G  L  I  N  I  N  M  L  K  N  I  L  A  E  Q  E  K  I  C  L  M  V  K  S  Y  G  S  E  S  I  K  Q  Y  Y    441

AAGCTTTTTTTCTAATGTGAATCTCTTGTTGTAAATTGTGATCTTGTGTCATTGAATGCGTTACGTATTTTTTTGGGACAGTTTCATGAGCTGAATTTTC 62000
  A  K  K  R  I  H  I  E  Q  Q  L  N  H  D  Q  T  M  S  H  T  V  Y  K  K  P  V  T  E  H  A  S  N  E    408

AAGTTCATTTAATAAAGTTTTTATTTTAGTTTGTATTTCATCTTCCTCTGCCAGCTTTGTAAAAACCGGATCTTCTTCTAAGTTTTTTATACTAACATTG 62100
 L  E  N  L  L  T  K  I  K  T  Q  I  E  D  E  E  A  L  K  T  F  V  P  D  E  E  L  N  K  I  S  V  N    375

GTAATAATATCAACGAGTCTGCAAGGGGAAATAAATCTCGATCCGATGTATAATTTGTCATCATTTGTAACATTTTCTTCGTTAAAATTGAAAACATCTT 62200
T  I  I  D  V  L  R  C  P  S  I  F  R  S  G  I  Y  L  K  D  D  N  T  V  N  E  E  N  F  N  F  V  D  Q    341

GAATAGTTTCACCATATAAAAACTTACTAATTTCATTTTGGAGATCAGCATAAAGTGAATGACAAAGATTTAGTTGATTTAGAGTTATTGAAGCTTTTTG 62300
  I  T  E  G  Y  L  F  K  S  I  E  N  Q  L  D  A  Y  L  S  H  C  L  N  L  Q  N  L  T  I  S  A  K  Q    308

AATAACAGATTCTGAAATTTTAGACCAATAGGTGAATTCACTCAACGAATAAATTGTATCTGGAATTTTGCTAAAAAGGTCAAAGTTGATTAAAGTGTTT 62400
 I  V  S  E  S  I  K  S  W  Y  T  F  E  S  L  S  Y  I  T  D  P  I  K  S  F  L  D  F  N  I  L  T  N    275

TCAGCTTCTTGTAAATGTGGAGGTTTTGAATATAGTTGCTGTTGAAAAATATTTGTGATTTTCGCCATATACTCTTTTACAAGACCATATTGTTTTGTGG 62500
E  A  E  Q  L  H  P  P  K  S  Y  L  Q  Q  Q  F  I  N  T  I  K  A  M  Y  E  K  V  L  G  Y  Q  K  T  S    241

AAAAATCCAAATCTTTCTCAATGATGTCAACGTTTTTTTCTAAATTGGATACCATATGTGTTTCAGTTAGATGATTACAAAATTTGCCAGCTAGTCTTTT 62600
  F  D  L  D  K  E  I  I  D  V  N  K  E  L  N  S  V  M  H  T  E  T  L  H  N  C  F  K  G  A  L  R  K    208

ACGAATAGACTTTCCTTGATTGGGCACTAAAGATAATTCTTCATAACATTTTAAACATGTAGTAGTTTCAGAAAAAACTTCTGGCTTCATCACTACGCAA 62700
 R  I  S  K  G  Q  N  P  V  L  S  L  E  E  Y  C  K  L  C  T  T  T  E  S  F  V  E  P  K  M  V  V  C    175

ACGCCACACAAAATTGTTAGAAATTCAATAATTTGACTGCAAGCATTAATGCCATCCAGTGGTGAAATTAAGCCAAAAATACAATTCATCTTACATAAAA 62800
V  G  C  L  I  T  L  F  E  I  I  Q  S  C  A  N  I  G  D  L  P  S  I  L  G  F  I  C  N  M  K  C  L  L   141

GTTTTTCAATATCATTGATCGTAGAAATGTCTACAGAAAAAGTCAAATAATTATTAAGAGCCAAGTCCAATTTTAAGTGTTGCTTGCATTCATTGAGTTG 62900
  K  E  I  D  N  I  T  S  I  D  V  S  F  T  L  Y  N  N  L  A  L  D  L  K  L  H  Q  K  C  E  N  L  Q    108

GAATATATTTTCAAAATATTTTTGTTTGTCATTATCTAATAGAAATTGATGAATTGAAGCATCGAGTAATAAGACTTGATTATACATAGCTTTTAATAAA 63000
 F  I  N  E  F  Y  K  Q  K  D  N  D  L  L  F  Q  H  I  S  A  D  L  L  L  V  Q  N  Y  M  A  K  L  L     75

ATCTGTAAATAAATGGTTACAGGAGATGCACAATTAAAATTTTGATGACAGAGTTCAGCAAATAATGTCTTAAACAAATGAATTATTACAACACCGTTTT 63100
I  Q  L  Y  I  T  V  P  S  A  C  N  F  N  Q  H  C  L  E  A  F  L  T  K  F  L  H  I  I  V  V  G  N  K    41
```

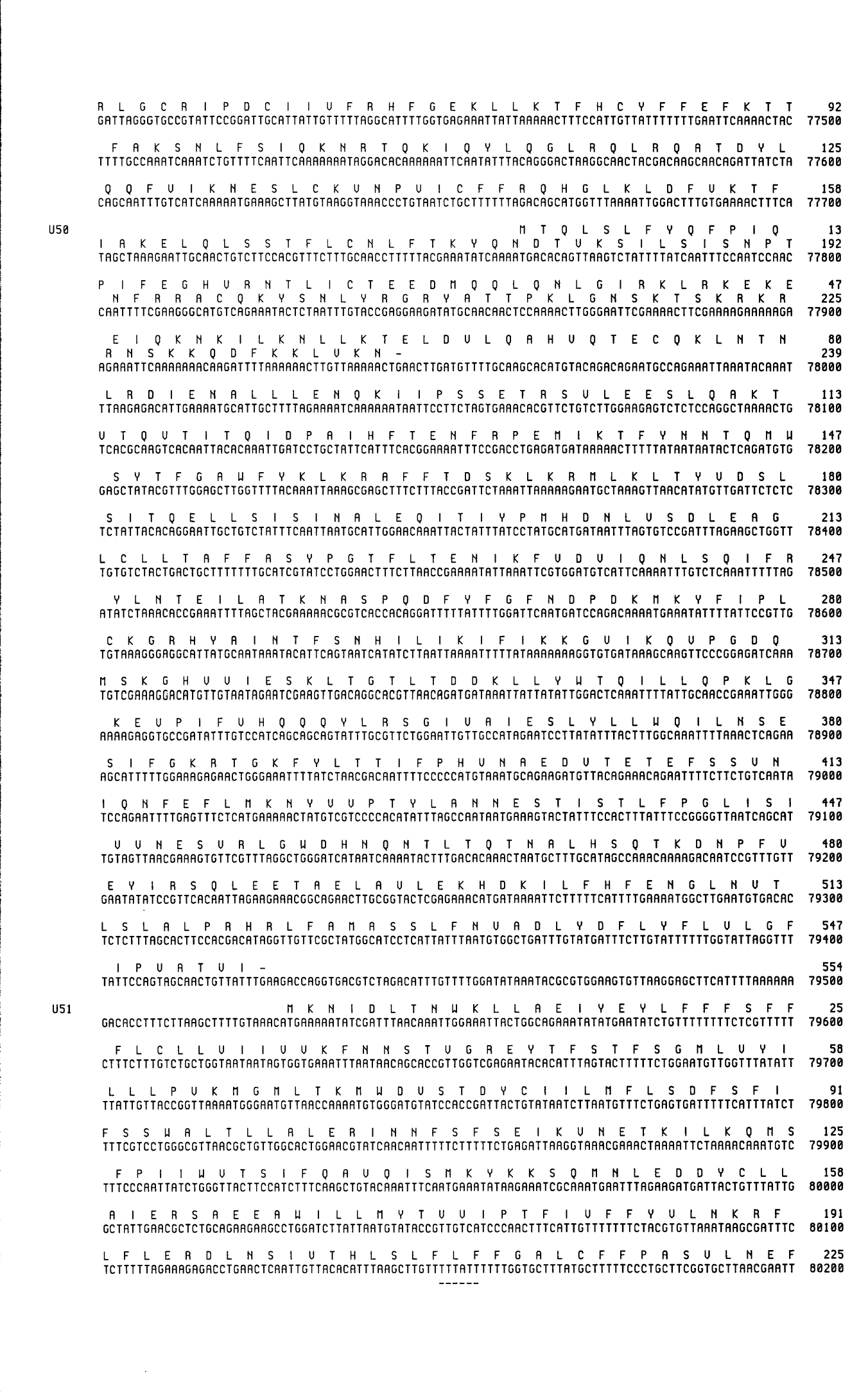


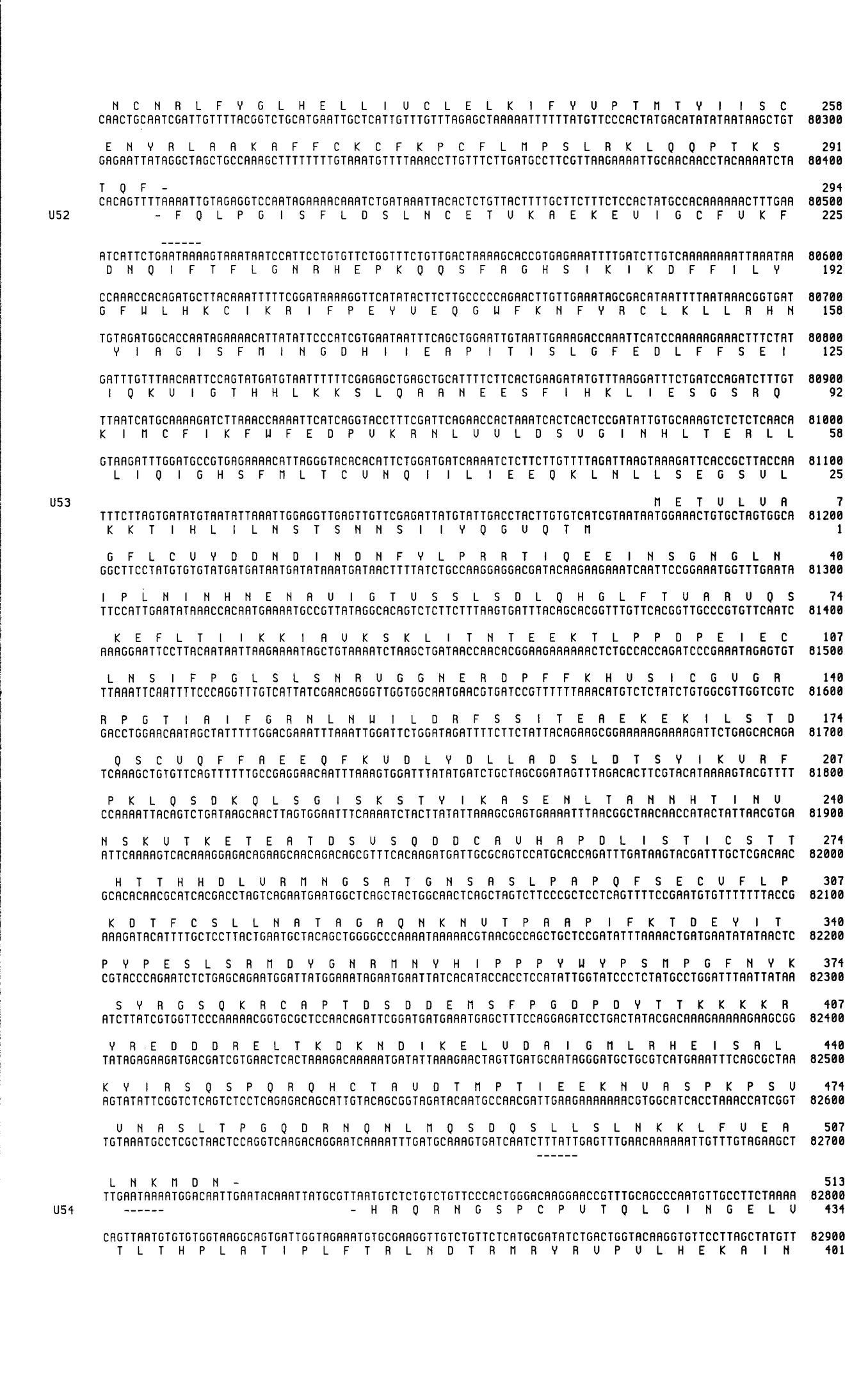

< start of fragment
containing origin of DNA replication
end of fragment containing origin of DNA repliation >
U42
U43

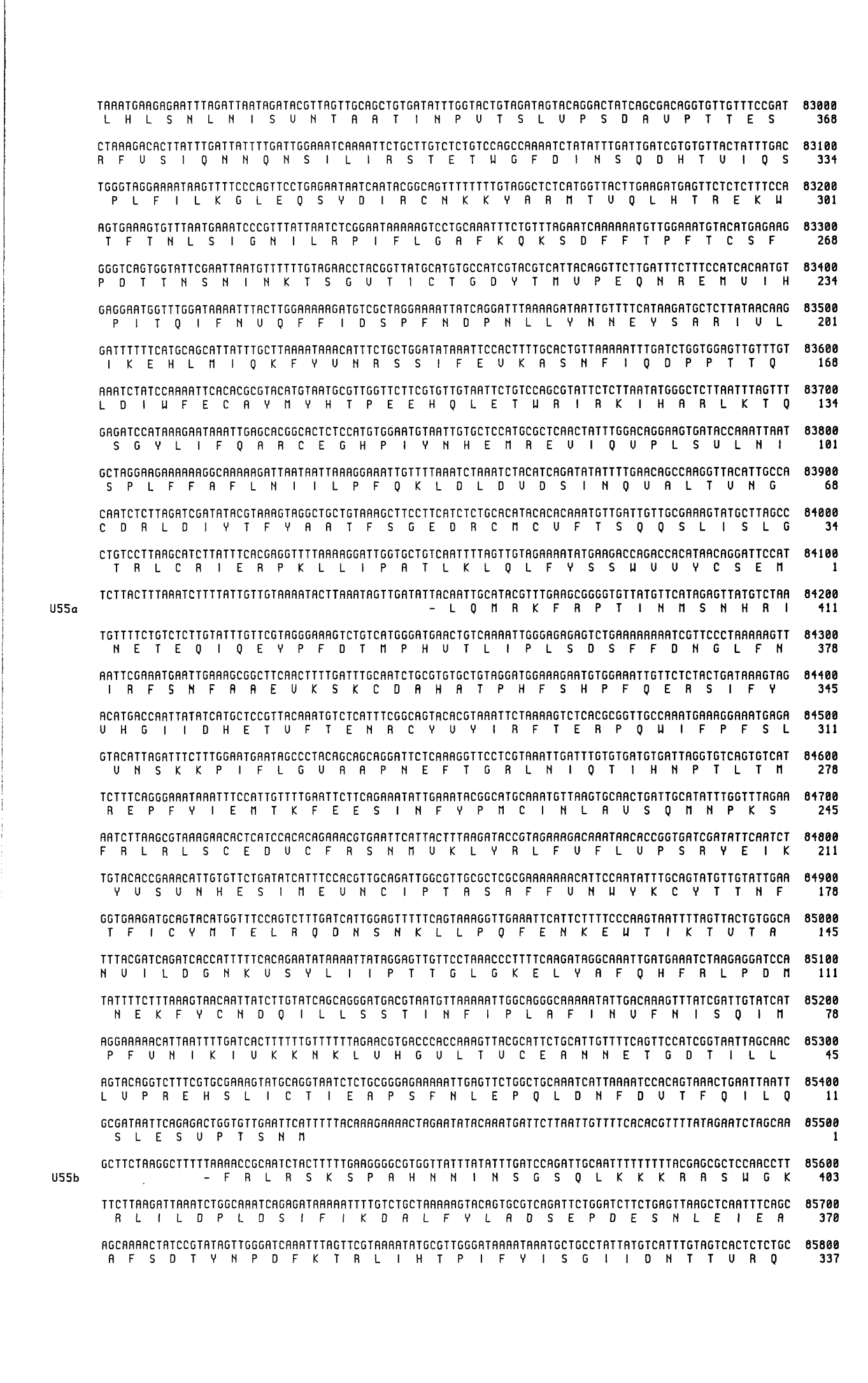
```
TTGCTCCATAATATTATCAAAATGCGATTGTTATCTACCTTAAATTCTAAGTACACTAAGACATTTTCACGGTTTCCCTTATGATTCCTTGTTAAACAAC  68900
K  S  W  L  I  I  L  I  R  N  N  D  V  K  F  E  L  Y  V  L  V  N  E  R  N  G  K  H  N  R  T  L  C  S    804

TGAAATCACTTAGTTTTCCGTTTTGCAATTTTTTTAGTTGTATATTTTTGTGATCAATTTTTAGAAAAGTTACATGTTGAAATTGTGTAGCTGTTCGTGC  69000
   F  D  S  L  K  G  N  Q  L  K  K  L  Q  I  N  K  H  D  I  K  L  F  T  V  H  Q  F  Q  T  A  T  R  A    771

TTCACAGTGTTGGATGTTCTTTTTTAACATTTGTGGCCAGGCAATTGAACGAGTAAAAGTCATTAACAAATCATCACGGGAAAGACAACTTGTTTCCGAA  69100
E  C  H  Q  I  N  K  K  L  M  Q  P  W  A  I  S  R  T  F  T  M  L  L  D  D  R  S  L  C  S  T  E  S    738

AAACCATTGACATGTAAATATTTCAAACATAAAGTCTCCAGTGAACTTTGTTTTCTGTGATACAGTTGCTTAGTCCTAGAATCTATAAAAGAAAACTCTG  69200
F  G  N  V  H  L  Y  K  L  C  L  T  E  L  S  S  Q  K  R  H  Y  L  Q  K  T  R  S  D  I  F  S  F  E  T    704

TTGGACAGGCCTTATCAAAAATGCTGTAGATTATCTCTGTAATGGTTTGGTCTTTAGTTCCATGATGTAAAATATTATTGAGATCTAGTTGCATGCTAAC  69300
   P  C  A  K  D  F  I  S  Y  I  I  E  T  I  T  Q  D  K  T  G  H  H  L  I  N  N  L  D  L  Q  M  S  V    671

GAAATCTTGAAAATTCTTTTTCATTTTTTCCGGGACAATAAAAATAGGCTTCAGACGACCATACAAAAATCTGTTATTCTCTTCGTCAACTTTATACATG  69400
F  D  Q  F  N  K  K  M  K  E  P  V  I  F  I  P  K  L  R  G  Y  L  F  R  N  N  E  E  D  V  K  Y  M    638

AAACCTAACCGTAAACAACGTCCCGTATTATAGATTCCTGTATCAAAACATTCACCAGGAAAAATTACCGAGTTAAGAATCTGCATTAAATCGTGGTTTA  69500
F  G  L  R  L  C  R  G  T  N  Y  I  G  T  D  F  C  E  G  P  F  I  V  S  N  L  I  Q  M  L  D  H  N  L    604

GACACATTAAATGATTGAATATTTTAGAAAGTTGTTTTAATGGCTCACTTCCAATTATTGCAGTACCTTGAGGGATTGGAATAGCAATTCGAAGACCAAT  69600
   C  M  L  H  N  F  I  K  S  L  Q  K  L  P  E  S  G  I  I  A  T  G  Q  P  I  P  I  A  I  R  L  G  I    571

TTTTTTTTTGCAAACACAAAAAGCTGTTTCAACTTCATTGTAACATACATTGTGGTTCAGGGGATTCTCCGGATTACTACAGGTAGTTTTGAAAAAAAAG  69700
K  K  K  C  V  C  F  A  T  E  V  E  N  Y  C  V  N  H  N  L  P  N  E  P  N  S  C  T  T  K  F  F  F    538

ATGGGGTAGGCGTCTTTGTCAATAAACGGAAATATTTTTTGCCATGCATTGATCAGCACATTCCGCATCAATTTACACATAGAGAAAAATAACGATTCCG  69800
I  P  Y  A  D  K  D  I  F  P  F  I  K  Q  W  A  N  I  L  V  N  R  M  L  K  C  M  S  F  F  L  S  E  T    504

TTATTATAGCTGAATCCTGTAATGGCAGATCTAAATCACCAATATAATTTGTAACCGGAAGGTATTCATTAAAGATTTCATGCCTGGTTACGTAAAACTG  69900
   I  I  A  S  D  Q  L  P  L  D  L  D  G  I  Y  N  T  V  P  L  Y  E  N  F  I  E  H  R  T  V  Y  F  Q    471

TTTTTCTATTTCGTCGCTTAAAAAGGTCGTGTCCTGCAACCAAAATGTACTTGTTAAAGATTCATCCGAAATGTACTCATTGGGTAAAAATTGCATAACC  70000
K  E  I  E  D  S  L  F  T  T  D  Q  L  W  F  T  S  T  L  S  E  D  S  I  Y  E  N  P  L  F  Q  M  V    438

TGGTCGAAACCCATATTTTTGTACCAACTTTCTGCACTTCCAAAACAATAAAGATTTTTGTAAGACATTTGAACTCTAAACATAGGAATGGGGTTGCTGA  70100
Q  D  F  G  M  N  K  Y  W  S  E  A  S  G  F  C  Y  L  N  K  Y  S  M  Q  V  R  F  M  P  I  P  N  S  F    404

ATTGGTCATGTAGATATTGTCTCCTTGTAACAGGTTTACCATCTATCAAAATAGATTTATTAGATGCCAAAATTTTTAAAAATCCTTGTAAGCCTAATTT  70200
   Q  D  H  L  Y  Q  R  R  T  V  P  K  G  D  I  L  I  S  K  N  S  A  L  I  K  L  F  G  Q  L  G  L  K    371

AATTGGTGTAAAATACCCTTCACCGTGTTTATTGCATTTGTTAACCATCTTTGAAATTTGGTCTGAATTTATTACAAGCATGATATTTCTATTACTGTTT  70300
I  P  T  F  Y  G  E  G  H  K  N  C  K  N  V  M  K  S  I  Q  D  S  N  I  V  L  M  I  N  R  N  S  N    338

ATGTCATAAACGTACTGTTGCATTTGAAATTTGTTGTTTAACATCCTTTTTATTTCTACATAATTTTGAAAATAACTGTCTTCGTTAAAGTACTGTTTCA  70400
I  D  Y  V  Y  Q  Q  M  Q  F  K  N  N  L  M  R  K  I  E  V  Y  N  Q  F  Y  S  D  E  N  F  Y  Q  K  M    304

