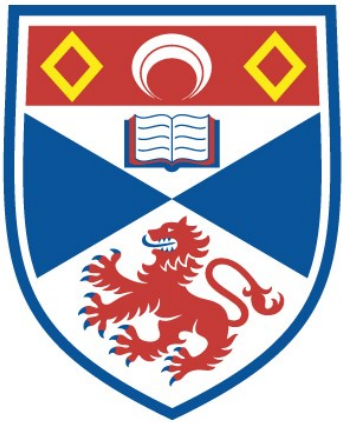
SYNTHETIC STUDIES ON THE ADENOVIRUS SEROTYPE 2 L3 23 KDA PROTEINASE

Thomas James Cromie

A Thesis Submitted for the Degree of MPhil
at the
University of St Andrews



1996

SYNTHETIC STUDIES ON THE ADENOVIRUS SEROTYPE 2 L3 23 kDa PROTEINASE.

A thesis presented by Thomas James Cromie

to the

UNIVERSITY OF ST. ANDREWS

in application for

THE DEGREE of MASTER OF PHILOSOPHY

St Andrews April 1996

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

I, ..., hereby certify that this thesis, which is approximately 21,000. words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 27-5-96 signature of candidate ....

I was admitted as a research student in OCTOBER 1993. and as a candidate for the degree of M.PHIL in OCTOBER 1993...; the higher study for which this is a record was carried out in the University of St. Andrews between 1993... and 1995.....

date 27-5-96. signature of candidate .

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of M.Phil.... in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date 27th May '96 signature of supervisor

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date 27-5-96 signature of candidate ..

Dedicated to my Mother and the memory of my Grandfather, Robert McAnoy.

CONTENTS

List of Figures

List of Schemes

List of Tables

Acknowledgements

First and foremost on the long list of family, friends and colleagues who must be mentioned is of course Dave Gani, to whom I am greatly indebted for his patience, advice and support, cheers Dave the beers are on me. Neck and neck with Dave must be my family, particularly Mum (Financial Services) and Dad (Management Services) and together for Quality Control, my Grandparents for an infinitely sympathetic ear and a constant flow of advice. My brother and sister, David and Sarah, for amusement and reality checks. Thanks folks.

At work, a round of cheers goes to Dr. M. Akhtar (Crackers), Stacey, Dr. A. Cole (George), Dr. N. Camp (Den), Dr D. Perry (Merry), Dr. A. Mehrotra (Smokes), Dr. J. Schulz (Jerry) and Dr. K. Morris. Not forgetting my postgrad. colleagues; Don (TGD) McNab, Roger (ROG) Pybus, Graham (Cool Running) Allen, (Marky) Mark Hillier and Jon (Newky Broon) Park. Cheers lads. A vote of thanks must go to the team in Biochemistry, for their help and advice on the biological side of things including the testing (Dr. G. Kemp, Heather Murray and Dr. Helen Reddy). Thanks also go to Marjory, Colin, Andrew, Bob, John, Melanja, Jim, Caroline, Iain, Sylvia, Arnott and Peter and all the technical staff for great amusement and keeping it all in perspective.

Out of work a huge thanks to all the guys and gals including; Scott, Brian, Duncan, Andrew, Stephen, Stuart, Malcolm, Gregor, the Darts Team and Fencing Club especially Steve, Richard (Bob), Mark (Grandad). Also, at the end much fun was had from Duncan FM, Michelle (Scouse) and Nikki. The boys back home deserve a mention also; Mark, Neil, Stuart, Conor, Jim, Charlie, Leo, Francis, Joanne and Fiona for therapy sessions whilst back in Norn Iron. Without my family I would surely have been a gibbering wreck long before now, thanks once again to one and all, it's been interesting!

Abstract

This Thesis details research undertaken on the continuing project investigating the novel catalytic mechanism of the adenovirus serotype 2 L3 23 kDa proteinase. The proteinase is a virally encoded endopeptidase which is essential for the successful replication, maturation and spread of infectious virus particles. Three tetrapeptide esters were synthesised and tested as potential inhibitors. These tetrapeptides were based on the N-terminal minimum substrate specificity requirement established by Kemp *et al.* and are of the form;

M(L)XGX↓G or M(L)XGG↓X.

From these motifs it is clear that there is a very strict requirement at P_2 for glycine. It is proposed that this requirement is, at least in part, steric and the three tetrapeptides synthesised and tested vary the residue at P_2 in order to establish if this is the case. The three tetrapeptides are;

Cbz-(2*S*)-Leu-(2*S*)-Ala-Gly-Gly-OPr^i (**9**),
Cbz-(2*S*)-Leu-(2*S*)-Ala-(2*S*)-Ala-Gly-OPr^i (**10**),
Cbz-(2*S*)-Leu-(2*S*)-Ala-(2*R*)-Ala-Gly-OPr^i (**11**).

Leucine was chosen over methionine for simplification of synthesis and the isopropyl ester was chosen partly for ease of synthesis and partly to complete a series of "wild-type" tetrapeptide esters which are turned over as substrates. We show that (**9**) does not inhibit the proteinase, that (**10**) inhibits by 37.41% at high concentrations and that (**11**) inhibits by 14.64% at the same high concentrations. The estimated K_i's for (**10**) and (**11**) are 2.1 mM and 5.3 mM respectively, neither are potent inhibitors.

ABBREVIATIONS

ABBREVIATION	MEANING
Ad	Adenovirus
^{t}Boc	tertiary-Butoxycarbonyl
Cbz	Carboxybenzyl
DCM	Dichloromethane
DMF	N,N'-Dimethylformamide
eq	Equivalents
NMM	N-Methyl morpholine
NMR	Nuclear Magnetic Resonance
p-	para-
THF	Tetrahydrofuran

AMINO ACID CODES

Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Methionine	Met	M
Proline	Pro	P
Arginine	Arg	R
Serine	Ser	S
Valine	Val	V

CHAPTER ONE

INTRODUCTION

1.0 INTRODUCTION

The isolation and identification of adenoviruses as distinct viral agents was achieved in 1953 by two groups who independently recognised their appearance in tissue cultures from different tissue sources.[1] Rowe *et al.* were evaluating tonsil and adenoidal tissue, removed from children, for the growth of polio viruses.[2] They observed that a transmissible agent was causing cytopathic degeneration of epithelial like cells in uninoculated cultures after a prolonged incubation, they coined the agent responsible for this toxicity "Adenoid-Degenerating Agent". Hilleman and Werner recovered similar cytopathic agents from respiratory secretions in cell cultures of human upper respiratory tissue taken from febrile military personnel during an epidemic of an influenza-type respiratory disease.[1,3] These agents were coined - "Acute Respiratory Disease" (ARD) agents. It was later discovered that there were multiple serotypes of these agents and that the initial serotypes were antigenetically related by a cross reacting family antigen which gave rise to antibodies detectable by compliment fixation.[4]

The nomenclature of adenoviruses was clarified and adopted in 1956.[5] The Commission of Acute Respiratory Diseases of the U.S. Armed Forces confirmed that adenoviruses were the cause of the majority of acute febrile respiratory syndromes amongst military recruits. Epidemics of ARD were well documented in World War II and occurred almost exclusively in new recruits during the winter months, seasoned personnel who were in contact with the recruits were not infected nor were college students under similar conditions.[1,6] This suggested that the factors of crowded sleeping conditions, fatigue and humid localised environments of barracks provided the ideal set of circumstances for wide spread infection. Controlled studies of routes of infection for ARD's have shown that aerosolised virus inhaled by volunteers

produced the disease whereas application to the mouth, nasal mucosa or the intestine did not. Considering that ARD affected 80% of new recruits with 20% - 40% being hospitalised, it is little wonder that the military expressed such an interest in the early findings and studies on adenoviruses. The symptoms of ARD had been recognised for up to one hundred years before the discovery of this new group of viruses and, indeed, German clinicians in the 1920's isolated a distinct clinical entity from outbreaks of epidemic keratoconjunctivitis which were now able to be identified as adenoviruses.[7]

Adenoviruses belong to the adenoviridae family and are divided into two genera; aviaadenovirus and mastoadenovirus.[8] Although there is a broad antigenic cross-reactivity within each of the genera, there is no common antigen that characterises the whole family. The mastoadenovirus genus include human, simian, bovine, equine, porcine, ovine, canine and opossum viruses. Human adenoviruses have been divided into six groups comprising some 47 serotypes.[9]

1.1 STRUCTURE OF THE VIRUS

Adenoviruses are nonenveloped, regular icosahedrons (20 equilateral triangular surfaces and 12 vertices) that are 60 to 90 nm in diameter. Each vertex has a fibre projecting from it and the length of this fibre varies with the serotype. The viral particle in its entirety is a large macromolecular assembly composed of at least 11 different structural proteins and a linear double-stranded DNA genome. The total mass of the particle is 150 MDa, which is roughly 18 times the mass of the picornaviruses and 6 times the mass of simian virus 40 (SV40), another DNA virus [See Figure 1, Page 3].[1,10,11]

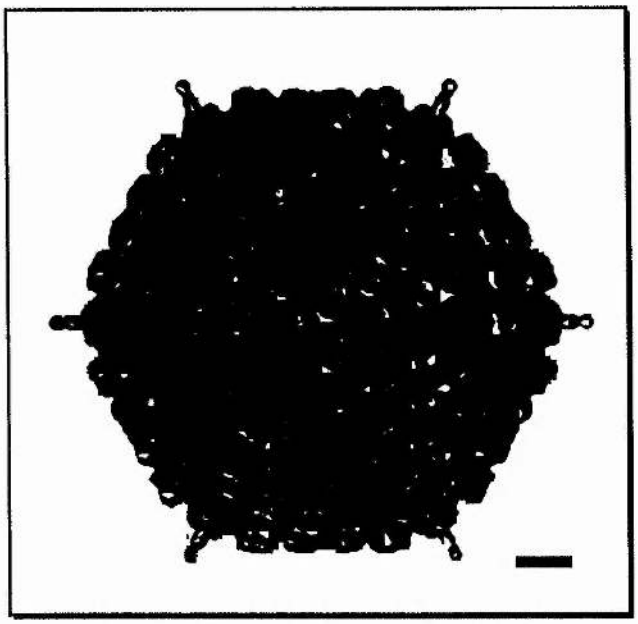


Figure 1: *Electron Micrograph of adenovirus 2.*

The particle is defined as two compartments; i) the outer coat or capsid which is the icosahedron and, ii) the core which contains the DNA genome and is surrounded by the capsid.