TGATTGAAATGAGATCCGAGTGTAAAGAATTTCCTAGGATGTACATATTTTCCTGACATTGAGAAGTATCAGATCTGGTAGGATCATAAGATGTCAGTTG  70500
   I  S  I  L  D  S  H  L  S  N  G  L  I  Y  M  N  E  Q  C  Q  S  T  D  S  R  T  P  D  Y  S  T  L  Q    271

ACTATACTTGATGAATTTATTTTTTTCTAATGCTGTACTGTAAGCAATGTAGATATAGTGGACTCGAAAACTGTTTGTGAGATGCACGCTGTTTCTAAAA  70600
S  Y  K  I  F  K  N  K  E  L  A  T  S  Y  A  I  Y  I  Y  H  V  R  F  S  N  T  L  H  V  S  N  R  F    238

ATATCTATTTGCTTATCAAGATAAAGAAATTCTTTTCGATCTCTTATGTTTTTTTGTTTTGCATATTCAGAAAAACTAACAAATTTTTTCCGATAGTTAG  70700
I  D  I  Q  K  D  L  Y  L  F  E  K  R  D  R  I  N  K  Q  K  A  Y  E  S  F  S  V  F  K  K  R  Y  N  S    204

ATTTCAAAGTAATATTAATGAATTCAGACATGGAGGAATAAGTAAATACCATATCTCCTAATTTATCAGTGAAACACTGAACATAATTCTTAGTTGTTGC  70800
   K  L  T  I  N  I  F  E  S  M  S  S  Y  T  F  V  M  D  G  L  K  D  T  F  C  Q  V  Y  N  K  T  T  A    171

TATGGTGCTTTTCTGCTCTTCAAATAAATAATAATACATCGTTAATAGTAGCATTCCTTCTGTTTGACCGTATGTTGATAAAAACCAAAACGGTGACGTG  70900
I  T  S  K  Q  E  E  F  L  Y  Y  Y  M  T  L  L  L  M  G  E  T  Q  G  Y  T  S  L  F  W  F  P  S  T    138

GGTAAGAGATACTTATTTGTTAGATGTTTAATGATTATAGTCTTGCAAAAAAAGACTAAGCTTTTAAATTCTATATTTTCACTTCCTTCTGGAACATGAG  71000
P  L  L  Y  K  N  T  L  H  K  I  I  I  T  K  C  F  F  V  L  S  K  F  E  I  N  E  S  G  E  P  V  H  S    104

AAAGTAAAAAATCTTTATACAGATTTTCTGTTTCTTCCGTATTTAAACAGTGAATAATATCCTCAGACTTAATGGGTAATGCATTTTTTACATAATAGCT  71100
   L  L  F  D  K  Y  L  N  E  T  E  E  T  N  L  C  H  I  I  D  E  S  K  I  P  L  A  N  K  V  Y  Y  S    71

GAAATTCAAAGATTCTGCGTCGCAAACAAATACCGTCTCAATTCTTTTGGAAAACTTGCATTTCTGTGTCTGCATACAAAAATATACATTTTTTGATGGT  71200
F  N  L  S  E  A  D  C  V  F  V  T  E  I  R  K  S  F  K  C  K  Q  T  Q  M  C  F  Y  V  N  K  S  P    38

TTATATTTAACGATTATTGGAAAAATTAAATGTTCACTAGGGTTCTTACATAAGATATTTACTATCACGTGCGCGGGGTCATATTCCGTTGCAAACACAG  71300
K  Y  K  V  I  I  P  F  I  L  H  E  S  P  N  K  C  L  I  N  V  I  V  H  A  P  D  Y  E  T  A  F  V  T    4

TAATAGTCATTGTGAATGTCTTTCAGTTTATCTTCCAACAGCTTTTAATATAAGACAATAGGAATCATGGGAAATTGCTTTGTAAAAAAAATCTCTTCTG  71400
   I  T  M                                                                                              1

U44                                  M  L  Y  S  E  L  S  E  Q  E  D  L  L  D  F  L  E  T  K  Y  T  D  F  G  I  L    26
ACATCTTCAACTCTAACTACCAGATGCTCTATTCTGAGCTCTCAGAGCAAGAGGATTTACTGGATTTTTTAGAGACAAAATATACAGATTTTGGAATTTT  71500

K  T  D  I  L  N  Y  E  R  D  S  E  T  F  K  T  L  L  Q  V  L  P  I  Y  K  K  T  K  L  R  Y  N  L    59
AAAAACCGATATTCTCAACTATGAAAGAGACTCCGAAACATTCAAAACTTTGTTGCAAGTGTTACCTATCTATAAAAAAAACAAAGCTGAGGTATAATTTG  71600

I  E  R  C  L  N  N  C  P  P  H  V  K  D  A  L  I  I  E  I  M  K  A  K  K  I  L  E  T  L  D  V  V    92
ATTGAACGCTGTTTAAATAATTGTCCCCCTCACGTTAAAGATGCATTGATTATTGAAATCATGAAAGCTAAGAAAATTTTAGAGACTCTGGATGTTGTGT  71700
```

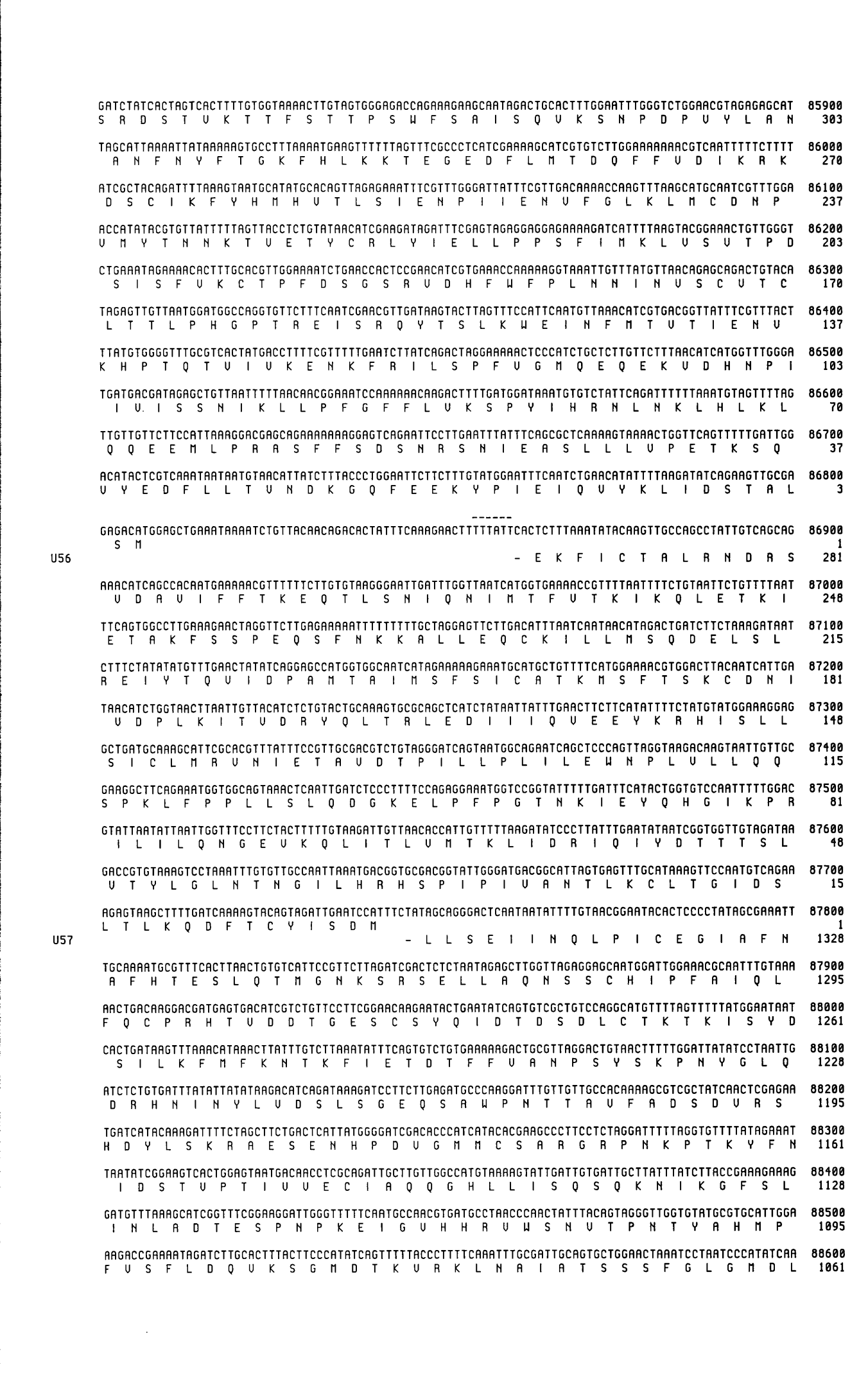






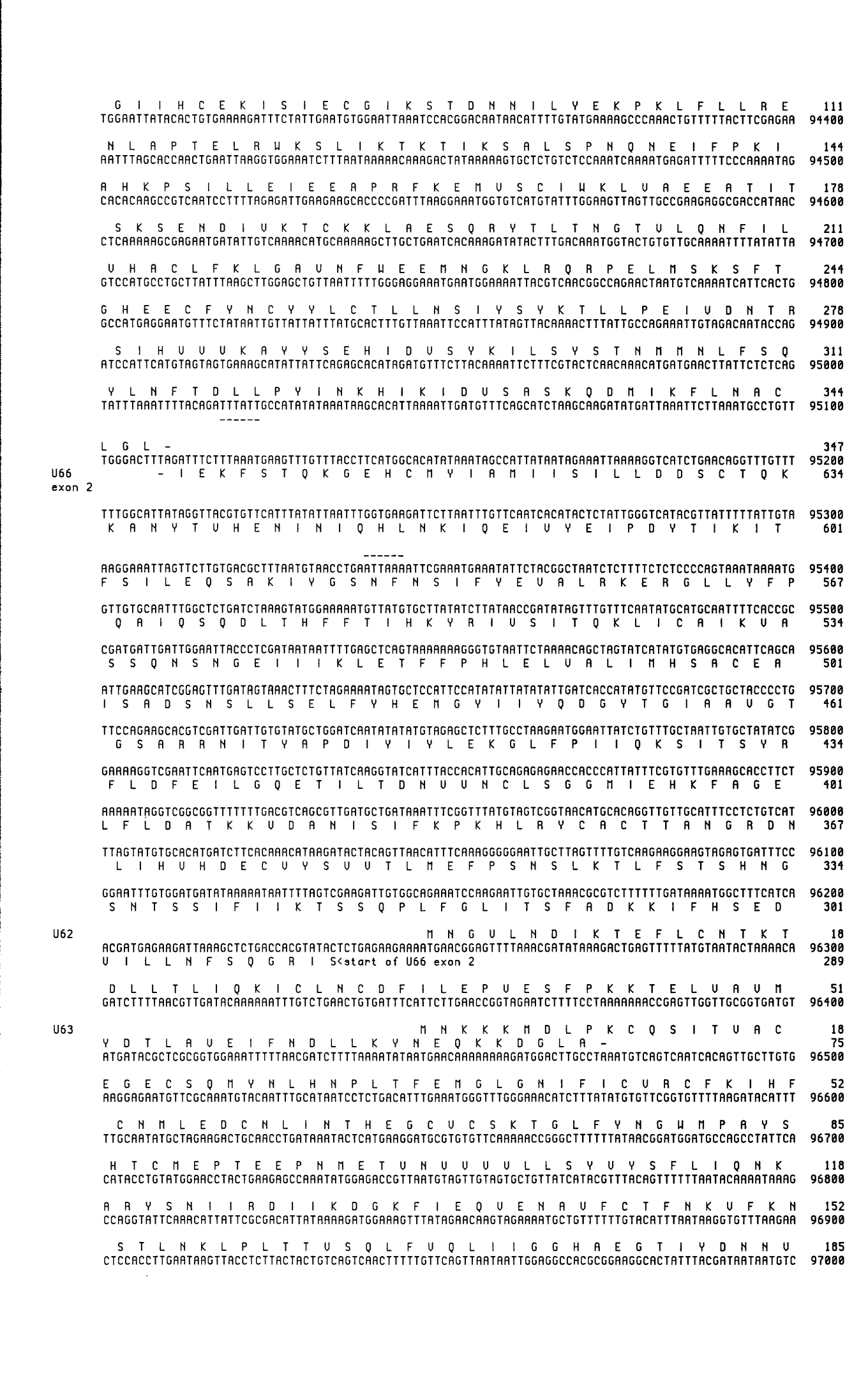


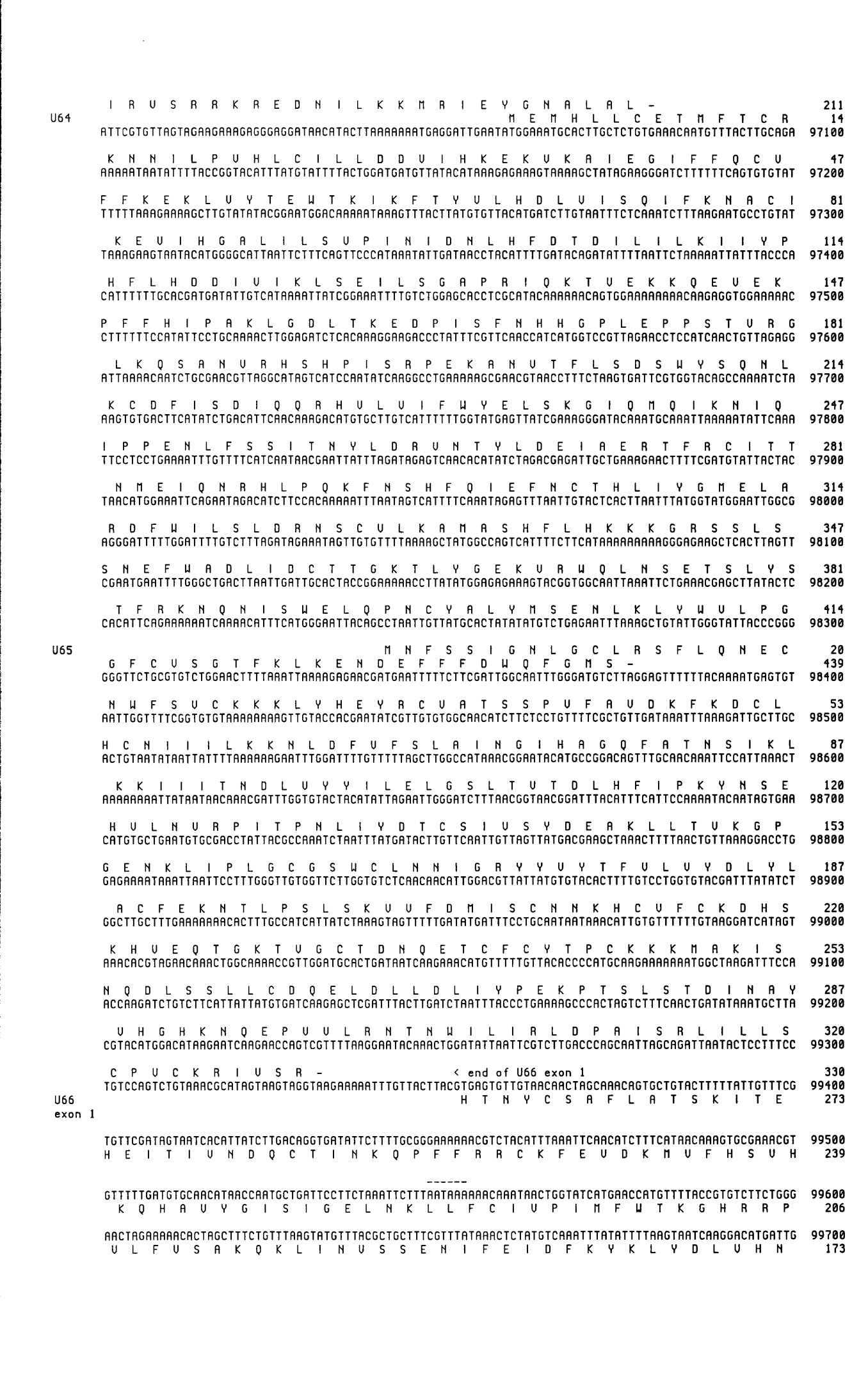




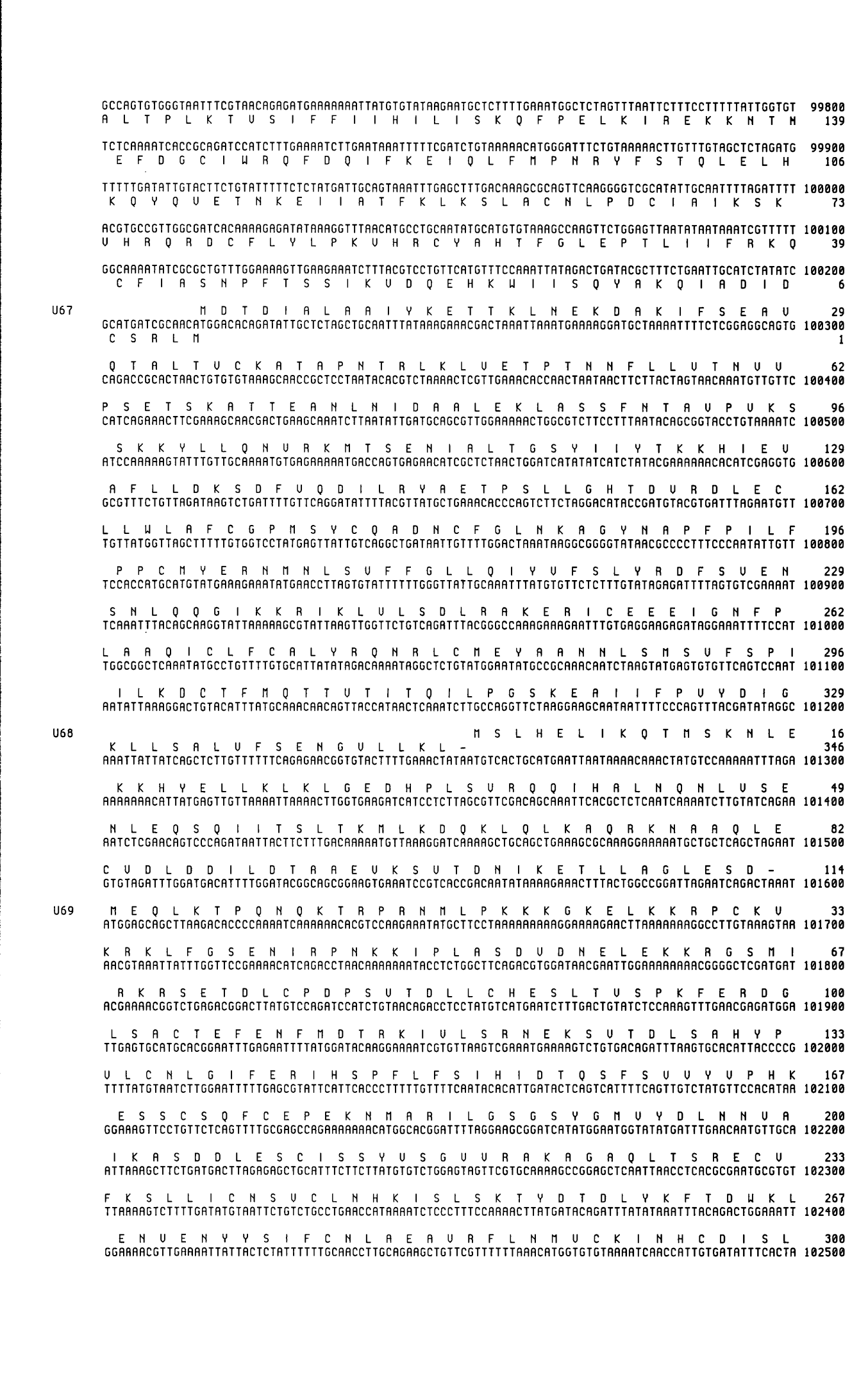
```
GAAAACTTATATGCTGCGTAAGGTTATATGTGGTTGAGATGTCTTTTACTTCTGTAGATACCGTTGGGTCATCAATCATTATGGATGTTGCAGATTTGGA  88700
  F  S  I  H  Q  T  L  N  Y  T  T  S  I  D  K  V  E  T  S  V  T  P  D  D  I  M  I  S  T  A  S  K  S   1028

GCTATACAGTAAGCAATTAATGTCGAAACAATCTGTTCTTACTAAAGTCGCAGCGAAACCAGGATGAATTTTTTGTCTGCTTTGAAGAATGATAGCTATT  88800
  S  Y  L  L  C  N  I  D  F  C  D  T  R  V  L  T  A  A  F  G  P  H  I  K  Q  R  S  Q  L  I  I  A  I   995

GGAGACAATTTGCAATGCATAGTAGCTAAAGTCATGATACTTAGCAATGAATTTGAACAACAAGAAGCATAGACAAAGAACGGTGTTCTTTGCCAATTAT  88900
  P  S  L  K  C  H  M  T  A  L  T  M  I  S  L  L  S  N  S  C  C  S  A  Y  V  F  F  P  T  R  Q  W  N  H   961

GGAATTCATGAGATAATGCTTGTGGTAGAGGAAAACCACCGTCATTTCTTTGATAGTGGGGAAATATGTTTAGATAGGTTTGTACTTCAATATTCATTAA  89000
  F  E  H  S  L  A  Q  P  L  P  F  G  G  D  N  R  Q  Y  H  P  F  I  N  L  Y  T  Q  V  E  I  N  M  L   928

GCCACAAATAATAGGGTCTGAGAAAAAACGATTAAAAGGAATTGGCTTGAAATACCGTTTCAACACATTAATTGGTGATACTAAACACAGTCCATTATAT  89100
  G  C  I  I  P  D  S  F  F  R  N  F  P  I  P  K  F  Y  R  K  L  V  N  I  P  S  V  L  C  L  G  N  Y   895

AAAACATGTTGTAGAGATTTGAAGTTTGAAGAACTACCATGTCTTTGAGAAGGATCTAAAGAAGATTCAATTTCCAAAATTTTTGTATTCTCCGTTAGGT  89200
  L  V  H  Q  L  S  K  F  N  S  S  S  G  H  R  Q  S  P  D  L  S  S  E  I  E  L  I  K  T  N  E  T  L  Y   861

ATAATATAGATTTAAAGAGTTGTTTACCAATGCAAGATAAATCTGTGAGTGCATGTGAGGGACCAACAGCTTCAGTTATCAATTCAATCAATACATTGTT  89300
  L  I  S  K  F  L  Q  K  G  I  C  S  L  D  T  L  A  H  S  P  G  V  A  E  T  I  L  E  I  L  V  N  N   828