The capsid is made up of 252 subunits or capsomeres. Of these capsomeres 240 are hexons, which make up the faces and edges of the triangles, and 12 are pentons, which form the vertices.[12] The pentons consist of a penton base and a fibre which projects from these base, surrounding the individual pentons are 5 hexons and around each hexon are 6 other hexons. Almost all of the structural data available for adenoviruses has been derived from detailed studies on serotypes 2 and 5, hence the following data are those which have been determined for Ad 2. The viral polypeptides that make up the virion have been numbered II to X (including IIIa) and terminal polypeptide.[11] Polypeptides II, III, IIIa, IV, VI, VIII and IX make up the capsid, the remaining peptides plus the viral DNA make up the core [See Figure 3, Page 7].

1.1.1 Capsid Polypeptides

The hexon capsomere is a trimer of polypeptide II and is held together non-covalently, the penton base is probably a pentamer of polypeptide III and is non-covalently linked to the fibre which is made up of three molecules of polypeptide IV, this assembly of penton base and fibre is known as the penton capsomere.[13,14] Polypeptides IIIa, VI, VIII and IX are all associated with these major proteins, polypeptides VI, VII and IX have been known to be associated with the hexon capsomere for sometime and more recently studies combining X-ray crystallography, electron microscopy and difference imaging have produced tentative spatial and structural assignments for all of these proteins.[11] Two monomers of IIIa penetrate the capsid and are sandwiched between hexons at adjacent facets, several copies of VI combine to form ring structures underneath the peripentonal hexons and polypeptide VIII is located on the inner surface of the hexon capsomere. The structure and location of peptide IX has been established more easily since it is located entirely on the surface of the capsid, 12 copies of IX monomers cluster into 4 trimers and each monomer extends along a hexon-hexon interface or groove. This agrees with experimental findings which showed that mutant viruses lacking this protein still form intact virions but are less stable. The protein is synthesised late in infection and is not required for virion assembly so it may be assumed that it attaches to already assembled virions and stabilises them. This type of stabilising effect or "gluing together" of the major capsid proteins is also believed to be the function of proteins IIIa and VI and demonstrates the more highly complex structure of adenoviruses than that seen with smaller previously studied viruses.[11]

1.1.2 Core Polypeptides

The absolute nucleo-protein core structure has not been established but several features have been deduced from studies of the disrupted virion. Polypeptides V and VII are rich in arginine like cell histones but, unlike histones, they also contain tryptophan. After disruption of the virion particle with either 5M urea, 10% pyridine, acetone or multiple freeze-thaw cycles these two polypeptides remain non-covalently linked to the viral DNA.[15,16] In each core there are 1,070 copies of protein VII and 180 copies of protein V, the exact function of these proteins is unclear, however, protein V can bind to a penton base and may therefore serve a positioning role in the packaging of the viral DNA.[17] It appears that polypeptide IIIa may serve a bridging function between the protein VIII of cores and the peripentonal hexons of the capsid.[11,17] The terminal polypeptide (TP) is covalently bound to the 5' ends of the double stranded viral DNA.[18] The core contains another protein - polypeptide X, previously termed μ - which is also rich in arginine and histidine. This protein is probably the cleavage product of one of the larger core proteins. However, its precise location and function are not understood.[19]

1.1.3 Virion DNA

The viral DNA for Ad2 is 23.85 MDa and varies slightly in size between serotypes.[20] it is approximately 11 mm in length and displays the unusual feature of the terminal polypeptide covalently linked to dCMP at each 5' end of the linear genome.[18,21] When the DNA molecules are released from virions (with 4M guanidine) they form circles due to non-covalent interactions between the terminal proteins at each end.[18] Adenovirus DNA also has inverted terminal redundancies at either end of each strand of DNA. The length of these repeats varies from 100 to 140 base pairs with the serotype. A

consequence of this is that after the terminal proteins have been removed by proteolysis, the denatured single strands of Ad DNA can form circles by base pairing these terminal redundancies to form panhandle structures, and it may be that these are important for DNA replication [See Figure 2, Page 6].[22]