CGTAATCGGTAATTCGCAAAAGTTGTCATCAGGTAATATAAAGGGTTCTGTGTAGAAGAAATCTAGAATGAAAGATTTCACATCAAAACCTGCACCACAC  89400
  T  I  P  L  E  C  F  N  D  D  P  L  I  F  P  E  T  Y  F  F  D  L  I  F  S  K  V  D  F  G  A  G  C   795

ATCTTGTTATTTGTCAGTGCTGGTAGAAAGCAGAAATAAAAAATTTTGCTTAAAACAATCGTTTCAAAAGATGGTCTATCTACGTCTGTGAATATTTCTG  89500
  M  K  N  N  T  L  A  P  L  F  C  F  Y  F  I  K  S  L  V  I  T  E  F  S  P  R  D  V  D  T  F  I  E  T   761

TTGAATCCAATAAATTCATTCTATTCAGGTCAGATACTTCATAGTTTCTTAATTTCACTGTATTTTGGGTAAGGGGAGTGTTACCAGCTGTGATTATTGC  89600
  S  D  L  L  N  M  R  N  L  D  S  V  E  Y  N  R  L  K  V  T  N  Q  T  L  P  T  N  G  A  T  I  I  A   728

ATTTGCTTCATTTTTTGGAAGACTGTTTAGAAATGGAGGGAATAGACGATTATCAAACAAAGCGTTCACAAATCCAACCAATGGTTCCCCACACAGTTGC  89700
  N  A  E  N  K  P  L  S  N  L  F  P  P  F  L  R  N  D  F  L  A  N  V  F  G  V  L  P  E  G  C  L  Q   695

TCATTAAAGTTTGAAATAGAAATTGTCCTTTTCACTAATTTTAAGATTGAAATTATATTTTTGTAATGTGTATAAGCAATTCCTGGAATTGACTCATCTC  89800
  E  N  F  N  S  I  S  I  T  R  K  V  L  K  L  I  S  I  I  N  K  Y  H  T  Y  A  I  G  P  I  S  E  D  G   661

CAAGATAGGTTGTAATTAACCAAATCATTTCAAAATTATTGCAAAACAATAATATGTGCTTTATATTATACCAATACGTAATACATTGGCTGACAACGTC  89900
  L  Y  T  T  I  L  W  I  M  E  F  N  N  C  F  L  L  I  H  K  I  N  Y  W  Y  T  I  C  Q  S  V  V  D   628

TTTAAGAATCTGAAACGCTGTCTTATTACCATGGATGAGAAGTTCTATGATATATGCTAGTTCGGGATAAGCTGTGTTAGTTAAACTTTCAACGACGAGT  90000
  K  L  I  Q  F  A  T  K  N  G  H  I  L  L  E  I  I  Y  A  L  E  P  Y  A  T  N  T  L  S  E  V  V  L   595

TTCAGGGTGTAGTCATAATTTAAATTATTTGTCTTGGCTTGTTCGATCATTTGATTGGTCCTAGCTTCATGAAAAGATGAAGGTGCTAACGGCAGAGGTA  90100
  K  L  T  Y  D  Y  N  L  N  N  T  K  A  Q  E  I  M  Q  N  T  R  A  E  H  F  S  S  P  A  L  P  L  P  I   561

TATTGCCTAACAAAATTCTTGGTGAACACAAAACATCAGTTGATCTGTTTTTTTGAATATAGGTGAAATCAAAAAATGGATGTAGCTCTGTCTTTAAAGT  90200
  N  G  L  L  I  R  P  S  C  L  V  D  T  S  R  N  K  Q  I  Y  T  F  D  F  F  P  H  L  E  T  K  L  T   528

AAAATTTTCAGATTTATAGAAATCTTCTGTGGTAAGTTCATTTTTAAGCATATTCGTCGTCTTAGGTATTTCTTTTTTCATTTCATATAAGTTGTATAAA  90300
  F  N  E  S  K  Y  F  D  E  T  T  L  E  N  K  L  M  N  T  T  K  P  I  E  K  K  M  E  Y  L  N  Y  L   495

TTTGATAAAAAGGTGCTTGGAGGTTCTCTAACAAATTCAACCTGACACAGTTGGTTAAAAGATTCCCCCGTAGGTTTTGCAGCATTTTTAAAAATTCGCT  90400
  N  S  L  F  T  S  P  P  E  R  V  F  E  V  Q  C  L  Q  N  F  S  E  G  T  P  K  A  A  N  K  F  I  R  E   461

CTGAGACACTACAATCAAAAATAATTGGATGGCACAATGATGGCAATAGATCACTATAGTCTATTCTCTGCAAAAGCTTATCCTTATTATAGAAATAGAT  90500
  S  V  S  C  D  F  I  I  P  H  C  L  S  P  L  L  D  S  Y  D  I  R  Q  L  L  K  D  K  N  Y  F  Y  I   428

AGATGTGGGTAGATTGTTTTCCATTGTATCATTTAGTTTAAGCCTGCTATCCATAGTGCTAAAACCACTATCCTTCGAAATGTATAAGCCCATTGGGAAA  90600
  S  T  P  L  N  N  E  M  T  D  N  L  K  L  R  S  D  M  T  S  F  G  S  D  K  S  I  Y  L  G  M  P  F   395

AAAAACGTCAATTCCAATTTTTGTTCCAAAGGATCCTCAATGTTAGTGTTTTTATACACTTTCTTTAAATGATCTAAGACAACTGTTTTATCACTCAGTT  90700
  F  F  T  L  E  L  K  Q  E  L  P  D  E  I  N  T  N  K  Y  V  K  K  L  H  D  L  V  V  T  K  D  S  L  Q   361

GAATAATGTTTGTCTTTAGATCAGCATGTTGAGTTTGGTTTTCAAAGATTTCAGATTTTTTGTTTTGGTCTTGTGTTTCAGTCGCAACGTTTTTGGTATA  90800
  I  I  N  T  K  L  D  A  H  Q  T  Q  N  E  F  I  E  S  K  K  N  Q  D  Q  T  E  T  A  V  N  K  T  Y   328

ATTTGAAAAATCCGCCATAATAGCCTGGTATGCTATTGCAGTAACAGCATTTTCTTTCCCCATAACAAATGTGCCATAAGAAACAGGTACAGACATTGTC  90900
  N  S  F  D  A  M  I  A  Q  Y  A  I  A  T  V  A  N  E  K  G  M  V  F  T  G  Y  S  V  P  V  S  M  T   295

ATTTGTGAAATGTGTTGCGATAGGGCATTGGACAAAATCTGAATCGTATTAGGAGTTCCGAGTAAAACGCCGTTAATAGGTGTTCCATCTTTTAAAACAT  91000
  M  Q  S  I  H  Q  S  L  A  N  S  L  I  Q  I  T  N  P  T  G  L  L  V  G  N  I  P  T  G  D  K  L  V  Y   261

AGCTATTGGTGTTATTTAAAATGCTTTCAGTGGTTGAATCAACCATTTCACACAGATATCTGTATATAAAGCTTACATTAGAAGTTCTGTTCAAAAAAAA  91100
  S  N  T  N  N  L  I  S  E  T  T  S  D  V  M  E  C  L  Y  R  Y  I  F  S  V  N  S  T  R  N  L  F  F   228

TAAGGAATTGATGAGTTTGTTTTTAAAGTTTTGAAACATATTGCTTCGCTGTACTTTAGAAAGAATCTGTTTACTGTTGACTTTGTTTTCTAACATGGTT  91200
  L  S  N  I  L  K  N  K  F  N  Q  F  M  N  S  R  Q  V  K  S  L  I  Q  K  S  N  V  K  N  E  L  M  T   195

TTCAGGATAAACTGTGGCGGTGCCTTCTTAAGAATAGTTTTAATGAAAGCATATATTAACCCTCGTTGTAACGAATCAGCAGAATTCTTCAGCGATCGTA  91300
  K  L  I  F  Q  P  P  A  K  K  L  I  T  K  I  F  A  Y  I  L  G  R  Q  L  S  D  A  S  N  K  L  S  R  L   161

ACACCGTATGAATCGCATTGATGTTGAGCATTTGATCTAAAATTGTGTTTTTAAAGGTGTTTTCTAAATGCGCTAAACAAGCAGCACTTAATTCAAACGA  91400
  V  T  H  I  A  N  I  N  L  M  Q  D  L  I  T  N  K  F  T  N  E  L  H  A  L  C  A  A  S  L  E  F  S   128

TATATTGATAGGATGCTTTTCTGAATACTTGGTAACCAGTACGGTTGTTTCTTTAGGCGCAGTTACGTCGTTTCCTGTAGCGACTCTAGGTAATTGTATA  91500
  I  N  I  P  H  K  E  S  Y  K  T  V  L  V  T  T  E  K  P  A  T  V  D  N  G  T  A  V  R  P  L  Q  I   95
```

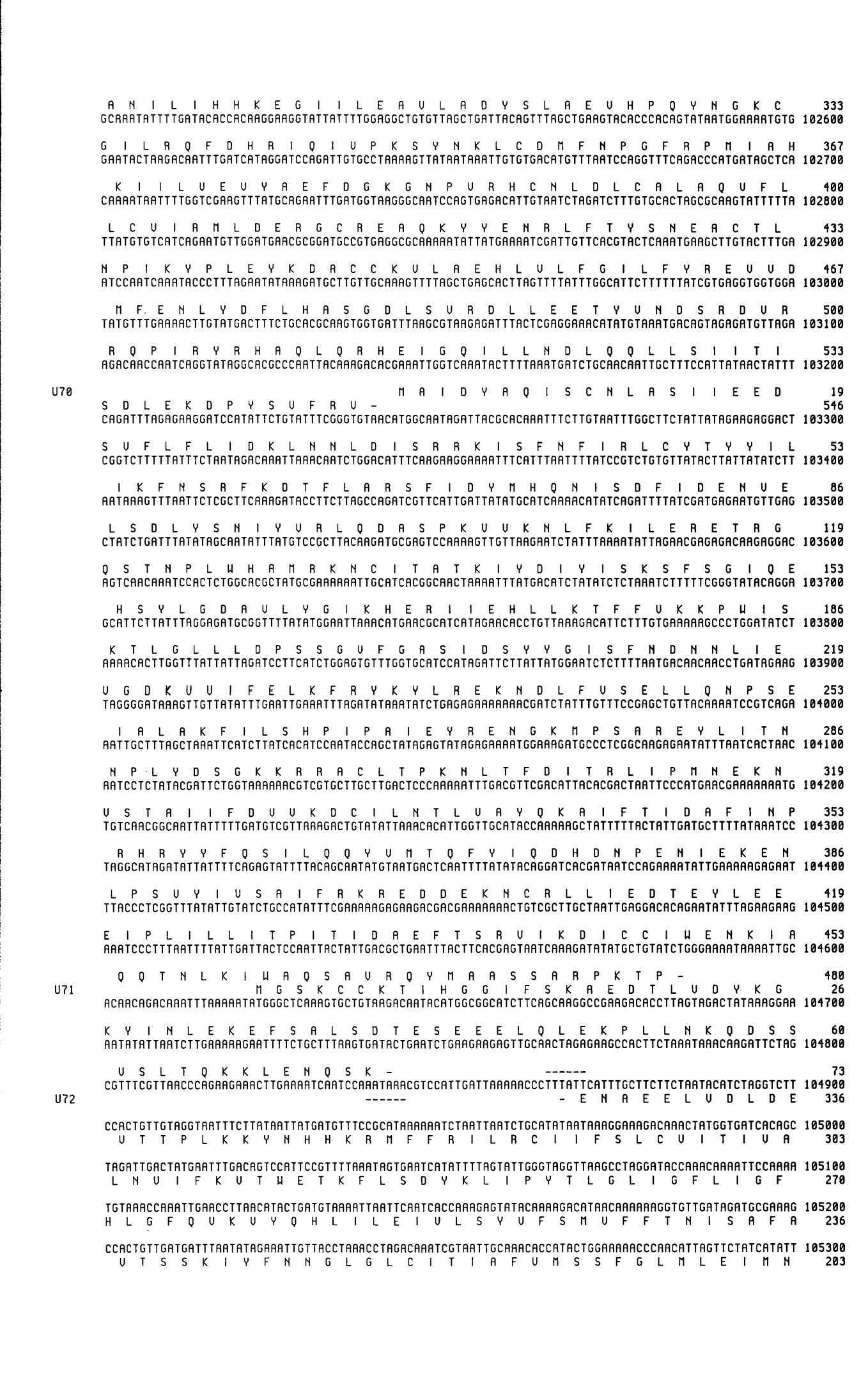
```
TAAAACAGAATTTTTCCCAGTGACATTTTGTCTAGGTCGTTGAAACGGATAACATTTGCAGCAACCGCTATGGACGTATGAAAAAAATCTATCCATTCAG  91600
Y  F  L  I  K  G  L  S  M  K  D  L  D  N  F  R  I  V  N  A  A  V  A  I  S  T  H  F  F  D  I  W  E  T    61

TCCGATTACAGTAAATCCCAAGTAATGCTTCGAAACTTATGTTATAACGATCAGAGTCGTCACCGTAATACAATCTTAAGTTTTCAAAAAGTTGTTCAGC  91700
  R  N  C  Y  I  G  L  L  A  E  F  S  I  N  Y  R  D  S  D  D  G  Y  Y  L  R  L  N  E  F  L  Q  E  A    28

U58                                                                                                  M  I  D     3
AGTTTGTGTTCTGATATCATCAAACACGTTTGGAGAAACATCTAGTTTTGGGAAAATTTCAGCTGTGCGCCAATTTTCCATGGTTAGTAATAATGATAGA  91800
 T  Q  T  R  I  D  D  F  V  N  P  S  V  D  L  K  P  F  I  E  A  T  R  W  N  E  M                      1

  S  I  S  E  E  T  L  I  V  K  S  Y  T  V  N  H  C  A  K  N  V  P  V  F  I  N  S  Y  D  L  T  A  E    36
CTCAATATCTGAAGAAACTTTAATAGTAAAGAGCTACACCGTTAATCACTGTGCTAAAAATGTTCCAGTGTTTATTAACTCCTATGATTTAACCGCAGAA  91900

 V  A  K  N  E  D  V  R  L  A  R  Q  V  Q  I  S  L  E  K  I  D  E  V  I  E  S  I  F  S  A  S  G  P    69
GTGGCCAAAAATGAAGACGTGCGATTAGCGCGACAAGTTCAAATTTCATTAGAGAAAATAGATGAAGTTATAGAATCAATTTTTTCTGCGTCTGGTCCTA  92000

S  V  E  N  V  K  D  Q  A  K  F  A  L  C  R  L  L  L  G  P  V  S  I  P  C  Y  C  E  E  W  D  V  N  F   103
GCGTTGAAAATGTAAAAGATCAGGCAAAGTTTGCTTTGTGTCGTTTACTGCTTGGTCCTGTGAGTATTCCGTGCTACTGCGAAGAATGGGATGTCAATTT  92100

  Y  L  T  K  C  S  Y  N  C  E  G  P  V  L  Y  I  Y  K  N  A  S  Q  C  C  E  S  T  Y  R  F  S  I  M   136
TTATCTGACAAAATGTAGTTATAATTGCGAAGGCCCGGTTCTATATATCTATAAAAATGCTTCTCAATGCTGTGAAAGCACATATCGTTTTTCTATCATG  92200

 T  N  Y  H  S  T  H  I  F  R  G  L  L  S  L  Q  E  W  N  S  H  L  S  N  I  L  C  T  C  S  N  V  T   169
ACTAATTATCATTCCACTCACATCTTTAGAGGATTATTATCATTACAAGAATGGAATAGTCATCTATCAAATATCTTATGTACTTGTTCGAACGTAACAG  92300

G  D  K  Y  T  A  T  I  F  P  N  N  A  S  I  Y  L  E  Y  Y  P  Y  F  L  C  Y  L  C  K  H  L  S  I  I   203
GTGATAAATATACTGCAACAATCTTTCCAAACAATGCTTCAATTTACTTGGAATATTATCCGTATTTTCTATGCTATCTATGCAAGCATCTGTCTATCAT  92400

  D  I  E  Q  C  T  N  E  L  I  A  F  L  G  P  K  T  S  Q  R  I  I  I  H  Y  K  L  L  F  G  F  R  S   236
TGATATTGAGCAATGTACTAATGAATTGATAGCTTTTCTTGGTCCAAAGACTTCTCAAAGGATTATAATTCACTATAAACTGTTATTCGGGTTTCGATCT  92500

 K  P  M  N  F  T  V  S  L  L  E  Q  V  F  T  L  E  I  Q  K  L  Y  Y  S  V  S  K  H  N  S  T  T  A   269
AAACCAATGAATTTCACTGTTTCTTTGTTAGAACAGGTTTTCACCCTTGAAATCCAAAAACTCTACTATTCCGTTAGTAAGCACAACAGTACAACAGCAG  92600

D  F  F  N  V  I  T  A  K  F  A  E  D  K  Y  F  V  L  R  T  F  K  L  S  A  Q  I  T  P  G  I  Q  S  F   303
ATTTTTTCAATGTCATTACCGCTAAATTTGCAGAGGACAAATATTTTGTTCTACGAACATTTAAATTGTCTGCGCAAATCACTCCTGGTATTCAAAGTTT  92700

  C  S  L  K  F  K  L  Q  T  L  Y  L  N  L  K  I  M  K  N  T  K  L  S  I  S  N  S  F  Y  H  G  K  T   336
TTGTTCATTGAAATTCAAACTCCAGACCTTATATCTAAATTTGAAAATTATGAAAAACACAAAATTATCCATTTCTAATAGTTTTTATCATGGTAAAACT  92800

 L  Y  T  L  D  E  K  Q  L  V  W  R  N  L  L  L  I  Y  Y  G  Y  N  L  K  D  N  V  K  Q  T  Q  E  E   369
TTATATACACTGGATGAAAAGCAACTTGTTTGGAGAAATTTATTGTTAATTTACTATGGTTACAATTTAAAAGACAATGTAAAACAAACACAAGAAGAGA  92900

S  L  L  S  M  H  Y  I  R  I  L  E  R  L  S  L  K  S  F  R  E  I  N  Q  Q  F  R  F  E  I  P  S  Y  Q   403
GTTTGTTGTCGATGCATTACATACGAATATTGGAAAGATTGTCTCTAAAAAGTTTTCGCGAAATTAATCAACAATTTAGATTTGAAATTCCGAGTTACCA  93000

  E  K  T  L  Q  F  I  P  G  G  N  D  F  A  E  I  T  S  V  T  H  G  E  T  T  V  N  A  F  N  T  N  R   436
AGAGAAAACTTTGCAGTTTATTCCAGGTGGAAATGATTTCGCAGAAATCACATCGGTCACGCATGGAGAAACAACTGTGAATGCATTTAATACAAATAGG  93100

 V  M  N  V  K  A  A  L  S  G  E  I  H  C  V  L  H  R  I  P  K  S  M  T  H  S  F  V  M  Y  K  R  T   469
GTCATGAATGTGAAAGCTGCTCTTTCAGGAGAAATACACTGTGTTTTACATCGTATTCCTAAAAGCATGACACATAGTTTTGTGATGTATAAGCGCACTT  93200

F  K  E  P  S  L  T  V  S  T  F  I  S  N  D  D  F  T  T  S  S  L  N  I  N  I  R  G  P  Y  C  D  F  L   503
TTAAAGAACCCTCTTTAACAGTGAGTACCTTCATTTCAAATGATGATTTCACCACAAGTTCATTGAACATTAACATTCGAGGTCCCTACTGTGATTTTTT  93300

  Y  A  L  G  V  Y  R  L  H  V  N  I  Q  D  F  F  L  P  A  F  V  C  N  S  N  N  S  M  D  L  H  G  L   536
ATATGCATTGGGTGTGTATAGATTACATGTTAACATACAAGATTTTTTTTTACCTGCTTTTGTTTGTAACAGTAATAATTCGATGGATTTGCATGGGCTA  93400

 E  N  Q  G  I  V  R  K  R  K  K  K  V  Y  W  I  T  N  F  P  C  M  I  S  N  S  E  K  V  N  V  G  W   569
GAAAATCAAGGAATTGTGCGAAAGCGTAAAAAGAAGGTTTACTGGATAACTAATTTTCCTTGTATGATCTCCAATTCTGAAAAAGTTAATGTTGGTTGGT  93500

F  K  A  G  T  G  I  I  P  K  V  S  G  T  D  L  K  N  V  L  L  K  E  L  I  S  I  G  E  I  P  N  I  T   603
TTAAAGCGGGAACAGGCATAATACCCAAGGTTTCTGGAACAGACTTAAAAAATGTTCTTTTGAAAGAACTCATAAGCATCGGAGAAATTCCCAATATTAC  93600

  F  D  M  D  L  H  A  L  L  T  L  L  E  K  R  N  M  H  Q  V  P  F  L  I  K  Q  F  F  M  F  L  R  L   636
TTTTGATATGGATTTACATGCTTTGTTAACTCTTTTAGAGAAACGAAATATGCATCAGGTTCCATTTCTTATTAAACAATTTTTTATGTTTCTTCGTTTA  93700

 G  L  L  V  G  Y  G  R  K  Q  E  R  K  V  H  H  I  M  L  F  L  I  Q  K  G  F  F  D  F  S  K  N  S   669
GGTCTTTTAGTGGGATATGGACGTAAACAGGAAAGAAAGGTCCATCACATTATGCTATTCTTAATACAAAAGGGTTTTTTCGATTTTTCGAAGAATTCAG  93800

V  A  N  S  K  I  K  H  A  C  A  L  V  G  S  R  L  A  N  N  V  P  K  I  L  S  K  Q  K  K  M  K  L  D   703
TTGCCAACAGTAAAATTAAACATGCATGCGCTCTAGTTGGAAGTCGACTTGCCAACAATGTGCCAAAAATTCTGTCTAAGCAGAAAAAAATGAAACTGGA  93900

  H  L  G  R  N  A  N  A  L  T  V  L  R  F  I  V  E  N  G  Y  Y  K  R  K  T  I  F  R  K  L  L  K  Y   736
TCATTTGGGTCGAAACGCAAACGCTTTAACCGTTTTACGTTTTATTGTAGAAAATGGTTATTATAAAAGGAAGACAATTTTTCGCAAACTTTTGAAGTAT  94000

U59                                                                                  M  T  A  K  Q  I  F  P  V  W  K    11
 L  A  T  T  S  F  N  A  H  V  Q  T  E  S  N  R  L  L  N  L  M  H  N  D  S  K  T  N  F  S  S  L  E   769
TTAGCTACTACATCTTTTAATGCTCATGTTCAAACTGAAAGTAATCGTTTACTCAACTTGATGCACAATGACAGCAAAACAAATTTTTCCAGTTTGGAAA  94100