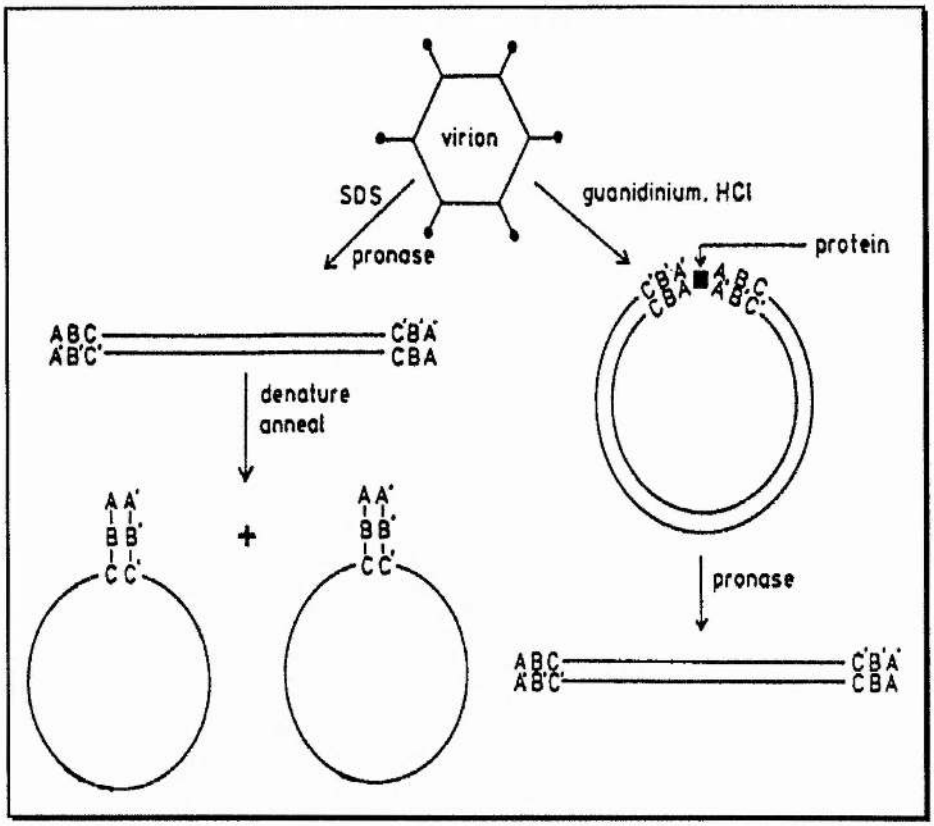


Figure 2: *Scheme of viral DNA digestion.*

Partial nuclease digestion of adenovirus cores results in a diffuse band of DNA of approximately 150 base pairs, that is associated with six polypeptide VII proteins in a nucleosome-like structure.[19] Each of these structures is separated from the next by a variable length of DNA containing one molecule of polypeptide V. The length of the DNA spacer plus nucleosome is approximately 200 base pairs on average but, unlike cellular or SV40 chromatin, the nucleosomes are irregularly placed on the DNA. The six copies of polypeptide VII that make up the nucleosome are arranged as three dimers and the 110 kDa molecular weight of this assembly is very similar to the 108

kDa of the histone octomer in each chromatin nucleosome, however the spacer protein V is much larger than its histone equivalent. Various models of the core packaging have been proposed describing interactions with the faces or vertices of the virion. Although the structure has not been established, the DNA associated polypeptides have been given hypothetical locations [See Figure 3, Page 7].[11,23] The viral DNA is probably compacted by about 5.5-fold in length by its association with these polypeptides.[19]