 D  C  T  L  Y  V  N  N  E  T  A  T  V  H  E  I  L  N  S  D  L  S  E  L  L  Q  L  K  T  E  F  V  S    44
R  L  Y  T  L  R  -                                                                                  775
GACTGTACACTTTACGTTAACAATGAGACAGCAACTGTGCATGAAATCTTGAATTCCGATTTAAGTGAACTGTTACAGTTAAAGACGGAATTTGTATCTA  94200

M  T  D  L  C  V  Y  I  T  G  C  I  N  Q  N  I  S  S  I  T  I  Y  W  H  A  Y  S  E  V  I  Y  A  L  T   78
TGACAGACCTATGTGTTTACATTACTGGATGTATAAATCAGAATATTTCCAGCATCACGATATATTGGCATGCTTACAGTGAAGTAATTTATGCTTTAAC  94300
```

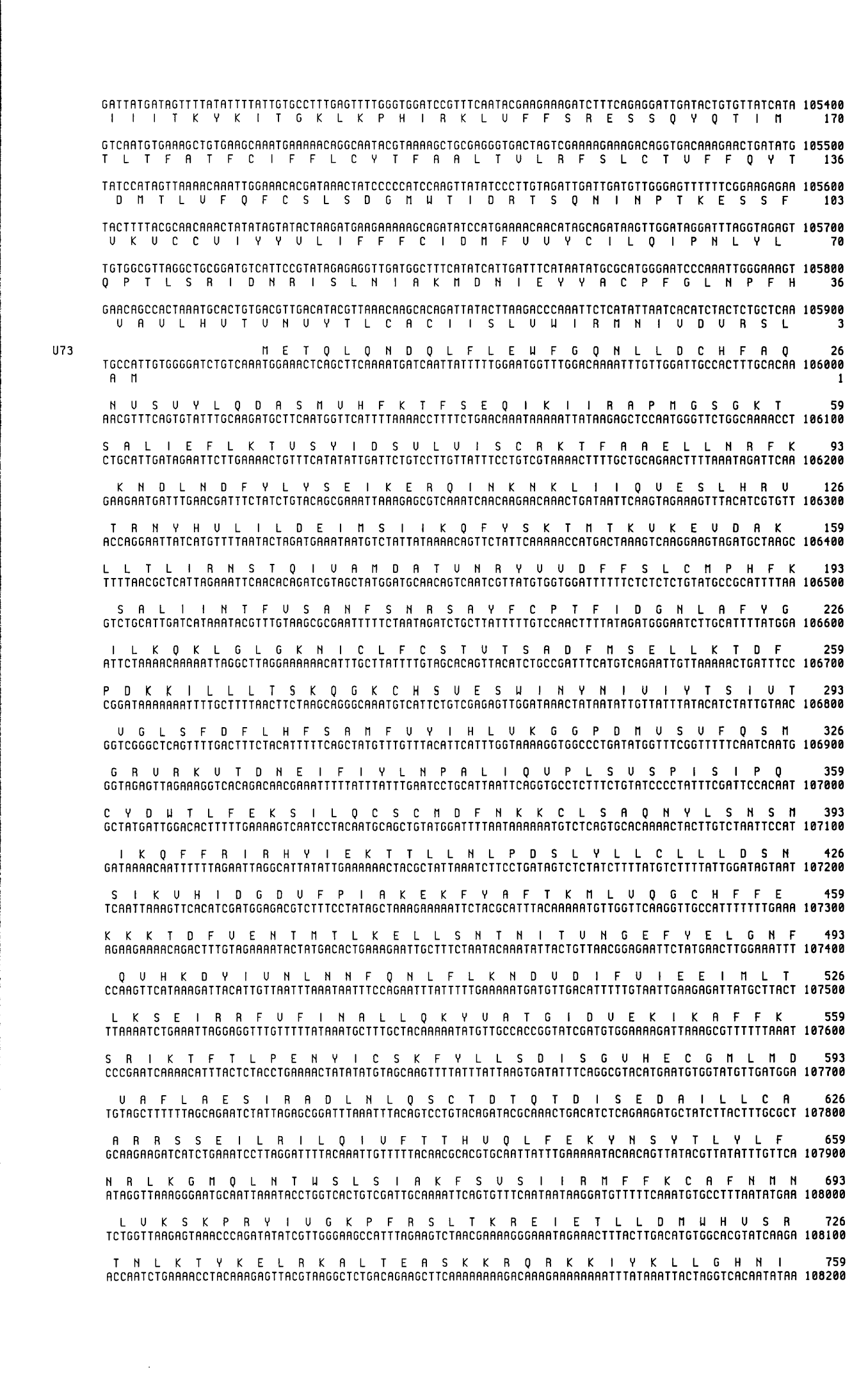


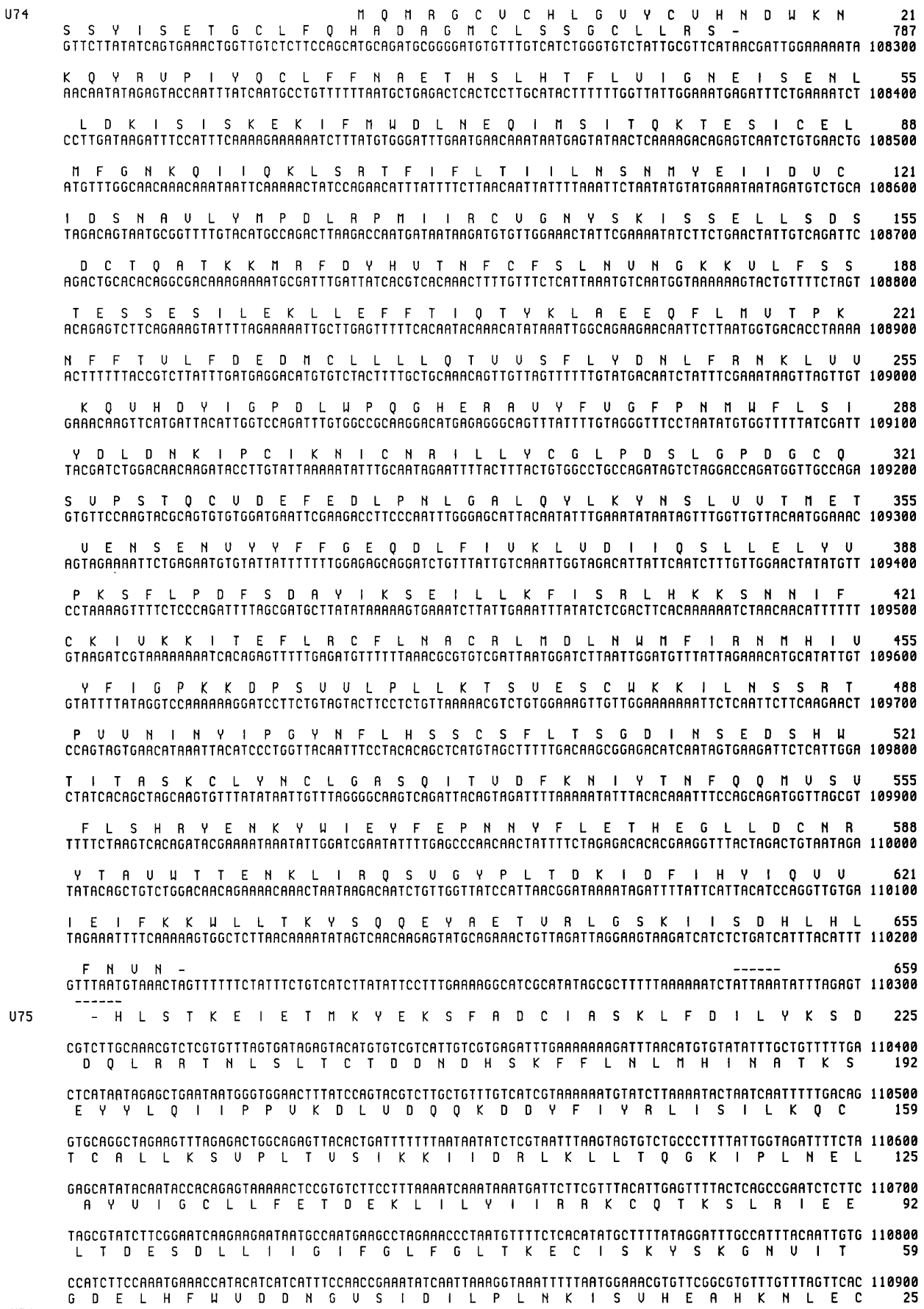

I R V S R R K R E D N I L K K M R I E Y G N A L A L - 211
U64 M E M H L L C E T M F T C R 14
ATTCGTGTTAGTAGAAGAAAGAGGGAGGATAACATACTTAAAAAAATGAGGATTGAATATGGAAATGCACTTGCTCTGTGAAACAATGTTTACTTGCAGA 97100
K N N I L P V H L C I L L D D V I H K E K V K A I E G I F F Q C V 47
AAAAATAATATTTTACCGGTACATTTATGTATTTTACTGGATGATGTTATACATAAAGAGAAAGTAAAAGCTATAGAAGGGATCTTTTTTCAGTGTGTAT 97200
F F K E K L V Y T E W T K I K F T Y V L H D L V I S Q I F K N A C I 81
TTTTTAAAGAAAAGCTTGTATATACGGAATGGACAAAAATAAAGTTTACTTATGTGTTACATGATCTTGTAATTTCTCAAATCTTTAAGAATGCCTGTAT 97300
K E V I H G A L I L S V P I N I D N L H F D T D I L I L K I I Y P 114
TAAAGAAGTAATACATGGGGCATTAATTCTTTCAGTTCCCATAAATATTGATAACCTACATTTTGATACAGATATTTTAATTCTAAAAATTATTTACCCA 97400
H F L H D D I V I K L S E I L S G A P R I Q K T V E K K Q E V E K 147
CATTTTTTGCACGATGATATTGTCATAAAATTATCGGAAATTTTGTCTGGAGCACCTCGCATACAAAAAACAGTGGAAAAAAAACAAGAGGTGGAAAAAC 97500
P F F H I P A K L G D L T K E D P I S F N H H G P L E P P S T V R G 181
CTTTTTTCCATATTCCTGCAAAACTTGGAGATCTCACAAAGGAAGACCCTATTTCGTTCAACCATCATGGTCCGTTAGAACCTCCATCAACTGTTAGAGG 97600
L K Q S A N V R H S H P I S R P E K A N V T F L S D S W Y S Q N L 214
ATTAAAACAATCTGCGAACGTTAGGCATAGTCATCCAATATCAAGGCCTGAAAAAGCGAACGTAACCTTTCTAAGTGATTCGTGGTACAGCCAAAATCTA 97700
K C D F I S D I Q Q R H V L V I F W Y E L S K G I Q M Q I K N I Q 247
AAGTGTGACTTCATATCTGACATTCAACAAAGACATGTGCTTGTCATTTTTTGGTATGAGTTATCGAAAGGGATACAAATGCAAATTAAAAATATTCAAA 97800
I P P E N L F S S I T N Y L D R V N T Y L D E I A E R T F R C I T T 281
TTCCTCCTGAAAATTTGTTTTCATCAATAACGAATTATTTAGATAGAGTCAACACATATCTAGACGAGATTGCTGAAAGAACTTTTCGATGTATTACTAC 97900
N M E I Q N R H L P Q K F N S H F Q I E F N C T H L I Y G M E L A 314
TAACATGGAAATTCAGAATAGACATCTTCCACAAAAATTTAATAGTCATTTTCAAATAGAGTTTAATTGTACTCACTTAATTTATGGTATGGAATTGGCG 98000
R D F W I L S L D R N S C V L K A M A S H F L H K K K G R S S L S 347
AGGGATTTTTGGATTTTGTCTTTAGATAGAAATAGTTGTGTTTTAAAAGCTATGGCCAGTCATTTTCTTCATAAAAAAAAAGGGAGAAGCTCACTTAGTT 98100
S N E F W A D L I D C T T G K T L Y G E K V R W Q L N S E T S L Y S 381
CGAATGAATTTTGGGCTGACTTAATTGATTGCACTACCGGAAAAACCTTATATGGAGAGAAAGTACGGTGGCAATTAAATTCTGAAACGAGCTTATACTC 98200
T F R K N Q N I S W E L Q P N C Y A L Y M S E N L K L Y W V L P G 414
CACATTCAGAAAAAATCAAAACATTTCATGGGAATTACAGCCTAATTGTTATGCACTATATATGTCTGAGAATTTAAAGCTGTATTGGGTATTACCCGGG 98300
U65 M N F S S I G N L G C L R S F L Q N E C 20
G F C V S G T F K L K E N D E F F F D W Q F G M S - 439
GGGTTCTGCGTGTCTGGAACTTTTAAATTAAAAGAGAACGATGAATTTTTCTTCGATTGGCAATTTGGGATGTCTTAGGAGTTTTTTACAAAATGAGTGT 98400
N W F S V C K K K L Y H E Y R C V A T S S P V F A V D K F K D C L 53
AATTGGTTTTCGGTGTGTAAAAAAAAGTTGTACCACGAATATCGTTGTGTGGCAACATCTTCTCCTGTTTTCGCTGTTGATAAATTTAAAGATTGCTTGC 98500
H C N I I I L K K N L D F V F S L A I N G I H A G Q F A T N S I K L 87
ACTGTAATATAATTATTTTAAAAAAGAATTTGGATTTTGTTTTTAGCTTGGCCATAAACGGAATACATGCCGGACAGTTTGCAACAAATTCCATTAAACT 98600
K K I I I T N D L V Y Y I L E L G S L T V T D L H F I P K Y N S E 120
AAAAAAAATTATAATAACAAACGATTTGGTGTACTACATATTAGAATTGGGATCTTTAACGGTAACGGATTTACATTTCATTCCAAAATACAATAGTGAA 98700
H V L N V R P I T P N L I Y D T C S I V S Y D E A K L L T V K G P 153
CATGTGCTGAATGTGCGACCTATTACGCCAAATCTAATTTATGATACTTGTTCAATTGTTAGTTATGACGAAGCTAAACTTTTAACTGTTAAAGGACCTG 98800
G E N K L I P L G C G S W C L N N I G R Y Y V Y T F V L V Y D L Y L 187
GAGAAAATAAATTAATTCCTTTGGGTTGTGGTTCTTGGTGTCTCAACAACATTGGACGTTATTATGTGTACACTTTTGTCCTGGTGTACGATTTATATCT 98900
A C F E K N T L P S L S K V V F D M I S C N N K H C V F C K D H S 220
GGCTTGCTTTGAAAAAAACACTTTGCCATCATTATCTAAAGTAGTTTTTGATATGATTTCCTGCAATAATAAACATTGTGTTTTTTGTAAGGATCATAGT 99000
K H V E Q T G K T V G C T D N Q E T C F C Y T P C K K K M A K I S 253
AAACACGTAGAACAAACTGGCAAAACCGTTGGATGCACTGATAATCAAGAAACATGTTTTTGTTACACCCCATGCAAGAAAAAAATGGCTAAGATTTCCA 99100
N Q D L S S L L C D Q E L D L L D L I Y P E K P T S L S T D I N A Y 287
ACCAAGATCTGTCTTCATTATTATGTGATCAAGAGCTCGATTTACTTGATCTAATTTACCCTGAAAAGCCCACTAGTCTTTCAACTGATATAAATGCTTA 99200
V H G H K N Q E P V V L R N T N W I L I R L D P A I S R L I L L S 320
CGTACATGGACATAAGAATCAAGAACCAGTCGTTTTAAGGAATACAAACTGGATATTAATTCGTCTTGACCCAGCAATTAGCAGATTAATACTCCTTTCC 99300
C P V C K R I V S R - < end of U66 exon 1 330
TGTCCAGTCTGTAAACGCATAGTAAGTAGGTAAGAAAAATTTGTTACTTACGTGAGTGTTGTAACAACTAGCAAACAGTGCTGTACTTTTTATTGTTTCG 99400
U66 H T N Y C S A F L A T S K I T E 273
exon 1
TGTTCGATAGTAATCACATTATCTTGACAGGTGATATTCTTTTGCGGGAAAAAACGTCTACATTTAAATTCAACATCTTTCATAACAAAGTGCGAAACGT 99500
H E I T I V N D Q C T I N K Q P F F R R C K F E V D K M V F H S V H 239

GTTTTTGATGTGCAACATAACCAATGCTGATTCCTTCTAAATTCTTTAATAAAAAACAAATAACTGGTATCATGAACCATGTTTTACCGTGTCTTCTGGG 99600
K Q H A V Y G I S I G E L N K L L F C I V P I M F W T K G H R R P 206
AACTAGAAAAACACTAGCTTTCTGTTTAAGTATGTTTACGCTGCTTTCGTTTATAAACTCTATGTCAAATTTATATTTTAAGTAATCAAGGACATGATTG 99700
V L F V S A K Q K L I N V S S E N I F E I D F K Y K L Y D L V H N 173



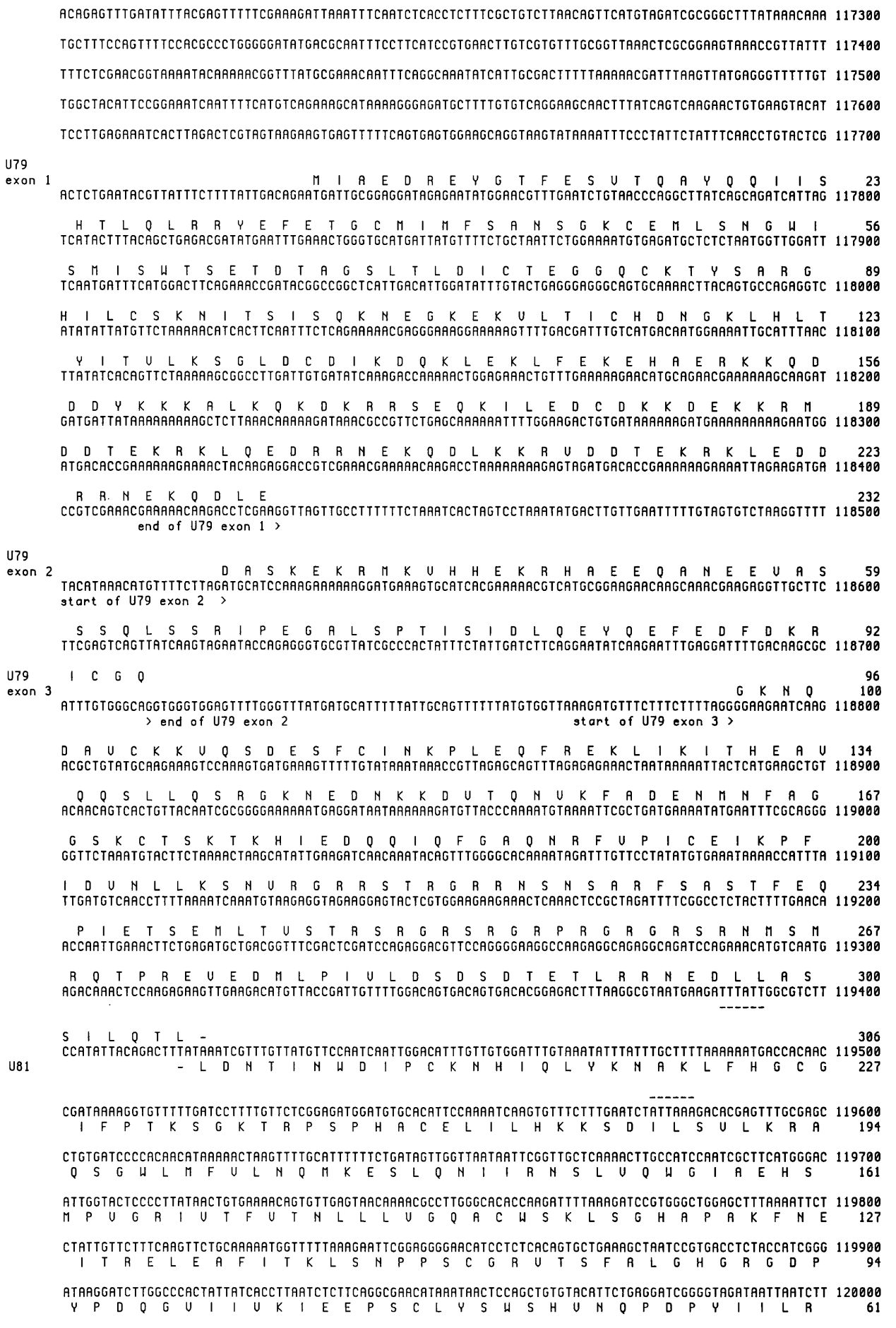




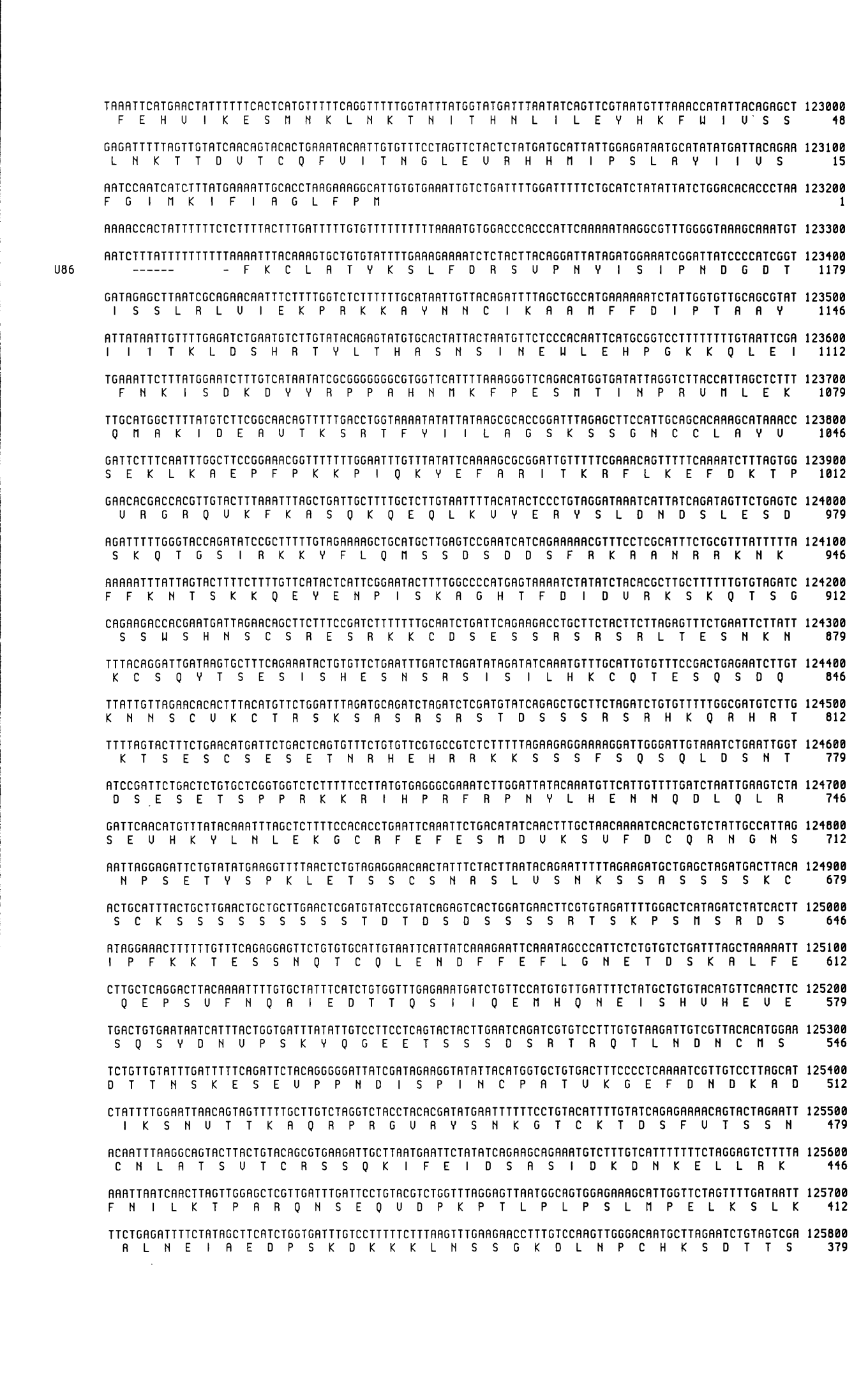


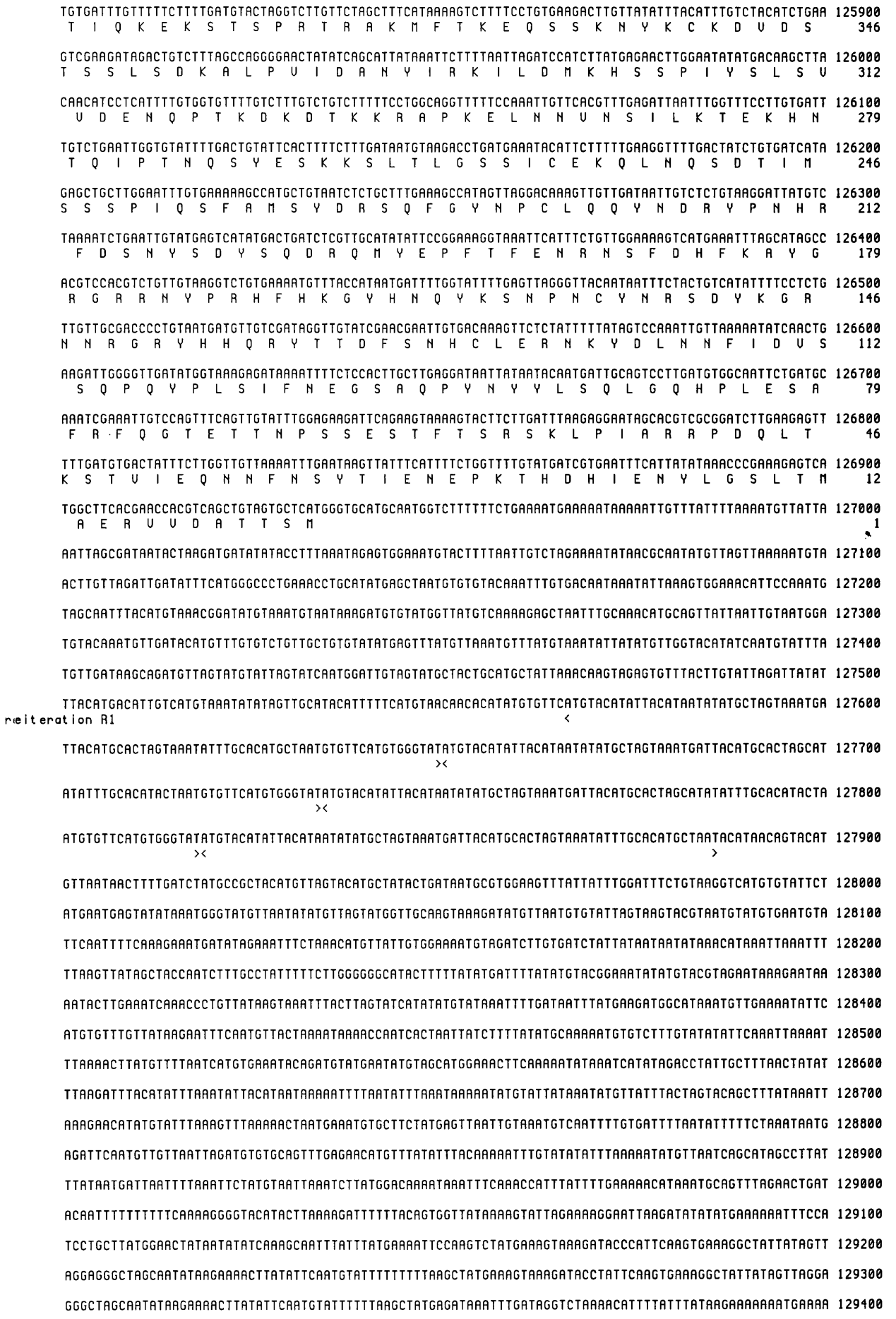
```
AGATATTCCGGACGCTTGATATATTGAATATGTCTAGAGTTTCGAATCCTTTTAAGCTTGCTGGCAACGTCATCTTGCGGATATTGAAAATCTTCTGGGT 111000
  I  N  R  V  S  S  I  N  F  I  D  L  T  E  F  G  K  L  S  A  P  L  T  M                                1
L  Y  E  P  R  K  I  Y  Q  I  H  R  S  N  R  I  R  K  L  K  S  A  V  D  D  Q  P  Y  Q  F  D  E  P  N   607

TTGTGGATCCTAAATCAGGTGTCAAAGTCAAATCACTTTGTTTCTGAGTATGTATTAAATGCGAAAGCTCGTTAATCTGCCGTATATAGATTAATCTTAA 111100
  T  S  G  L  D  P  T  L  T  L  D  S  Q  K  Q  T  H  I  L  H  S  L  E  N  I  Q  R  I  Y  I  L  R  L   574

ATAATGCTTAATTTTACAGTTTGAGTACACAGCTAACATAATCTCTTTTAGAGAGAGGAACAGAAACTCTTCTGGGAGCAGTTCAAAATGTCTTAATTTT 111200
 Y  H  K  I  K  C  N  S  Y  V  A  L  M  I  E  K  L  S  L  F  L  F  E  E  P  L  L  E  F  H  R  L  K    541

AGAACATGTTTCATGAAGGGTATAAGAAGCAATTCTAGCGTGTAATTTGAATATGAGATGCTGTCTTCCTGTCCTTGATTGTTGACATTTTTTCGTAGAC 111300
L  V  H  K  M  F  P  I  L  L  L  E  L  T  Y  N  S  Y  S  I  S  D  E  Q  G  Q  N  N  V  N  K  R  L  R   507

GGAATGTCCTTGTGTATTCAGTTTCCCATAATTGATCAATTTTTTTGTCTGCGTATTCAATATCTGGAATATATTGTGAAAAAAAACTGTTTGCCACAGC 111400
  F  T  R  T  Y  E  T  E  W  L  Q  D  I  K  K  D  A  Y  E  I  D  P  I  Y  Q  S  F  F  S  N  A  V  A    474

TCTGGTGTCGTCTATTGATACACTAGTGAATGGGAGACTTTGAACTTTGTTCAGAGCTTTAGCGAGAGACAAGCTTTCAATATCTGAATCTAAAGTAAAG 111500
 R  T  D  D  I  S  V  S  T  F  P  L  S  Q  V  K  N  L  A  K  A  L  S  L  S  E  I  D  S  D  L  T  F    441

TCATGTTTAGAAGTTTCGGATAAATTTATTAATGAATGCATGTCTTTAATCTTATTTTCAAATAGCCGATTTGCAGTTTTCAGTGTTTCGATTTCGTCCA 111600
D  H  K  S  T  E  S  L  N  I  L  S  H  M  D  K  I  K  N  E  F  L  R  N  A  T  K  L  T  E  I  E  D  L   407

GTTGTGATTGAATCTGTTCCTCCATGCATTTTATGATTTGTTTTTTGAAGATATCTCTAATATCGCGATCATTATTGTCCTGCCCACGACGTTGCAAATT 111700
  Q  S  Q  I  Q  E  E  M  C  K  I  I  Q  K  K  F  I  D  R  I  D  R  D  N  N  D  Q  G  R  R  Q  L  N    374

TCCTTTGCTGATTAATAATTTATTTTGGTCAATTATAGACGGCGTGATGTCTTGTAAAAAACCCTCAATACTATCTGTTATTCCAATTTTAGATTTACTA 111800
 G  K  S  I  L  L  K  N  Q  D  I  I  S  P  T  I  D  Q  L  F  G  E  I  S  D  T  I  G  I  K  S  K  S    341

TCGGACAAGTTTAATAGAAACTTTATAAGACTTTTTTTCGCATCGCTGGTTTTTTCCTCTCTTTCTATAAGTTCTAGAATTTTTTTGTTATCAACACTGT 111900
D  S  L  N  L  L  F  K  I  L  S  K  K  A  D  S  T  K  E  E  R  E  I  L  E  L  I  K  K  N  D  V  S  N   307

TCTTTGATGTAGTTAAAACTTTAATTGGAAAAGTATTCAGCAATTGACACATTTTACGGTGCCTGTGTAGGTTTTCGTGTTTATTCAATTCTGAGTATAA 112000
  K  S  T  T  L  V  K  I  P  F  T  N  L  L  Q  C  M  K  R  H  R  H  L  N  E  H  K  N  L  E  S  Y  L    274

GTGAGCTAGGGGTGTTGTAAACAGCAATTGATTATTTATGAAATATGGTGCACTGTAAACAAAGTACTCTAGTTCCTTGTTATTGGAGAGACAGGTTATA 112100
 H  A  L  P  T  T  F  L  L  Q  N  N  I  F  Y  P  A  S  Y  V  F  Y  E  L  E  K  N  N  S  L  C  T  I    241

TTGCCGTTTGACTTGTTTCGATAAATGTTTATGTTGGTGAAATCGGCAATAGATGTTGAGGAGATCTGTGTTAAAATGTGCAGTAGAAAGTTCTGTTTGT 112200
N  G  N  S  K  N  R  Y  I  N  I  N  T  F  D  A  I  S  T  S  S  I  Q  T  L  I  H  L  L  F  N  Q  K  N   207

TTCGTATGATTTCTTGAATCAATCTAGAGGAAATACTTTTTGCTAAGGGTGTCAGCGTAACCTTGTTTTTTGGTAATGTGACAATTGGTGTTATTCCAAT 112300
  R  I  I  E  Q  I  L  R  S  S  I  S  K  A  L  P  T  L  T  V  K  N  K  P  L  T  V  I  P  T  I  G  I    174

CGTACAAATCCAATCAATGTATTTATTATAAAATGTCGTTATCGAGATTTTGAAGGATGTAGTTTATCGTTTGCAAAATGCCTTCCAGGATTACCTGTTTG 112400
 T  C  I  W  D  I  Y  K  N  Y  I  D  N  D  L  N  Q  L  I  Y  N  I  T  Q  L  I  G  E  L  I  V  Q  K    141

ATTGCATAATACCAGATATTAAAATATTCTTGAAATGACATGCCTTTACTGCGGAACGGAAATGCTTTCAGATGACGTTGCCATTCAGCATTGATCTGCT 112500
I  A  Y  Y  W  I  N  F  Y  E  Q  F  S  M  G  K  S  R  F  P  F  A  K  L  H  R  Q  W  E  A  N  I  Q  Q   107

GGATGGATATACTATTCAAGAGAGTGTGTGACAGTGTGTAAAACATTTGCTTCTTAAAGACAGCCGAGTCTCTAAGAACACTGTGTATAGATTGTCCCTT 112600
  I  S  I  S  N  L  L  T  H  S  L  T  Y  F  M  Q  K  K  F  V  A  S  D  R  L  V  S  H  I  S  Q  G  K    74

U77                                                                M  S  L  S  S  L  F  N  G  K  Y  D    12
AAAGAATTCTTTTTTACCGTGTAGCATTTCTATAAATTTTATAGTCTTAATAGAAGGAAATAAAATGTCATTGTCTTCATTATTCAATGGTAAATATGAC 112700
 F  F  E  K  K  G  H  L  M  E  I  F  K  I  T  K  I  S  P  F  L  I  D  N  D  E  N  N  L  P  L  Y  S    41

 T  K  F  L  L  N  M  S  S  A  A  K  V  E  L  I  V  E  K  V  A  A  L  A  D  A  C  L  E  T  P  L  P    45
ACAAAATTCCTTTTAAACATGTCATCTGCTGCTAAGGTGGAGTTGATTGTTGAGAAAGTTGCAGCTTTAGCTGATGCCTGTTTAGAGACGCCGTTACCAA 112800
V  F  N  R  K  F  M  D  D  A  A  L  T  S  N  I  T  S  F  T  A  A  K  A  S  A  Q  K  S  V  G  N  G  V     7

T  D  W  F  R  N  I  L  D  P  E  L  E  F  N  S  N  F  E  E  I  H  S  I  G  D  E  E  F  A  Q  P  L  P    79
CTGACTGGTTTCGTAACATTCTTGATCCTGAATTAGAATTTAATAGCAATTTTGAAGAAATTCATTCTATAGGTGATGAAGAATTTGCTCAACCGTTGCC 112900
  S  Q  N  R  L  M                                                                                      1

  F  L  P  F  R  V  L  L  I  T  G  T  A  G  A  G  K  T  S  S  I  Q  T  L  A  A  N  S  D  C  L  I  T   112
ATTTTTACCTTTTAGAGTATTATTAATAACCGGTACTGCGGGTGCGGGCAAAACAAGCAGCATTCAAACCTTAGCAGCTAATAGTGATTGTCTTATAACT 113000

 A  T  T  S  I  A  A  Q  N  L  S  G  L  L  N  R  T  K  S  A  Q  V  K  T  I  F  K  T  F  G  F  N  S    145
GCTACCACTTCCATTGCTGCTCAAAATCTTAGTGGTTTATTGAACAGAACCAAATCTGCGCAAGTGAAAACAATTTTTAAAACATTTGGTTTTAATAGTT 113100

S  H  V  S  M  N  E  R  I  S  C  S  V  T  T  L  D  S  I  A  D  Q  Q  K  H  D  L  S  T  Y  W  N  V  I   179
CACATGTATCTATGAATGAAAGAATTAGTTGTTCAGTAACAACTTTAGATTCGATTGCGGATCAGCAAAAACATGATTTATCTACATATTGGAACGTCAT 113200

  A  D  I  A  E  R  A  L  N  A  A  N  G  K  A  K  V  I  P  D  L  C  E  S  S  V  I  V  I  D  E  A  G   212
CGCAGATATAGCGGAAAGAGCTTTGAATGCAGCAAATGGGAAAGCCAAAGTGATACCTGATCTATGTGAAAGCAGTGTGATTGTGATTGATGAAGCAGGA 113300

 V  I  L  R  H  I  L  H  T  V  V  F  F  Y  W  F  Y  N  G  L  H  K  T  Q  L  Y  K  N  R  V  I  P  C    245
GTGATCTTAAGGCACATTTTGCATACAGTTGTCTTTTTTTATTGGTTTTATAACGGTCTTCATAAAACACAGCTCTATAAAAATAGAGTTATTCCTTGTA 113400

I  V  C  V  G  S  P  T  Q  S  G  A  L  I  S  S  F  N  P  L  T  Q  N  K  D  V  K  K  G  F  D  I  L  S   279
TTGTATGTGTTGGATCTCCTACACAAAGTGGGGCTTTAATTTCATCGTTTAATCCGTTAACACAAAATAAAGACGTAAAGAAAGGATTTGATATCTTATC 113500

  A  L  I  C  D  D  I  L  S  N  Y  C  K  I  S  E  N  W  V  I  F  V  N  N  K  R  C  T  D  V  E  F  G   312
CGCTTTAATATGTGACGACATTCTATCAAATTACTGTAAAATATCAGAAAATTGGGTGATTTTTGTTAATAATAAACGATGCACTGATGTGGAATTCGGC 113600

 E  F  L  K  H  I  E  F  G  L  P  L  K  P  E  L  I  E  Y  V  D  R  F  V  R  P  A  T  Y  I  R  N  P    345
GAATTTTTAAAACATATAGAATTTGGTTTGCCATTGAAACCTGAATTGATTGAGTATGTTGATAGGTTTGTTAGACCGGCAACTTATATTAGAAATCCTA 113700
```

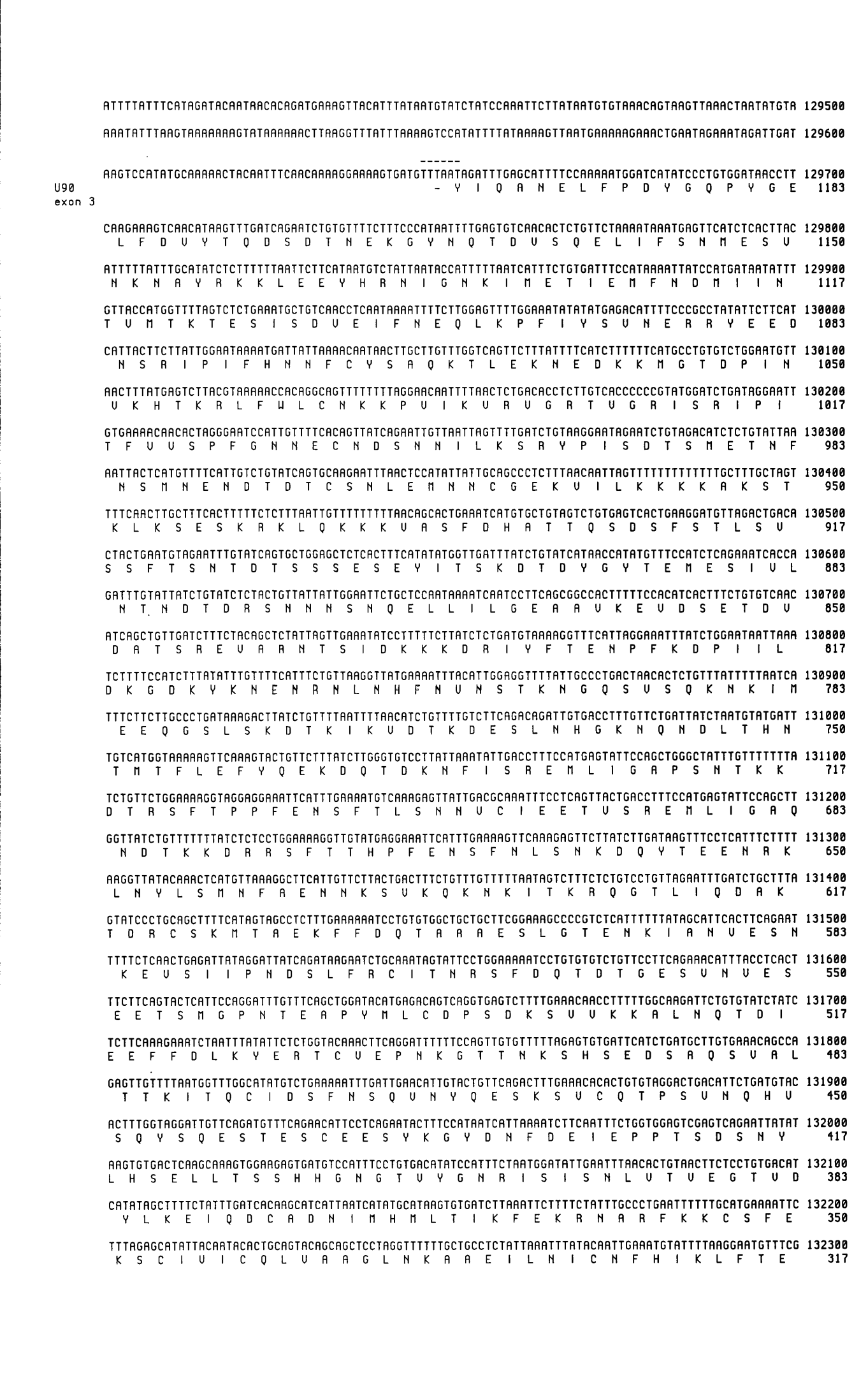
```
T  N  E  I  G  M  T  R  L  F  L  S  H  Y  E  V  K  S  Y  F  K  V  L  H  E  Q  V  E  L  T  N  K  D  N    379
CAAACGAAATTGGAATGACGCGTTTATTTTTATCACATTACGAAGTTAAGTCATATTTTAAAGTTTTACATGAGCAGGTCGAACTGACAAATAAAGATAA 113800

  L  F  T  F  P  V  Y  F  I  I  Q  N  K  A  F  E  D  Y  K  N  E  I  S  N  F  T  L  E  I  E  P  W  F    412
TCTTTTTACTTTTCCTGTTTATTTCATAATTCAGAATAAAGCATTTGAAGATTACAAAAATGAAATTTCAAATTTTACTTTAGAGATTGAACCATGGTTT 113900

 K  T  N  L  H  R  L  N  T  Y  S  Q  F  A  D  Q  D  L  S  K  T  I  Q  I  E  E  I  V  L  D  D  G  S    445
AAAACTAATTTACACAGATTAAATACTTATTCCCAATTCGCTGATCAAGATTTGTCTAAGACCATACAAATCGAAGAAATTGTATTAGATGATGGTTCGG 114000

V  E  E  T  L  I  T  C  H  L  K  H  I  K  H  S  S  I  G  V  T  S  R  T  K  S  S  T  V  G  F  S  G  T    479
TAGAGGAAACTTTGATAACGTGTCATTTGAAACATATAAAACATAGTTCTATTGGCGTTACTTCCAGAACAAAATCTTCAACTGTTGGGTTTTCAGGGAC 114100

  Y  E  K  F  V  E  L  L  Q  S  D  L  F  I  E  K  T  A  C  E  Y  S  V  H  A  Y  S  F  L  T  G  L  M    512
ATATGAAAAATTTGTGGAGCTTCTGCAAAGTGATTTATTTATTGAAAAAACAGCATGCGAATATAGCGTGCACGCCTATTCCTTCTTGACAGGTTTAATG 114200

 Y  G  G  M  Y  S  F  C  L  S  E  F  T  T  S  E  V  M  T  E  I  R  K  I  K  L  P  N  I  D  F  L  Q    545
TATGGGGGTATGTATTCTTTCTGTTTATCCGAATTCACGACTTCTGAAGTAATGACAGAAATAAGAAAGATCAAATTGCCCAATATTGATTTTCTACAAA 114300