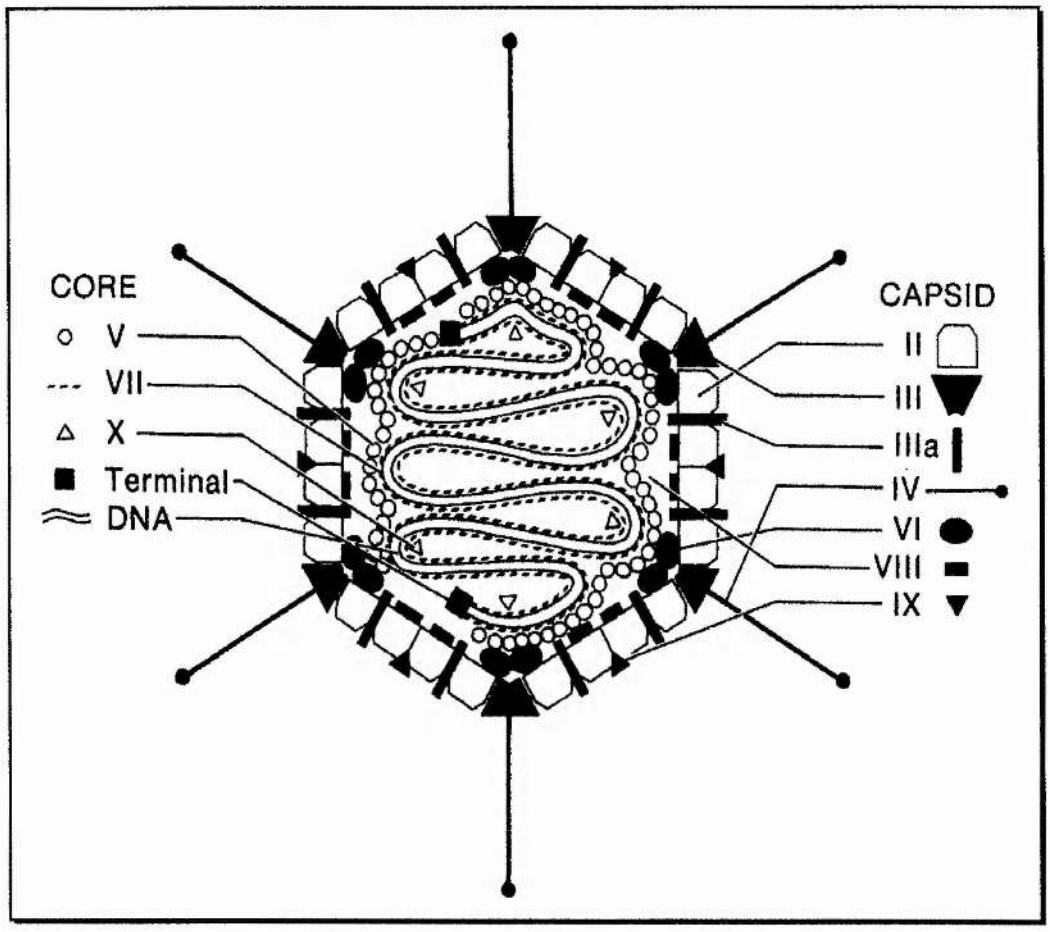


Figure 3: *Pictoral Representation of Viral particle.*

1.2 SUBGROUP AND SEROTYPE CLASSIFICATION

To date at least 5 methods of subdividing human adenoviruses into subgenera have been used [See Table 1, Page 8].[1,9] The first sub-classification was based

on the haemagglutination reaction patterns with rat and rhesus monkey red blood cells.[24] This method, which defines six groups (A-F), compares the extent and selectivity of haemagglutination induced in the two different types of cell as a result of adenovirus infection. With the discovery that Ad12 could induce tumours in rodents a second sub-classification was developed as it was quickly realised that adenoviruses varied considerably in their oncogenic potential and with few exceptions the groups defined by haemagglutination patterns displayed similar oncogenic properties.[4,25]

The current method of classification uses DNA base composition and homology.

Sub-genus	Serotype	DNA %	G+C %	Rat	Haemagglutination Rhesus Monkey	Oncogenicity
A	12,18,31	48-69	48	+/-	-	High
B1 B2	3,7,16,21 14,11,34,35	89-94	51	-	+	Weak
C	1,2,5,6	99-100	58	+/-	-	Nil
D	8-10,13,15,17 19,20,22-30 32,33,36,37 38,39,42-47	94-99	58	+	+/0	Nil
E	4	--	58	+/-	-	Nil
F	40,41	62-69	52	+	-	Nil

Table 1: *Classification of currently recognised adenovirus serotypes.*

Base composition refers to the percentage G+C content and homology refers to conserved sequences within a single group of serotypes.[26] Group A adenoviruses have a comparatively low G+C content of 48% and it was

thought that this was related to tumour induction and that the high G+C content (58%) of group C resulted in non-oncogenic viruses. However, this generalisation does not hold for the simian viruses and is, therefore, probably indicative of the genetic origin or evolutionary variation of a serotype. Homology within single groups, measured by DNA-DNA hybridisation, is greater than 85% except for group A which is 48-69% between its three members.[27,28] Across different groups there is less than 20% homology as determined by the hybrid method. Divergence of the DNA sequence between members of a group is at a maximum in the regions coding for the hexon capsomere and the fibre. These are the capsid proteins to which many of the hosts antibodies are directed.[27]

Most of the adenovirus serotypes show distinct patterns for DNA cleavage fragments after digestion of the viral DNA by restriction endonucleases. Thus, serotypes have been grouped by common, related fragments. This method has also been particularly useful in classifying Ad isolates removed from clinical specimens.[29] Another criterion for group classification is the size of the virion polypeptides using Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE).