T  M  T  A  E  V  S  L  Q  T  F  D  E  S  D  E  Y  Y  D  L  H  I  A  P  T  D  E  E  M  L  A  S  D  P    579
CAATGACAGCTGAAGTTTCTTTGCAAACCTTCGATGAATCAGACGAATACTATGATCTACACATTGCACCTACAGATGAAGAAATGTTAGCTTCGGATCC 114400

  C  P  D  P  F  F  L  K  Y  K  Q  L  P  L  T  N  V  L  T  F  E  E  I  S  Y  L  Y  T  V  F  K  E  I    612
GTGCCCAGATCCTTTTTTTTTAAAGTACAAGCAACTTCCATTAACGAATGTTCTAACATTTGAAGAAATCAGTTACCTTTATACAGTATTTAAAGAGATT 114500

 F  I  S  R  F  A  I  L  Q  R  H  S  K  E  M  F  G  K  S  N  L  I  T  Y  N  R  N  N  V  S  S  K  R    645
TTCATTTCTAGATTTGCAATTCTACAAAGGCACAGTAAAGAAATGTTCGGCAAGAGTAATTTAATCACATATAATAGGAATAATGTTTCGAGCAAAAGAT 114600

C  G  E  I  C  S  H  V  K  S  F  Y  G  M  L  T  Y  A  V  P  A  N  N  Y  T  L  E  G  Y  T  Y  D  N  V    679
GTGGGGAGATATGTTCACATGTTAAAAGCTTCTACGGCATGTTAACATATGCTGTGCCTGCTAATAATTATACTTTGGAAGGATATACTTATGATAATGT 114700

  I  F  L  G  T  D  K  M  L  P  P  I  I  Y  K  R  G  L  P  K  I  V  I  K  D  E  M  G  F  I  S  I  L    712
GATTTTCTTAGGGACGGATAAAATGCTTCCTCCGATAATTTACAAAAGAGGTTTACCCAAAATTGTTATCAAGGATGAAATGGGCTTTATTTCGATCTTA 114800

 D  N  N  V  S  K  L  T  D  T  V  N  G  N  S  F  H  I  C  T  T  I  D  Y  A  I  V  S  K  V  A  M  T    745
GACAACAATGTATCAAAACTTACTGACACTGTCAATGGTAACAGTTTTCATATCTGTACAACCATAGATTACGCCATAGTTTCTAAAGTTGCAATGACTG 114900

V  T  K  S  Q  G  L  S  I  Q  R  V  A  L  D  F  G  N  D  P  K  N  L  K  L  S  S  I  Y  V  G  M  S  R    779
TGACAAAAAGTCAAGGCTTATCTATACAAAGGGTTGCCTTAGATTTTGGAAATGATCCTAAAAATTTAAAGTTGAGTTCTATATATGTTGGTATGTCAAG 115000

  V  V  D  P  N  N  L  I  M  N  L  N  P  L  R  L  N  Y  E  N  D  N  I  I  A  S  H  I  V  K  A  L  K    812
AGTTGTTGATCCAAATAATTTAATTATGAATCTGAACCCTTTACGGTTGAACTATGAAAATGATAATATTATTGCTTCACATATTGTAAAGGCCTTAAAA 115100

 N  K  D  T  M  L  I  F  -                                                                              820
AATAAAGATACTATGCTTATTTTTTAAATGCCAGAGTTTGTGTTGATTGACGTATTTTTTTAAACTAAGTTGAAATTAAGTTTTTCATGTATCGATGTGC 115200
------

                                                  ------
TTTTTTCTGGTATGTTTTTAATGGTTGGCAATTGTAAAAATAAAGACATGAATCAATGATTTGGTATACACTATTTTATTGCGAAAATGATAAATACTCT 115300

GGGTCCGGAAGTTGAGGCTTTTAGAAATCTTTCATCTTCTGAATCTGCAAAGAAAAAAAATGCTTAGAATCTTTTGCACATCTAGATAAAGTAGGTAAAA 115400

------
AATAAATATAAACAAAATTTACCGTCTGCGTAGTTATTTGTTCCAGTTGAATGAGGACTGCAGTATTCTTATGCCAGGTTTTTCAAAGACGCAGTCTAAA 115500

AGTAAAGGGAAAAACTTTTCCGGAGTATACAAATGTACTCCGGATTGCGCAGGCGTGTTTATTCTACATTTAAAATTGATGGCAGTTAGCGGTGTCAGGT 115600

GGTATGTATGGTGTCTATTTTAAGTATGTCAGAAATCAACTAAAGTTCCTTTTCCACACACTACTGAAGGTTGTATGAATATCACAATTCACAATTTTAT 115700

AACAATCTAGAGTTGCTTTTTAAGGTATGGAAAAAGCCTATGAAATCATGCAAAATGCTGATAATTGTACATTCAGAAAGCTAAAACTACTTATTCTTAA 115800

TATCTATGACTTAACCTTAAAAATCTATGACAAAACCTATTGATTATTCTTAACAGCTATTATTAAAATTAAAAGTGATTAACTATTCTTAAAAACTATG 115900

ATTTTACCCCATTCCCCACCCCCACCCCCACCCTTCTTGATTATATCCCGGACCCCACTAAATGGATTGTCTTTGTATTATGTTGGTGTGTTGCCCACTA 116000

CACACTGAAACATAGTACAAGGGCTAACTCTTAATGGTTACAAGTATGCAGCTCAAAATCGTAATGGTTCTGATCTTTCTTGTACATGTTCAAGTTTTCG 116100

GCGGTTAAGATGCTTTAGAGATAGTTCAAGATTTTAGATTCAAAACATCACTTATCTCCACAGATCATTTCCAAGCAATTGTACAGAAAGATAGAATGTC 116200

AATGAGAACAGCCGTTCTCATTGTCTTCACAAAGTGAGGTTCAGATCATGACATAGTAGCCGTAAATGGCTACTGTGGTTTACATTATAAATGAGAACAG 116300

CCGTTCCCATAGTTTTAAAATGAGGTTTAAAACATACAGTAGCCATTACATGGTTACTGTGGTCCAAAACGCGGTATCCGAATCGCAATCCACAATACAC 116400

AGTCCCGTTGTTCAAAAAGTTATGCAGTTGGCGGTCTTAATCAGGTATGAGATGGTATGGCAATACACGTTGTAAAAATTCTGAGATTTGAAACAATGAT 116500

GTCACTGTTGTAGGATATCCTGTAGTTTTTGATGGTTTTTTGGTTGATGACTTACACAAGATCAAAAGTTACTGGTAATCCTTAGTGTTAGATCTTACAT 116600

GATATGACAAGAGAAAAAAAGCTATATATGGGAAGGGATAACCCATATCAAACAAGAAAAGTTAAAGAGAAGAAACTATGAGATGGGAGAACCTATAACA 116700

AATACACTTTTAAGATTGGGGTGGGTGAACAGGGCATAACTGTAGCAATAAAACCAGTGACCGACAGTTACATTCCTTCATAAAAACTAACGTGATAGCA 116800

AAAAAAATATACATAAATAAAGAAGATGGCTTTTCAAGTGATAATGGCAGTCTAGATACACTTATATTCGTGAGAGGTGTGTCAGTCCATATACAACTTT 116900

TCCTATTACCTCGTCTAAAAAAGGCAACCTAAAACTGGCTGTTTTACCTGGGATGATTTAAATATAGAATAAACATTCACCTACCTTGTTTTCTTCTCTG 117000

AAGAGTAGTAGTGTTGACTGCATTGTCTCCGTGTATCTTGATCCAATAGAGCTTTATCTAATAAATACAAAAAAGGAAAAATCCGTATATTCCATAACTT 117100

GTTAAGTACAATTTCCATAAAAATTGAAAAAATATATCGTTTAGATGTTATACAAAAGTTTATCGTTTTAACTAATCATTGTCTATTTTCAGACATCGAT 117200
```

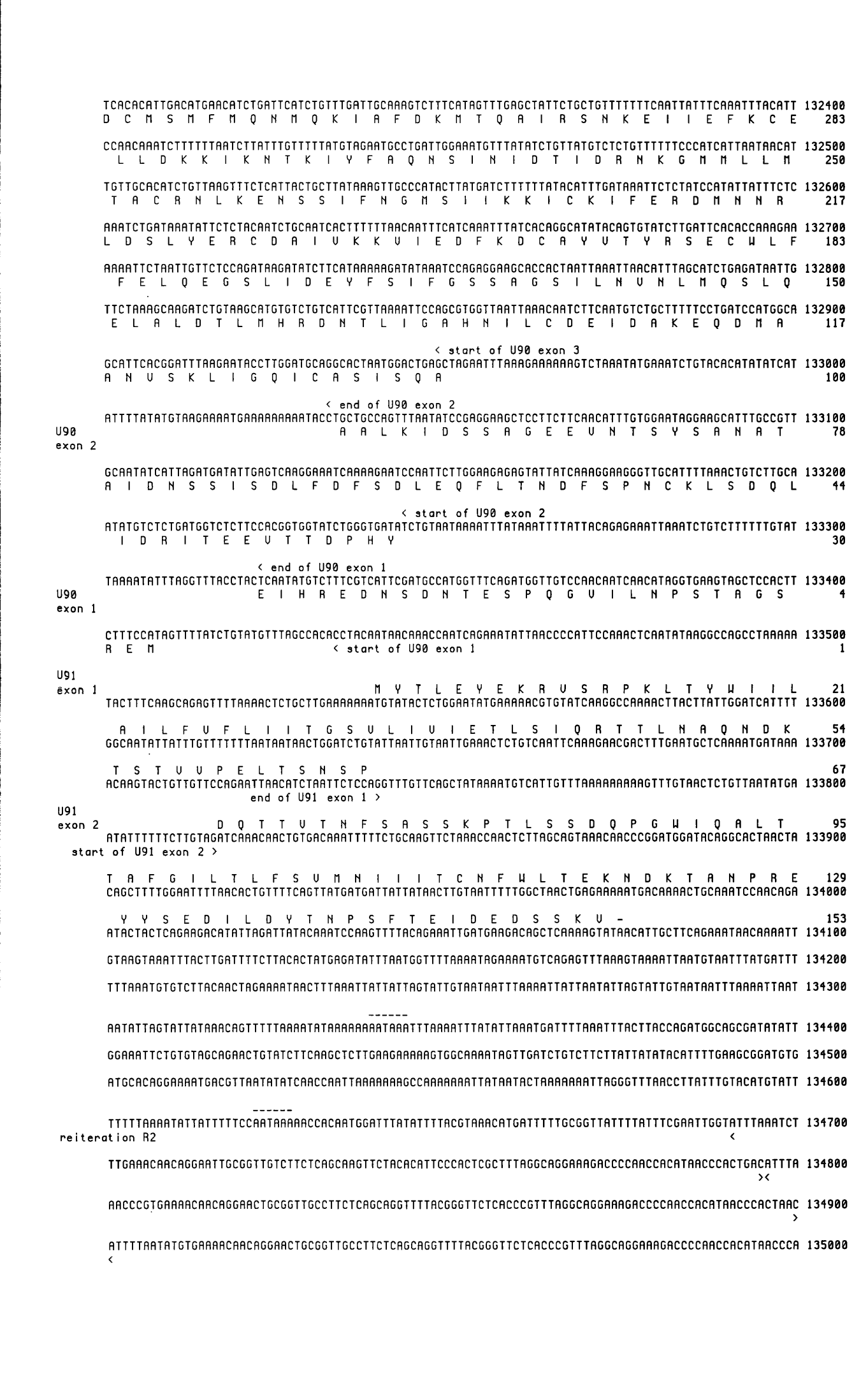


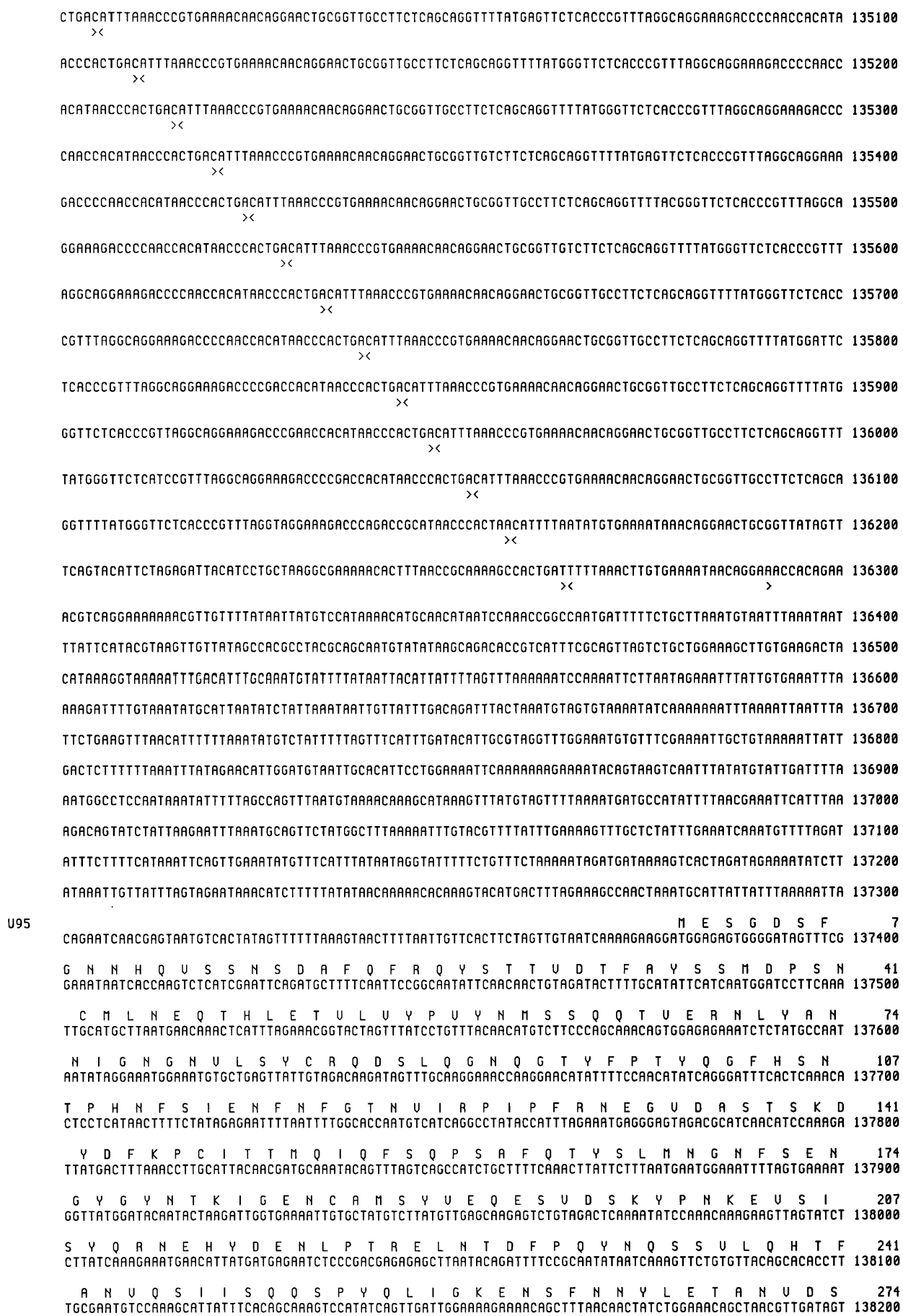
```
TCTCTGTCACGCTGAACCAATTTGTATACATTTTGTAATTGTACAATGTCAGAATCTGATAAGTTTAAAAACTTAAGCCACTTCACATTGATTTGAAACG 120100
E R D R Q V L K Y V N Q L Q V I D S D S L N L F K L W K V N I Q F T    27

TTTCATGTTGTTCTTCAAGAGACATTTTTATTGAGTTGTCATCTGAAATGTTTTCCAGCATCCACTGTAGTAATGCCATTATTGAATGGGAGAAGCAGTG 120200
  E H Q E E L S M K I S N D D S I N E L M W Q L L A M                1

TTCATGGGCTTTTTATTAATGTTGCCTGAACACGTGACACATAAGTTGAGCGATCTATAAAGCATGCGTCAGTAAGTTCTAATGAGCAAATTCGAAAACT 120300
U82                                                  - Y T R I L L N S F E   237

CGCTTCGGAATGGAAGCACTTCATTATATTTTTCCATTTCTTGAATTAACGGGAAACTAAAGTTGTAATACCTAGAAAAATATCTCAAACAATTTAACAA 120400
  S R F P L V E N Y K E M E Q I L P F S F N Y Y R S F Y R L C N L L   204

AGCATCAAAAACATTAATTCCATGGTGATCAATTCTCACAATTCTCTGTGTGCGAGTCTCTTCATTTTCAAGGTAAATCAAAATTGACATGTTAAAATTT 120500
 A D F V N I G H H D I R V I R Q T R T E E N E L Y I L I S M N F N    171

ATCATATTTGATGGTATATTGAATTCAAATCCGATTATATTTTCTACAAAGATGTTGCCCGTGAATTTGAGTTTTTCATAAGAGTATGATTTACACACAT 120600
I M N S P I N F E F G I I N E V F I N G T F K L K E Y S Y S K C V N   137

TATCTCTACAGGTTAAAAGTGTGTTATCTGAAAGACATCTTTCAAAGTTATTTAAAAAGCCAAGCCAGTTGTGCCCTGATTCTGAATCTTTAATTAAGGA 120700
  D R C T L L T N D S L C R E F N N L F G L W N H G S E S D K I L S   104

AAGTAAGGTGTGTAGTTGATTAGGATTATTATACAGTAGGTAAAATGTTTCTAAAAAATCTTCATCAATTAAAACTTGATCTCTCCTTGAGATATAGGGA 120800
 L L T H L Q N P N N Y L L Y F T E L F D E D I L V Q D R R S I Y P    71

GTATGATATAATAAGCGAGAAATTATATCCGAATTTGGTATAGGTGCCTGATCAATAGGTTCCAGAAGCGACAGAGTTTTTGATTGGGTACAATTTAGAA 120900
T H Y L L R S I I D S N P I P A Q D I P E L L S L T K S Q T C N L I    37

TGTTTTTGATACATTCTTGTTCTAAATTAGAATAATAAGAGTTATTAAATAAAGCATATATTTGGTTTAGAATGGAGATAAGAAATATAAAAAAGATGTT 121000
  N K I C E Q E L N S Y Y S N N F L A Y I Q N L I S I L F I F F I N     4

AGTTTTCATGTTCCATAAAGTTGTATACAAAACACCAGGAACTTAAATATGGAAGACACATTTAAATGGAATCTAATAATTTACATACCGTTTTTGAAGT 121100
 T K M                                                              1

GAGCTTTCTAATTATTTCGTGATTATAGAAATTATCATTCTGTTAAATACTCTGAAACAGAAAATCTGGAATTCTTAACAGTAGGTTTTCCTTACCTGTA 121200

TTAGTAGAAGTTGTATTTTTAGAAGTCGAATCTGAGTCAAGATCTATAGAAAATATAAACTTTATATCATTCCAAAAGTTAAGCAATAAATATATAACGT 121300

                                                  ------
AATTCGAGTCAATCAGAGTACATAAGTTTTAAAATATACTATGTCTTTAATAGTTTACCATAATCGTCCCAGTAGATGTATCTATTGAAAGTATTATAGA 121400
U84                                      - T K L L K G Y D D W Y I Y R N F T N Y I   281

TGATTAAATGCTCTTTAAGTAATAAAGGTTGTGTTAACTTTGAGACATTTAGTAAAATGTCAGCAGGGTAACGAATTGAAAATTGGACACAGAGATGAGC 121500
  I L H E K L L L P Q T L K S V N L L I D A P Y R I S F Q V C L H A   258

AGCACAAAATAGATCTTTTTCAGTTGGGGCATTTATGTACAGAGAGTAGACACCACTATTTTTCCTGTGTTTGCTGTTGACATTCACAGGCATCATAGCT 121600
 A C F L D K E T P A N I Y L S Y V G S N K R H K S N V N V P M M A    225

TCTTTTCTAGTCAGTTCAGTATGTGGCACTGTTCCTCTGAGAAATCTGTTAATTGCTATGGCAGTTTTCTGTTCATGGAAAAAGATCTCAGAAACGGTTT 121700
E K R T L E T H P V T G R L F R N I A I A T K Q E H F F I E S V T Q    191

GTTCGCCTTTGTTAACAATCAGATTAGCTATTTCCATCAATTTTTCGTGAAGTGAGTTTACAAGTAATCCGTTACGTGAATAAACTTGAATTTGAACTAC 121800
  E G K N V I L N A I E M L K E H L S N V L L G N R S Y V Q I Q V V   158

AGTCGAATTATAAAATCGAAGTAATTTATAAACTTCCGATAAGAAATTTTTCGGCGAACATAAACCGCATGGATGCATAAAATCACAATTAAAATCTGTA 121900
 T S N Y F R L L K Y V E S L F N K P S C L G C P H M F D C N F D T    125

TAATTTTTTATTTCAGGAGAGTGGATAGGTTTGGAATCGTTAAAATTGCTGTGAACAATACCCTTAGATTTTGTTGCAAAGTTTGTCGTTGAGAGATTTT 122000
Y N K I E P S H I P K S D N F N S H V I G K S K T A F N T T S L N E    91

CTTTTACAGCCTTAAATATTTTTCTCAATTTAGAATGTTTATAGATGGCATTCGTATATTTGTCCGCCTTAAGTAACATTTCTGATTTTGAAATGTTTAT 122100
  K V A K F I K R L K S H K Y I A N T Y K D A K L L M E S K S I N I    58

TTCAGAATTTATTTTTTCAGGAATTGTACATTTTATGGCTTTTTCATTAAACGGTAACGTTTTTGAAAAACAGTTACATGTATATCGAGCATTGGCTTTA 122200
 E S N I K E P I T C K I A K E N F P L T K S F C N C T Y R A N A K    25

GTACACATGCAGCTTTTAGAAGTGAGTTTACCTTTGGCCTTACAAACAGTTTTACGTTTTTTGCCCCTCATATTCAAGTACTGAAGCTTTCAAAATTTTC 122300
T C M C S K S T L K G K A K C V T K R K K G R M                     1
U85                                                                -

AATGTTGATGAGTTCCTAGTTTACACAGAAGGGTTGAAGCTATTATAACGAAAAGCAGGACACTGGAGATTATTAAAATTATCCACGCCAAATCGTTGTT 122400
  H Q H T G L K C L L T S A I I V F L L V S S I I L I I W A L D N N   248

AATTAACTCTGACTGTTTAAGAGAATCATCACTTAATATTGTATCATTATAGGGGTTTTCAAATGGTTGAATAATCGAGGTAACATTCTCGTCTATCTCT 122500
 I L E S Q K L S D D S L I T D N Y P N E F P Q I I S T V N E D I E    215

CCGCCGTAAGACAGACAAGTGATCGGTTTGCCAAAGAACGTATTGCTGCCTTGTCTTTTAAAAAGAAAGCTGAGTGTCGTTGTCTGCGAGCCGTTGTAAT 122600
G G Y S L C T I P K G F F T N S G Q R K F L F S L T T T Q S G N Y N   181

TTTCACGCACTGTTGGACTGTCCTGTCTATATATCTTATCTAAAAATAAAATGGCTACTAATGGCTTTGGATACGACGTCACAGAACATGTTACATCCAA 122700
  E R V T P S D Q R Y I K D L F L I A V L P K P Y S T V S C T V D L   148

AAAATTTTGCAGATGGCGATAGTACAAAGCAATTATTGGCTTCATTGAAAGATGCACACAACTCATAGTAGACAATCGAGTCTTGGCGAAGAATGCACAT 122800
 F N Q L H R Y Y L A I I P K M S L H V C S M T S L R T K A F F A C    115

GTTAAGCAACCAGCATCGTTAATTGTTCGAGATTTAAATTTCAGCAATGCTGTGATATTGTTAGCATTGAAGTAAATAATGCTTTCTTTGTCACGATGTT 122900
T L C G A D N I T R S K F K L L A T I N N A N F Y I I S E K D R H K    81
```

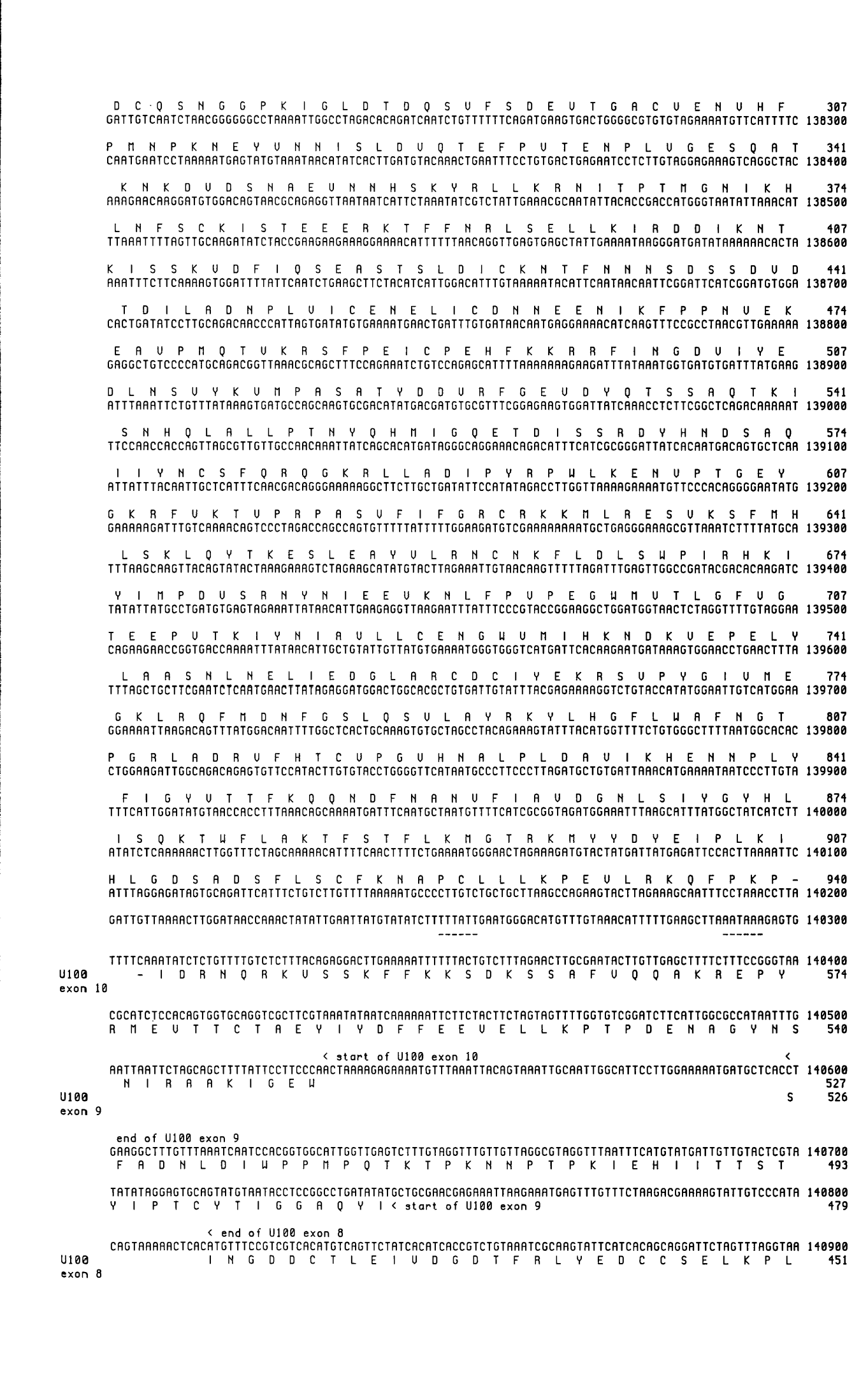


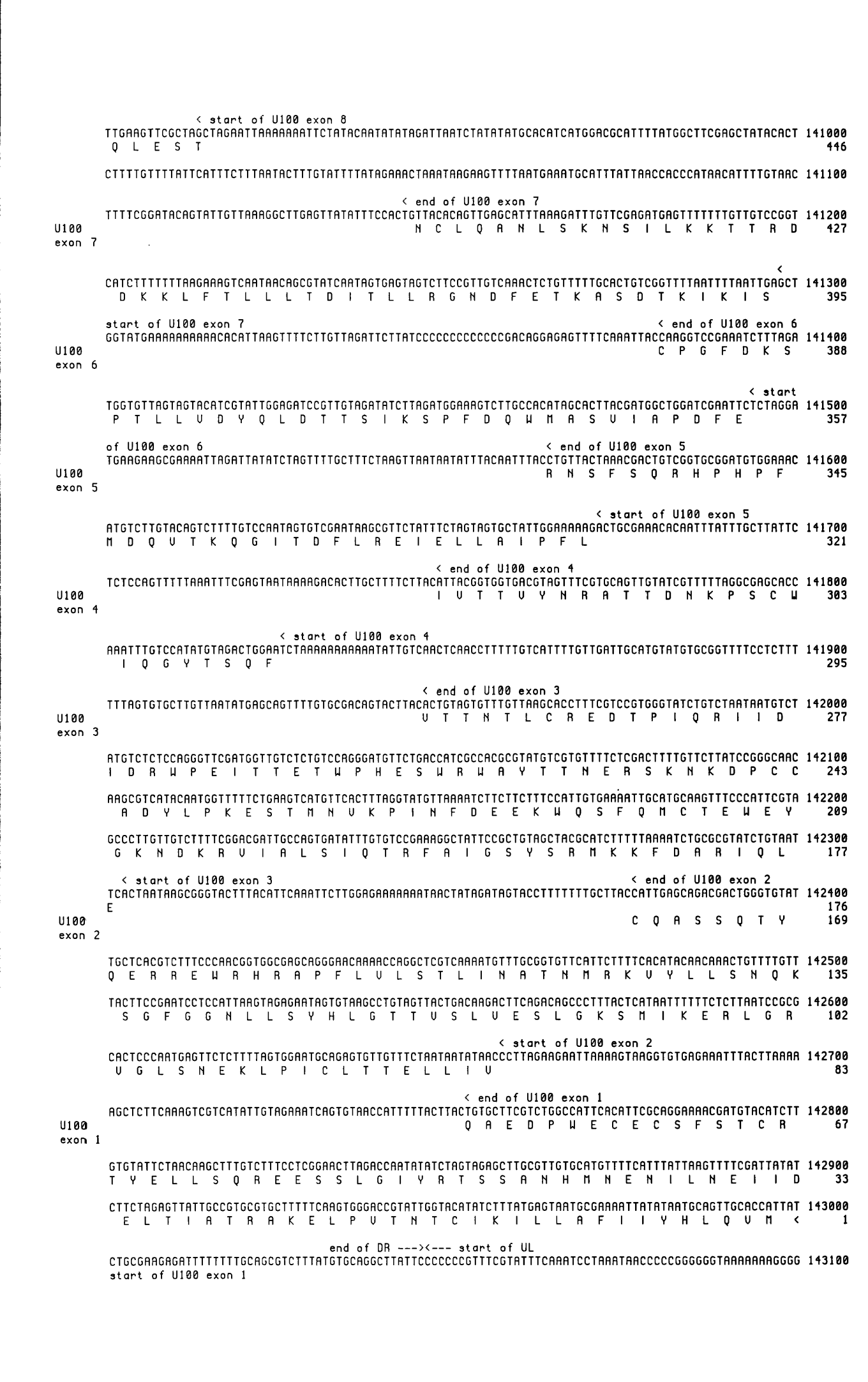

TGTGATTTGTTTTTCTTTTGATGTACTAGGTCTTGTTCTAGCTTTCATAAAAGTCTTTTCCTGTGAAGACTTGTTATATTTACATTTGTCTACATCTGAA 125900
T I Q K E K S T S P R T R A K M F T K E Q S S K N Y K C K D V D S 346
GTCGAAGATAGACTGTCTTTAGCCAGGGGAACTATATCAGCATTATAAATTCTTTTAATTAGATCCATCTTATGAGAACTTGGAATATATGACAAGCTTA 126000
T S S L S D K A L P V I D A N Y I R K I L D M K H S S P I Y S L S V 312
CAACATCCTCATTTTGTGGTGTTTTGTCTTTGTCTGTCTTTTTCCTGGCAGGTTTTTCCAAATTGTTCACGTTTGAGATTAATTTGGTTTCCTTGTGATT 126100
V D E N Q P T K D K D T K K R A P K E L N N V N S I L K T E K H N 279
TGTCTGAATTGGTGTATTTTGACTGTATTCACTTTTCTTTGATAATGTAAGACCTGATGAAATACATTCTTTTTGAAGGTTTTGACTATCTGTGATCATA 126200
T Q I P T N Q S Y E S K K S L T L G S S I C E K Q L N Q S D T I M 246
GAGCTGCTTGGAATTTGTGAAAAAGCCATGCTGTAATCTCTGCTTTGAAAGCCATAGTTAGGACAAAGTTGTTGATAATTGTCTCTGTAAGGATTATGTC 126300
S S S P I Q S F A M S Y D R S Q F G Y N P C L Q Q Y N D R Y P N H R 212
TAAAATCTGAATTGTATGAGTCATATGACTGATCTCGTTGCATATATTCCGGAAAGGTAAATTCATTTCTGTTGGAAAAGTCATGAAATTTAGCATAGCC 126400
F D S N Y S D Y S Q D R Q M Y E P F T F E N R N S F D H F K A Y G 179
ACGTCCACGTCTGTTGTAAGGTCTGTGAAAATGTTTACCATAATGATTTTGGTATTTTGAGTTAGGGTTACAATAATTTCTACTGTCATATTTTCCTCTG 126500
R G R R N Y P R H F H K G Y H N Q Y K S N P N C Y N R S D Y K G R 146
TTGTTGCGACCCCTGTAATGATGTTGTCGATAGGTTGTATCGAACGAATTGTGACAAAGTTCTCTATTTTTATAGTCCAAATTGTTAAAAATATCAACTG 126600
N N R G R Y H H Q R Y T T D F S N H C L E R N K Y D L N N F I D V S 112
AAGATTGGGGTTGATATGGTAAAGAGATAAAATTTTCTCCACTTGCTTGAGGATAATTATAATACAATGATTGCAGTCCTTGATGTGGCAATTCTGATGC 126700
S Q P Q Y P L S I F N E G S A Q P Y N Y Y L S Q L G Q H P L E S A 79
AAATCGAAATTGTCCAGTTTCAGTTGTATTTGGAGAAGATTCAGAAGTAAAAGTACTTCTTGATTTAAGAGGAATAGCACGTCGCGGATCTTGAAGAGTT 126800
F R F Q G T E T T N P S S E S T F T S R S K L P I A R R P D Q L T 46
TTTGATGTGACTATTTCTTGGTTGTTAAAATTTGAATAAGTTATTTCATTTTCTGGTTTTGTATGATCGTGAATTTCATTATATAAACCCGAAAGAGTCA 126900
K S T V I E Q N N F N S Y T I E N E P K T H D H I E N Y L G S L T M 12
TGGCTTCACGAACCACGTCAGCTGTAGTGCTCATGGGTGCATGCAATGGTCTTTTTTCTGAAAATGAAAAATAAAAATTGTTTATTTTAAAATGTTATTA 127000
A E R V V D A T T S M 1
AATTAGCGATAATACTAAGATGATATATACCTTTAAATAGAGTGGAAATGTACTTTTAATTGTCTAGAAAATATAACGCAATATGTTAGTTAAAAATGTA 127100
ACTTGTTAGATTGATATTTCATGGGCCCTGAAACCTGCATATGAGCTAATGTGTGTACAAATTTGTGACAATAAATATTAAAGTGGAAACATTCCAAATG 127200
TAGCAATTTACATGTAAACGGATATGTAAATGTAATAAAGATGTGTATGGTTATGTCAAAAGAGCTAATTTGCAAACATGCAGTTATTAATTGTAATGGA 127300
TGTACAAATGTTGATACATGTTTGTGTCTGTTGCTGTGTATATGAGTTTATGTTAAATGTTTATGTAAATATTATATGTTGGTACATATCAATGTATTTA 127400
TGTTGATAAGCAGATGTTAGTATGTATTAGTATCAATGGATTGTAGTATGCTACTGCATGCTATTAAACAAGTAGAGTGTTTACTTGTATTAGATTATAT 127500
TTACATGACATTGTCATGTAAATATATAGTTGCATACATTTTTCATGTAACAACACATATGTGTTCATGTACATATTACATAATATATGCTAGTAAATGA 127600
reiteration R1
TTACATGCACTAGTAAATATTTGCACATGCTAATGTGTTCATGTGGGTATATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCAT 127700
ATATTTGCACATACTAATGTGTTCATGTGGGTATATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGCATATATTTGCACATACTA 127800
ATGTGTTCATGTGGGTATATGTACATATTACATAATATATGCTAGTAAATGATTACATGCACTAGTAAATATTTGCACATGCTAATACATAACAGTACAT 127900
GTTAATAACTTTTGATCTATGCCGCTACATGTTAGTACATGCTATACTGATAATGCGTGGAAGTTTATTATTTGGATTTCTGTAAGGTCATGTGTATTCT 128000
ATGAATGAGTATATAAATGGGTATGTTAATATATGTTAGTATGGTTGCAAGTAAAGATATGTTAATGTGTATTAGTAAGTACGTAATGTATGTGAATGTA 128100
TTCAATTTTCAAAGAAATGATATAGAAATTTCTAAACATGTTATTGTGGAAAATGTAGATCTTGTGATCTATTATAATAATATAAACATAAATTAAATTT 128200
TTAAGTTATAGCTACCAATCTTTGCCTATTTTTCTTGGGGGGCATACTTTTTATATGATTTTATATGTACGGAAATATATGTACGTAGAATAAAGAATAA 128300
AATACTTGAAATCAAACCCTGTTATAAGTAAATTTACTTAGTATCATATATGTATAAATTTTGATAATTTATGAAGATGGCATAAATGTTGAAAATATTC 128400
ATGTGTTTGTTATAAGAATTTCAATGTTACTAAAATAAAACCAATCACTAATTATCTTTTATATGCAAAAATGTGTCTTTGTATATATTCAAATTAAAAT 128500
TTAAAACTTATGTTTTAATCATGTGAAATACAGATGTATGAATATGTAGCATGGAAACTTCAAAAATATAAATCATATAGACCTATTGCTTTAACTATAT 128600
TTAAGATTTACATATTTAAATATTACATAATAAAAATTTTAATATTTAAATAAAAATATGTATTATAAATATGTTATTTACTAGTACAGCTTTATAAATT 128700
AAAGAACATATGTATTTAAAGTTTAAAAACTAATGAAATGTGCTTCTATGAGTTAATTGTAAATGTCAATTTTGTGATTTTAATATTTTTCTAAATAATG 128800
AGATTCAATGTTGTTAATTAGATGTGTGCAGTTTGAGAACATGTTTATATTTACAAAAATTTGTATATATTTAAAAATATGTTAATCAGCATAGCCTTAT 128900
TTATAATGATTAATTTTAAATTCTATGTAATTAAATCTTATGGACAAAATAAATTTCAAACCATTTATTTTGAAAAACATAAATGCAGTTTAGAACTGAT 129000
ACAATTTTTTTTTTCAAAAGGGGTACATACTTAAAAGATTTTTTACAGTGGTTATAAAAGTATTAGAAAAGGAATTAAGATATATATGAAAAAATTTCCA 129100
TCCTGCTTATGGAACTATAATATATCAAAGCAATTTATTTATGAAAATTCCAAGTCTATGAAAGTAAAGATACCCATTCAAGTGAAAGGCTATTATAGTT 129200
AGGAGGGCTAGCAATATAAGAAAACTTATATTCAATGTATTTTTTTTTAAGCTATGAAAGTAAAGATACCTATTCAAGTGAAAGGCTATTATAGTTAGGA 129300
GGGCTAGCAATATAAGAAAACTTATATTCAATGTATTTTTTAAGCTATGAGATAAATTTGATAGGTCTAAAACATTTTATTTATAAGAAAAAAATGAAAA 129400

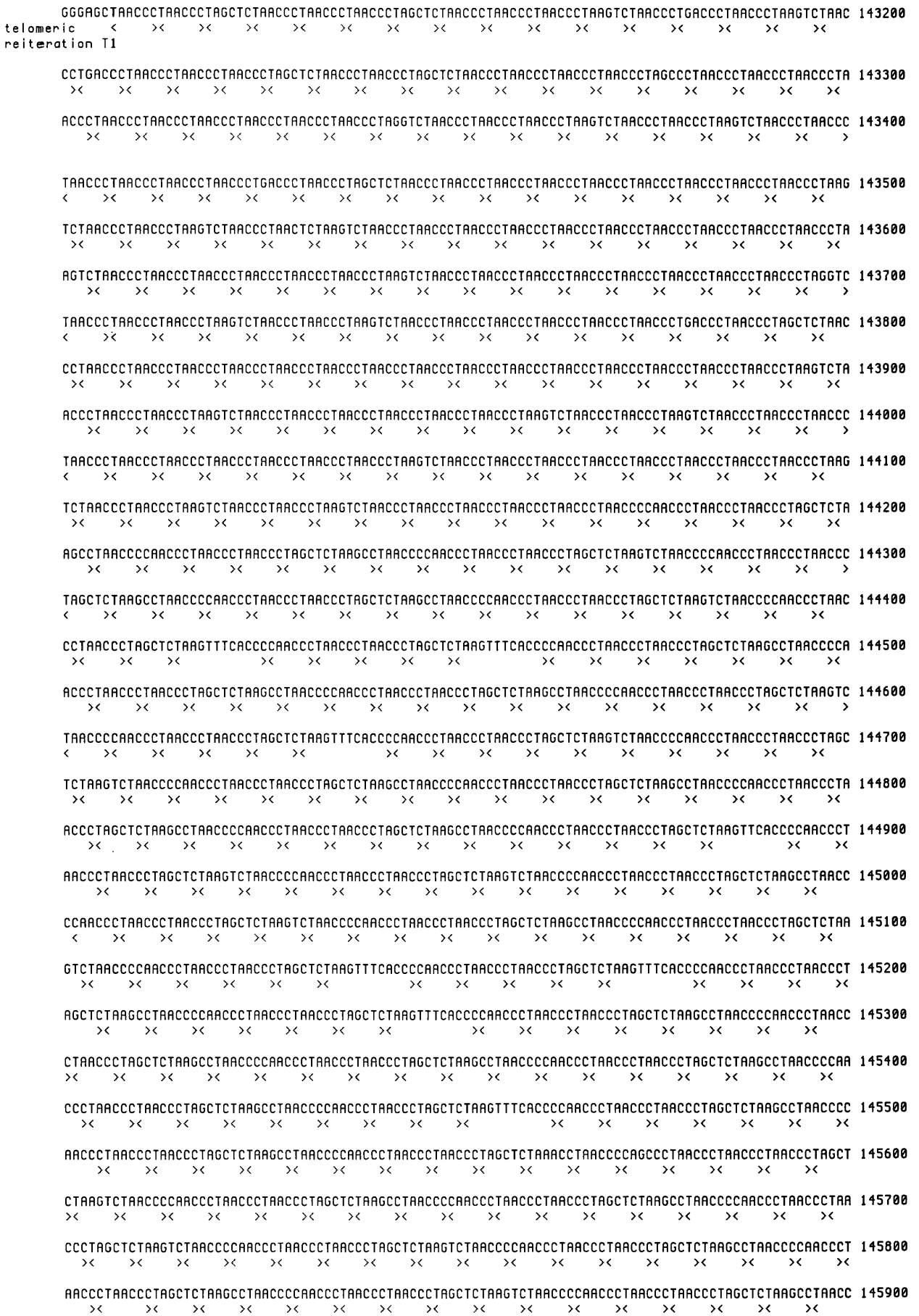


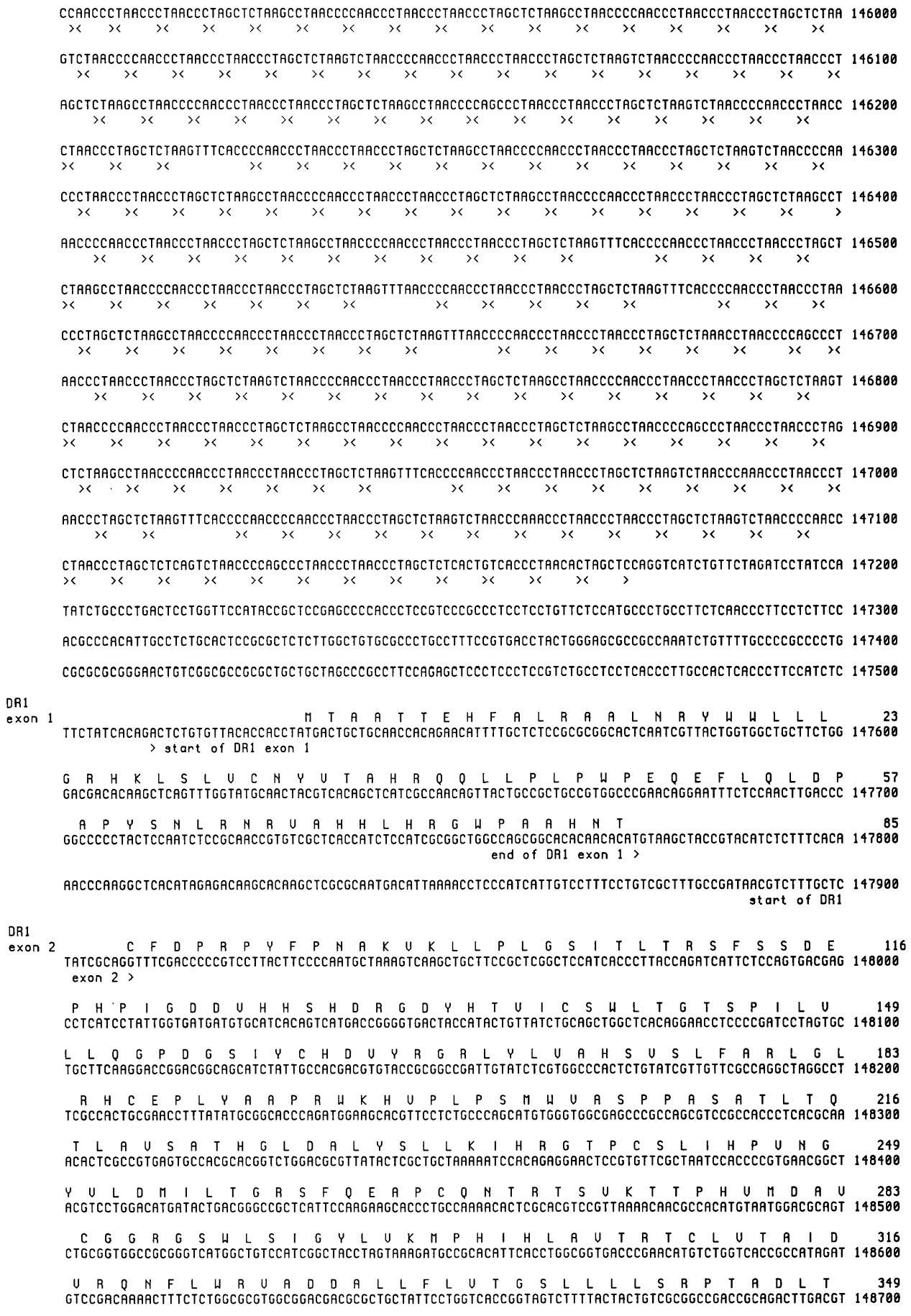







< start of U100 exon 8
TTGAAGTTCGCTAGCTAGAATTAAAAAAATTCTATACAATATATAGATTAATCTATATATGCACATCATGGACGCATTTTATGGCTTCGAGCTATACACT 141000
Q L E S T 446
CTTTTGTTTTATTCATTTCTTTAATACTTTGTATTTTATAGAAACTAAATAAGAAGTTTTAATGAAATGCATTTATTAACCACCCATAACATTTTGTAAC 141100
< end of U100 exon 7
TTTTCGGATACAGTATTGTTAAAGGCTTGAGTTATATTTCCACTGTTACACAGTTGAGCATTTAAAGATTTGTTCGAGATGAGTTTTTTTGTTGTCCGGT 141200
U100 N C L Q A N L S K N S I L K K T T R D 427
exon 7
<
CATCTTTTTTTAAGAAAGTCAATAACAGCGTATCAATAGTGAGTAGTCTTCCGTTGTCAAACTCTGTTTTTGCACTGTCGGTTTTAATTTTAATTGAGCT 141300
D K K L F T L L L T D I T L L R G N D F E T K A S D T K I K I S 395
start of U100 exon 7 < end of U100 exon 6
GGTATGAAAAAAAAAACACATTAAGTTTTCTTGTTAGATTCTTATCCCCCCCCCCCCCGACAGGAGAGTTTTCAAATTACCAAGGTCCGAAATCTTTAGA 141400
U100 C P G F D K S 388
exon 6
< start
TGGTGTTAGTAGTACATCGTATTGGAGATCCGTTGTAGATATCTTAGATGGAAAGTCTTGCCACATAGCACTTACGATGGCTGGATCGAATTCTCTAGGA 141500
P T L L V D Y Q L D T T S I K S P F D Q W M A S V I A P D F E 357
of U100 exon 6 < end of U100 exon 5
TGAAGAAGCGAAAATTAGATTATATCTAGTTTTGCTTTCTAAGTTAATAATATTTACAATTTACCTGTTACTAAACGACTGTCGGTGCGGATGTGGAAAC 141600
U100 R N S F S Q R H P H P F 345
exon 5
< start of U100 exon 5
ATGTCTTGTACAGTCTTTTGTCCAATAGTGTCGAATAAGCGTTCTATTTCTAGTAGTGCTATTGGAAAAAGACTGCGAAACACAATTTATTTGCTTATTC 141700
M D Q V T K Q G I T D F L R E I E L L A I P F L 321
< end of U100 exon 4
TCTCCAGTTTTTAAATTTCGAGTAATAAAAGACACTTGCTTTTCTTACATTACGGTGGTGACGTAGTTTCGTGCAGTTGTATCGTTTTTAGGCGAGCACC 141800
U100 I V T T V Y N R A T T D N K P S C W 303
exon 4
< start of U100 exon 4
AAATTTGTCCATATGTAGACTGGAATCTAAAAAAAAAAATATTGTCAACTCAACCTTTTTGTCATTTTGTTGATTGCATGTATGTGCGGTTTTCCTCTTT 141900
I Q G Y T S Q F 295
< end of U100 exon 3
TTTAGTGTGCTTGTTAATATGAGCAGTTTTGTGCGACAGTACTTACACTGTAGTGTTTGTTAAGCACCTTTCGTCCGTGGGTATCTGTCTAATAATGTCT 142000
U100 V T T N T L C R E D T P I Q R I I D 277
exon 3
ATGTCTCTCCAGGGTTCGATGGTTGTCTCTGTCCAGGGATGTTCTGACCATCGCCACGCGTATGTCGTGTTTTCTCGACTTTTGTTCTTATCCGGGCAAC 142100
I D R W P E I T T E T W P H E S W R W A Y T T N E R S K N K D P C C 243
AAGCGTCATACAATGGTTTTTCTGAAGTCATGTTCACTTTAGGTATGTTAAAATCTTCTTCTTTCCATTGTGAAAATTGCATGCAAGTTTCCCATTCGTA 142200
A D Y L P K E S T M N V K P I N F D E E K W Q S F Q M C T E W E Y 209
GCCCTTGTTGTCTTTTCGGACGATTGCCAGTGATATTTGTGTCCGAAAGGCTATTCCGCTGTAGCTACGCATCTTTTTAAAATCTGCGCGTATCTGTAAT 142300
G K N D K R V I A L S I Q T R F A I G S Y S R M K K F D A R I Q L 177
< start of U100 exon 3 < end of U100 exon 2
TCACTAATAAGCGGGTACTTTACATTCAAATTCTTGGAGAAAAAAATAACTATAGATAGTACCTTTTTTTGCTTACCATTGAGCAGACGACTGGGTGTAT 142400
E 176
U100 C Q A S S Q T Y 169
exon 2
TGCTCACGTCTTTCCCAACGGTGGCGAGCAGGGAACAAAACCAGGCTCGTCAAAATGTTTGCGGTGTTCATTCTTTTCACATACAACAAACTGTTTTGTT 142500
Q E R R E W R H R A P F L V L S T L I N A T N M R K V Y L L S N Q K 135
TACTTCCGAATCCTCCATTAAGTAGAGAATAGTGTAAGCCTGTAGTTACTGACAAGACTTCAGACAGCCCTTTACTCATAATTTTTTCTCTTAATCCGCG 142600
S G F G G N L L S Y H L G T T V S L V E S L G K S M I K E R L G R 102
< start of U100 exon 2
CACTCCCAATGAGTTCTCTTTTAGTGGAATGCAGAGTGTTGTTTCTAATAATATAACCCTTAGAAGAATTAAAAGTAAGGTGTGAGAAATTTACTTAAAA 142700
V G L S N E K L P I C L T T E L L I V 83
< end of U100 exon 1
AGCTCTTCAAAGTCGTCATATTGTAGAAATCAGTGTAACCATTTTTACTTACTGTGCTTCGTCTGGCCATTCACATTCGCAGGAAAACGATGTACATCTT 142800
U100 Q A E D P W E C E C S F S T C R 67
exon 1
GTGTATTCTAACAAGCTTTGTCTTTCCTCGGAACTTAGACCAATATATCTAGTAGAGCTTGCGTTGTGCATGTTTTCATTTATTAAGTTTTCGATTATAT 142900
T Y E L L S Q R E E S S L G I Y R T S S A N H M N E N I L N E I I D 33
CTTCTAGAGTTATTGCCGTGCGTGCTTTTTCAAGTGGGACCGTATTGGTACATATCTTTATGAGTAATGCGAAAATTATATAATGCAGTTGCACCATTAT 143000
E L T I A T R A K E L P V T N T C I K I L L A F I I Y H L Q V M < 1
end of DR ---><--- start of UL
CTGCGAAGAGATTTTTTTTGCAGCGTCTTTATGTGCAGGCTTATTCCCCCCCGTTTCGTATTTCAAATCCTAAATAACCCCCGGGGGGTAAAAAAAGGGG 143100
start of U100 exon 1


telomeric reiteration T1