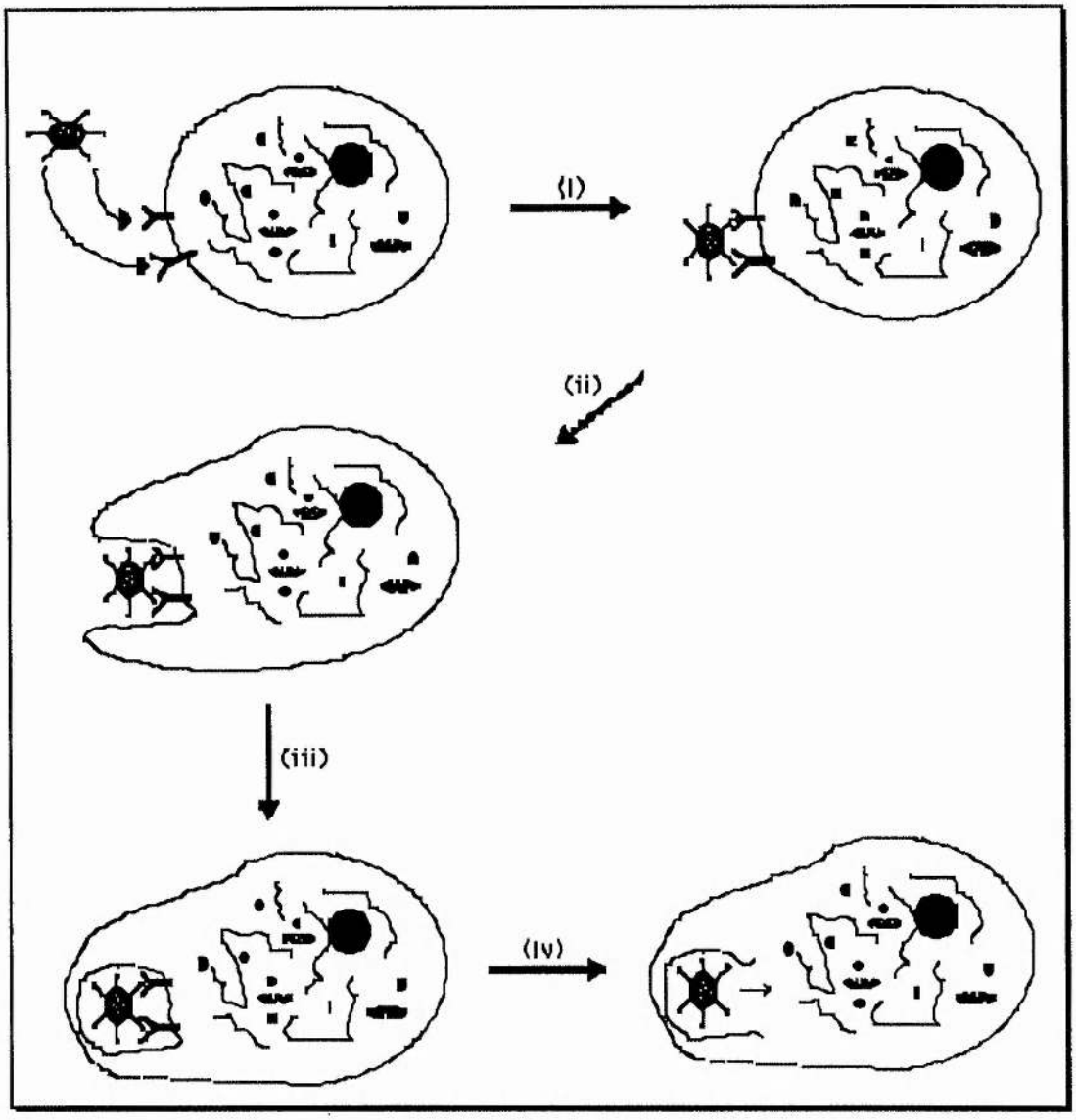
Since these different techniques rely on the properties of many different chemical and biochemical characteristics of the various serotypes, nowadays multiple serologic and biochemical techniques are used for the classification of new adenovirus isolates.

1.3 ADENOVIRUS REPLICATION

The study of lytic infection by adenoviruses has been aided by two major factors associated with adenoviruses. First, most adenoviruses grow well in

suspension cultures of HeLa or KB cells and, second, human adenoviruses remain cell-associated well after the production of new virions.[30] HeLa cells are a continuous, immortal epithelial cell line removed from a cervical tumour from Henrietta Lacks in 1951, KB cells are also a continuous epithelial cell line and both types of cell are used extensively in microbiology as standard cell lines. This makes it possible to concentrate large volumes of virus infected cells and study their biochemistry at high multiplicity ensuring synchronicity of infection and of subsequent biochemical events in the replicative pathway. The replicative cycle consists of several stages; invasion of the cell and transport to the nucleus, onset of replication (termed Early (E) Phase), viral DNA replication, start of Late (L) Phase, virion assembly and cell lysis with the release of new virus particles.[1] Early Phase and Late Phase are defined as the events which occur before and after viral DNA replication respectively but these are not absolute definitions. Indeed some polypeptides that are synthesised early are also synthesised late and vice versa.[31]

The virus is adsorbed slowly onto the cell surface *via* the fibre attaching to specific receptors on the cell membrane.[9,32] This fibre-cell receptor interaction does not involve carbohydrate recognition and is in fact an amino acid interaction.[33] The virion is then phagocytosed into receptosomes over several hours.[9] The penton polypeptide lowers the pH within the receptosome to 5.5 which probably serves to alter the configuration of the virion and results in rupture of the phagocyte membrane and release into the cytoplasm [See Scheme 1, Page 11].[1] In vitro studies have shown that the cytoplasmic virions are attached to microtubules by the hexon and the process of transport to the nucleus probably involves these structures.



Scheme 1: *Cartoon showing the adsorbtion of the virus into a cell. (i) Virus attachment to cell membrane receptors; (ii) + (iii) Phagocytosis of particle into cell receptosome; (iv) Penton lowers pH to 5.5 and virus released into cytoplasm.*

The virus is uncoated and the viral core enters the nucleus, many of the viral proteins are left in the cytoplasm.[34] Release into the cytoplasm, transport and insertion into the nucleus takes place over 1-2 hours. After uncoating and before the onset of replication the virus cannot be recognised cytopathically or antigenically nor can it be observed microscopically.[9] This is the eclipse phase

which is quite long and therefore the time taken for adenovirus infection to be detected in cell culture is also long. The early phase of replication may be further subdivided into immediate early (IE) and early (E). Before DNA replication commences the IE and E mRNA's are transcribed. This transcription takes place from 5 distinct regions of the genome.[9] The immediate early gene products promote and enhance the transcription of the early genes. These early genes are translated from early mRNA into early proteins required to initiate viral DNA replication. Of these delayed early mRNA molecules is E2B, which codes for a 140 kDa DNA polymerase and an 80 kDa precursor to the terminal protein.[35] Viral DNA replication commences in the nucleus approximately 6-8 h post-infection signifying the end of the early phase. By 20 h DNA replication is at a peak rate, stopping after a further 4 h signifying the start of the late phase.[9] The synthesis of host cell DNA is reduced by more than 90% after the onset of viral DNA replication *i.e.* 6-8 h post-infection.[36,37] The new DNA strand elongates in a continuous manner from the 5' end to the 3' end, *i.e.* one parent strand is replicated whilst the other is displaced and free. This single strand of free DNA spontaneously circularises into the panhandle structure previously described.[1] From this panhandle a second daughter strand elongates from the 5' end to the 3' end separating and replicating the panhandle and further into the "pan" to form a second molecule of viral DNA. The Terminal Protein which is covalently linked to the first nucleic acid serves as a primer for DNA replication initiation so both the 140 kDa DNA polymerase and the 80 kDa pTP coded by E2B are required in the first instance for replication. With the onset of viral DNA synthesis comes the transfer from early phase to late phase. During this phase the majority of the viral polypeptides are synthesised and the synthesis of host macromolecules is compromised. The late viral polypeptides are synthesised at a maximum at 20 h post-infection, host protein synthesis having been halted for some 4 h previously or 16 h post-infection.[9]

This inhibition of host macromolecular synthesis may be partially explained by the depletion of host mRNA on cytoplasmic ribosomes. However, primary transcription of host cell RNA is still active in the nucleus. There is a breakdown of the transport of these cell mRNA precursors to the cytoplasm. As there is no rapid breakdown of cell mRNA on ribosomes, it is still difficult to explain the cessation of host protein synthesis in terms of the inhibition of transport or processing of nuclear transcripts.[1] These features, therefore, point to some form of translational control. The inhibition of new ribosome formation is also profound and similar to that caused by inhibitors of protein synthesis. This process of host translation shut down in the late phase has been shown to prevent cell division and promote destruction of the intermediate filament network.

Many mRNA molecules transcribed in the early phase are also transcribed in the late phase. The structural viral polypeptides are mainly synthesised in the late phase but IVa2 and IX are also synthesised early. In the late phases some of the transcripts from L2 and L3 have been found early, particularly the L3 23 kDa proteinase which is always found in small amounts early in infection.

1.3.1 Assembly of Virions

Virion assembly takes place in the nucleus but the requisite capsomere structures are formed in the cytoplasm.[38] The first capsomere to be assembled is the hexon, a trimer of identical polypeptides held together non-covalently. This assemblage occurs with the assistance of a 100 kDa non-structural viral protein and occurs in minutes.[39] The hexons then self assemble through the formation of ninemers into the 20 faces of an empty shell capsid. The penton capsomere assembles from the penton base pentamer and three fibre polypeptides over a somewhat more prolonged period. The first 25% of

pentons are assembled within 20 min with the rest forming over some 10 h.[38] The details of penton assembly are not clear. However, it is likely that many are not incorporated into the capsid until a later stage.

The next recognisable stage is the capsid or light intermediate stage.[40] During this phase the capsid associated polypeptides VI and VIII become incorporated into the hexon shell as their precursors (pVI and pVIII) along with IIIa and a 50 kDa and 39 kDa viral protein. The 50 and 39 kDa polypeptides act as scaffolds in the formation of the light capsid and are removed shortly after.[1,40] Next comes the insertion of the viral DNA with the terminal protein precursor (pTP) then come the core polypeptides V and precursor VII. Since it is believed that the penton capsomere has not been fully incorporated at this point the DNA presumably inserts through one of the open 12 vertices. Protein IX becomes associated at this stage also. This state is known as the young virion, maturation from young to mature occurs with the action of the L3 23 kDa proteinase cleaving the precursor proteins into their active or mature forms in the core.[41] The fragments of these polypeptides pVI, pVII, pVIII and pTP are degraded or ejected as the particle firms up its mature conformation with final association and consolidation of the pentons. The proteinase contains over 200 amino acid residues and its sequence is highly conserved between serotypes (72% between Ad2 and Ad4). Sequence analysis has shown that there is no primary structure homology with any other known proteinases, therefore, this proteinase probably represents a new class of enzyme.[42] Only about 15% of new viral proteins and DNA are incorporated into virions.[9] The viral material not incorporated acts to promote the cytopathic effect (CPE) by shutting off host macromolecular synthesis and by disruption of both the cytokeratin network of the cytoplasm and the nuclear membrane.[1,9,43,44] This results in cells ceasing to divide, subsequent cell death and lysis releasing the new virus particles. The penton capsomere is

directly toxic to cells other than those infected with adenovirus.[1,9] Although its role is not clear in this area, pentons have been found in the blood of fatal cases of adenovirus related pneumonia.[45]