```
S  W  S  C  L  Q  Q  E  P  V  W  R  N  C  L  D  T  R  G  E  Q  D  E  T  E  D  Q  E  M  K  Q  S  T  S    383
CTTGGTCATGTTTACAGCAAGAACCTGTGTGGAGGAACTGTCTAGATACGCGCGGAGAACAGGATGAGACAGAAGACCAAGAGATGAAACAAAGCACAAG 148800

  K  K  Q  N  E  N  K  K  L  N  T  S  K  K  H  T  R  V  S  S  A  I  P  T  F  P  L  S  L  R  E  T  P    416
CAAAAAGCAAAATGAGAATAAAAAACTCAACACCTCAAAAAAACACACCCGCGTATCGTCGGCAATTCCGACCTTTCCCCTGAGTCTCCGAGAAACGCCG 148900

 P  E  A  R  S  P  A  V  L  A  A  A  T  Q  S  H  K  T  R  A  I  S  T  H  N  A  T  T  T  I  R  I  P     449
CCAGAAGCCAGGAGCCCAGCCGTCCTCGCCGCCGCCACCCAGTCTCACAAAACTCGAGCGATCTCGACGCATAATGCCACGACAACAATAAGAATACCGC 149000

R  L  P  S  Y  L  L  E  A  R  L  L  S  V  T  A  I  L  K  D  T  K  K  K  K  T  Q  P  Q  A  -            479
GCCTTCCCAGTTACCTGCTGGAAGCGCGTCTGTTGTCCGTGACAGCTATCCTGAAAGACACAAAGAAAAAAAAAAACCCAGCCTCAGGCGTAGCAGCTGCG 149100

ACGCTCAGCGCGGTGTCTGAAAGCTCGCCAAGGTCTCGCGTAAAAGAACAGATGTGAACTTCAGATGTACCAACCAAATAATACGGGTTCCGCTATAAAA 149200

AGTGCACCTCTATTCCCGTTCTTATCCCCGTTCTAACTCTTCCTTGTATCATACCTTGCATGTTAACCGGATCCCGTGGATCTTACACACATACACACAC 149300

ACACAAACTTGGTGAGGTAAACACAGAAATCTCACTAACTCATAATCCCCTACACGCTTACCACCACCTAAAAATGGTTATGACCAAAACTGGCAATAGT 149400

CTATCTTCTTTTTCTTTCCATTCACAGCCAATGTGCAGTACTCGTGGGTCCACAACAACGAAAGAGACTGTAGAGACACTTCCTTTAAGTAGACCTTAGA 149500

DR6
exon 1                                                                M  S  A  E  M  L  R  A  V  Q  L  Q  P  R  R      15
GACACACCAAATACAACCACAACCAAAAAAAAAACAGAAAACAACACAAAGCCAATGAGTGCAGAAATGCTCCGCGCTGTTCAGCTCCAGCCAAGACGCC 149600

R  G  H  S  S  S  P  T  S  P  P  L  E  G  E  P  S  P  K  R  L  Q  S  S  N  S  H  Q  G  R  R  G  R  P     49
GGGGACATTCCTCATCTCCCACTTCCCCTCCACTCGAAGGAGAGCCCAGTCCCAAGAGACTCCAATCGAGCAACAGTCACCAAGGGCGTAGAGGCAGACC 149700

  K  P  R  A  K  T  W  S  E  A  L  S  H  R  S  F  L  N  I  Y  A  W  L  S  L  S  R  G  S  P  R  K  V      82
TAAACCCAGAGCTAAAACATGGAGCGAAGCTTTATCCCACCGGTCCTTCCTCAACATTTACGCGTGGCTGTCTTTGAGTCGAGGGTCTCCGCGAAAAGTG 149800

 Y  G  Y  A  F  R  H  R  G  E  L  V  A  L  P  W  P  P  N  W  S  L  E  L  H  H  D  P  Y  R  D  A  R     115
TACGGATATGCCTTCAGGCACAGAGGAGAACTCGTAGCATTGCCATGGCCGCCTAACTGGAGCCTGGAACTTCACCACGATCCCTATCGAGACGCCAGAG 149900

A  Q  T  V  W  S  H  R  W  G  W  P  A  T  H  V  T  A  R  T  V  R  D  C                                139
CACAAACCGTTTGGAGTCACCGCTGGGGATGGCCTGCAACACACGTGACAGCTCGCACGGTGCGGGACTGCGGTGAGTGTAAGCAGTGTGACACATTGTT 150000
                                                    end of DR6 exon 1 >

ATCGCAATTGTCTTACCCGATTAACTTTTTATTAATGTATTAAGCACTCTTTCTTCACGTGTGACTGTTGTGTTTTTTGTTGTTATCTACATCCCGGCAG 150100
                                                                      start of DR6 exon 2

DR6
exon 2 A  L  D  T  H  M  Y  V  C  C  G  R  G  E  K  L  Q  P  V  G  Y  V  R  N  R  A  A  P  S  D  L  N  S  L    173
CCCTCGACACGCATATGTACGTGTGCTGCGGACGCGGAGAAAAGTTGCAGCCCGTCGGATACGTACGCAACAGAGCCGCGCCTTCAGACCTGAACTCGTT 150200
>
  R  V  L  L  I  A  R  D  G  A  M  Y  V  H  H  M  R  T  A  R  L  C  R  L  A  S  S  V  T  E  F  A  R    206
ACGCGTCCTCCTCATAGCCAGGGACGGAGCAATGTATGTGCATCACATGAGAACGGCGCGACTGTGCCGCCTAGCCAGCAGTGTGACCGAATTCGCGCGA 150300

 R  G  L  Q  R  E  S  E  V  Y  E  D  D  V  S  L  P  D  R  R  V  G  S  A  T  A  I  H  L  F  D  V  I     239
CGAGGGCTGCAGCGAGAATCCGAGGTTTATGAAGATGATGTTTCCTTGCCAGACCGTCGAGTAGGTTCGGCAACGGCCATTCACCTGTTTGACGTAATTA 150400

T  Q  A  A  D  V  H  D  L  L  T  V  A  G  L  C  Q  T  H  T  G  V  S  C  Q  L  W  Y  T  D  H  D  P  H    273
CCCAGGCAGCCGATGTCCACGACCTACTCACCGTGGCCGGACTGTGTCAGACTCACACCGGCGTCAGCTGCCAACTGTGGTATACAGACCACGATCCCCA 150500

  T  V  A  G  A  A  R  F  T  L  T  V  A  R  Q  Q  Y  R  L  W  P  N  A  R  R  K  L  L  Q  H  L  H  P    306
CACCGTCGCTGGGGCGGCACGCTTCACACTGACGGTCGCACGGCAGCAGTATCGATTGTGGCCAAACGCACGACGCAAACTGCTGCAGCACCTACATCCG 150600

 D  H  P  L  G  L  W  L  L  C  A  V  L  T  Y  D  A  K  E  T  N  R  A  V  P  P  V  T  P  G  A  E  T     339
GACCACCCACTTGGGCTGTGGCTGTTGTGTGCCGTGCTCACGTACGATGCAAAAGAGACGAATCGCGCAGTGCCACCCGTAACGCCAGGGGCCGAAACCG 150700

V  W  V  I  V  T  G  R  G  A  I  L  G  F  W  P  E  S  A  K  M  C  R  L  A  S  S  M  K  G  L  W  K  N    373
TGTGGGTGATAGTTACTGGCAGGGGTGCCATTCTAGGATTCTGGCCAGAGAGCGCCAAAATGTGCAGATTGGCCTCGTCTATGAAAGGACTCTGGAAAAA 150800

  G  A  R  A  L  K  G  H  W  T  Y  A  A  P  G  R  H  R  A  G  E  A  W  P  L  C  A  H  Y  Q  S  P  R    406
CGGAGCGCGGGCGCTAAAAGGTCACTGGACATACGCAGCACCCGGCCGGCATAGAGCGGGAGAGGCCTGGCCTTTGTGTGCACACTACCAATCTCCTAGA 150900

 -
TAGAACAAAATTAAAAAGATTAAAAAAAAAAAGAAAAAAAGTACAAGAGTGTTATCGCGAAACAGCGTGTCAAAAAAAAAAACAATCCACATACTCTAGA 151000
       ------   ------

ACAAACTGTACCCAAAAATAAGTCCGTGTGCAAAACTGGGAAAAAAAAAATCACCTTCCTCGTTGCCACTAGAGGGAGTACCGAAAGTGTAGGCAAGAAG 151100

GCCACGCTGTAAATGACTGTCAGCGTTTGGCGCTGAAAACATTGCTGTTCTTGCTGGCTCAAGCACAATCACGTGATTAAGATTCCTTTCGTTTTCAAAG 151200

TGTGCCCGGGAGGCAGACATGCCCTTTCTCGTGAGACATTATGAGATTTGCCTGCCAGAGAACCACGTGACTTGGACTTACTTTCGTTTTCTAAACGTGC 151300

CCTCTAGGCATGAATGCTCTTTAGCGTTAGCCATGAGGCTAGCGTGATCCTGTATAGTACATAAGTTTCTAAGAATATGTTTTTAACAATAATCATGTCC 151400

CAAAAAGTCGCGAGTGACTAAAATTCTCTGTAAATGAAGGCAAATTAAACAGGATACAGACAGTTGTGGCAGTGGTCCGTTTCGTCTTTCTGTGTTTTCC 151500

TTACGCGGCTGACGAGGTAAAGTGTCTCAGTCCATATTGTTGTCTGTGCCACCGTAGTTAGCGGTGGCATACTAAAAACTCCGATAGATGCAGAACAATA 151600

ACACCGAAAACCACGCTGTGGAACCAGACCACACTTTATAAACAAAACGGCCTTATCACCTGGAAAAAAAAACTAAAAATAAGGCAATGATACACCTGAC 151700

TTTCCATTGGAAACCTGCCGTAACCCTGACCACAAATCCCATGCTAAATCCCCTGAAACACTGCCAAACGTCGCTACAAGGTTTTTCCGGGATCGAGCCG 151800

CAGCAAGCTTAAACTGAGGTCACACACGACTTTAATTACGGCAACGCACAGCTGTAAGCTGCAGGAAAGATACTATCGTAAGCAAATGTAGTCCTACAAT 151900
```

```
        CAAGCGAGGTTGTAGACGTTACCTACAATGAACTACACCTCTAAGCATAACCTGTCGGGCACAGTGAGACACGCAGCCGTAAATTCAAAACTCAACCCAA 152000

        ACCGAAGTCTAAGTCTCACCCTAATCGTAACAGTAACCCTACAACTCTAATCCTAGTCCGTAACCGTAACCCCAATCCTAGCCCTTAGCCCTAACCCTAG 152100
telomeric                                                                 <    >      <    ><     ><    ><
reiteration T2

        CCCTAACCCTAGCTCTAACCTTAGCTCTAACTCTGACCCTAGGCCTAACCCTAAGCCTAACCCTAACCGTAGCTCTAAGTTTAACCCTAACCCTAACCCT 152200
          ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><          ><    ><    ><

        AACCATGACCCTGACCCTAACCCTAGGGCTGCGGCCCTAACCCTAGCCCTAACCCTAATCCTAATCCTAATCCTAGCCCTAACCCTAGGGCTGCGGCCCT 152300
            ><    ><    ><    ><      ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><      ><

        AACCCTAGCCCTAACCCTAACCCTAGCCCTAGGGCTGCGGCCCTAAACCTAACCCTAACCCTAGAGCTGCGGCCCTAACCCTAACCCTAGGGCTGCGGCC 152400
            ><    ><    ><    ><    ><      ><    ><    ><    ><    ><      ><    ><    ><    ><      ><

        CTAACCCTAACCCTAGGGCTGCGGCCCGAACCCTAGCCCTAACCCTAACCCTAGCCCTAGGGCTGCGGCCCGAACCCTAGCCCTAACCCTAACCCTAGCC 152500
        ><    ><    ><      ><    ><    ><    ><    ><    ><    ><    ><      ><    ><    ><    ><    ><

        CTAGGGCTGCGGCCCTAACCCTAACCCTAGCCCTAGGGCTGCGGCCCTAAACCTAACCCTAGCCCTAGGGCTGCGGCCCTAACCCTAACCCTAGGGCTGC 152600
        ><      ><    ><    ><    ><    ><    ><      ><    ><    ><    ><      ><    ><    ><    ><      ><

        GGCCCTAACCCTAACCCTAGGGCTGCGGCCCTAACCCTAACCCTAGGGCTGCGGCCCGAACCCTAACCCTAACCCTAACCCTAACCCTAACCCAAACCCT 152700
            ><    ><    ><      ><    ><    ><    ><      ><    ><    ><    ><    ><    ><    ><    ><    ><

        AACCCTAGGGCTGCGGCCCTAACCCTAACCCTAGGGCTGCGGCCCGAACCCTAACCCTAACCCTAACCCTAGGGCTGCGGCCCTAACCCTAACCCTAGGG 152800
            ><      ><    ><    ><    ><      ><    ><    ><    ><    ><    ><      ><    ><    ><    ><

        CTGCGGCCCTAACCCTAACCCTAACTCTAGGGCTGCGGCCCTAACCCTAACCCTAACCCTAACCCTAGGGCTGCGGCCCGAACCCTAGCCCTAACCCTAA 152900
          ><    ><    ><    ><    ><            ><     ><   ><    ><    ><      ><    ><    ><    ><    ><

        CCCTGACCCTGACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCC 153000
          ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    ><    >

        GCCCCCACTGGCAGCCAATGTCTTGTAATGCCTTCAAGGCACTTTTTCTGCGAGCCGCGCGCAGCACTCAGTGAAAAACA                     153080
```

PREDICTED LAYOUT OF HHV-7 RK GENES