1.4 CLINICAL SYNDROMES

Adenoviruses infect and replicate in a wide range of tissues producing many different diseases most of which are subclinical [See Tables 2 and 3, Pages 23 and 24].[9] These tissues include, most commonly, the respiratory tract, eye, gastrointestinal tract and bladder.[45] Sporadically they cause disease in more remote organs, the central nervous system for instance, where meningoencephalitis may be induced and complicated by an adenovirus infection.[9] Adenovirus infections occur world-wide and are species specific with patterns of infection varying from sporadic to epidemic. The virus is spread *via* respiratory or ocular secretions and by the faeco-oral route in young children.[9] Other means of trans-infection can be swimming pools and improperly sterilised ocular equipment. In 40-60% of young children the most common antibodies present are to types 1, 2 & 5 which are endemic in most populations, but few have been exposed to types 3, 4 or 7 and these are the serotypes responsible for epidemic infections.[9]

1.4.1 Respiratory Diseases

Adenoviruses are responsible for 5% of the acute respiratory disease in young children with symptoms including nasal congestion and coughing.[46] In addition to these symptoms there are often further effects such as general malaise, fever, chill and headache. Isolated cases are often indistinguishable from influenza, parainfluenza and other viral respiratory infections. Types 1, 2, 5 and 6 (group C or HA group III) and occasionally type 3 (group B or HA

group I) are the most common causes and are endemic in most populations. When conjunctivitis in one or both eyes is coupled with the above symptoms the disease is known as pharyngoconjunctival fever.[47] This is usually mild and most commonly caused by type 3 but types 7 and 14 within the same HA group have been found in such patients.[48] In more severe instances pneumonia may be the result and adenoviruses account for approximately 10% of childhood pneumonias.[45] These lower respiratory infections tend to be fairly trivial but where there have been epidemics of Ad 7, mortality rates were considerable. Of those who recover bronchictasis may clinically manifest itself years after primary infection. Acute respiratory disease (ARD) of military recruits as described previously (Page 1) is similar to that of the respiratory infection of children. Ad 4 and Ad 7 are responsible and less often Ad 3, this syndrome occurs under conditions of fatigue and crowding which is characteristic in young military recruits.[48]

Adenovirus infection has been linked in many cases to cases of whooping cough and some believe that it is this rather than Bordetella Pertussis that is responsible.[49] Ad 5 was isolated from multiple organs of a patient with whooping cough and lymphocytosis that ended fatally, however, adenoviruses alone are probably not responsible for the syndrome. As with many other adenovirus infections, symptoms related to the virus appear with concurrent infection by other pathogens. Latency has been shown to be prolonged and often disease never results, except in the case of immunosuppression by other infections.[50]

1.4.2 Ocular Diseases

Swimming Pool conjunctivitis is probably most commonly due to adenoviruses and acute follicular conjunctivitis can occur with respiratory-pharyngeal

syndrome or by itself. Significant preauricular lymphadenopathy often accompanies this disease.[47,51] This is a rather mild disease and complete recovery is the most common result. Infection may be sporadic and affect only family members or in the cases where the source is a lake or swimming pool the disease can be widespread amongst children and young adults.[45] The water-borne outbreaks occur almost entirely during the summer months and are caused by Ad 3 and 7 although a wide number of serotypes spanning groups B, C, D and E or HA groups I, II and III have been associated with this syndrome.[51] A more serious, highly contagious disease caused by adenoviruses is epidemic keratoconjunctivitis (EKC).[45] This was identified as a clinical entity before adenoviruses and occurred amongst German workers at the turn of the century. It was subsequently found in ship-yard workers in the continental United States and Hawaii and coined "Ship-yard Eye".[7] It was almost certainly transferred through the medical facilities that catered for trauma to the eyes of workers. Symptoms include follicular conjunctivitis, edema of the eyelids, pain, photophobia and lacrimation. The disease is often unilateral with preauricular lymph node hypertrophy. In some instances a haemorrhagic conjunctivitis develops which must be distinguished from the enterovirus 70 acute haemorrhagic conjunctivitis.[45] EKC was originally caused by Ad 8 then through the 1970's Ad 19 was more common and since 1977 Ad 37 has been the prominent progenitor.[52,53] EKC has occurred as an epidemic in parts of Taiwan, Vietnam and Japan during the autumn months and has accompanied endemic trachoma in Saudi Arabia.[45,54]

1.4.3 Renal Diseases

American children rarely have type 8 antibodies whereas half the Japanese or Taiwanese children do. Acute haemorrhagic cystitis is an illness that occurs almost solely in male children and has been linked to adenovirus type 11. The

problem with this infection is its confusion with more serious diseases of the kidney such as glomerulonephritis.[45] Type 21 can also cause haemorrhagic cystitis and belongs to the same group as type 11, group B or HA group I. Cases of acute haemorrhagic cystitis in Japanese males between the ages of 6 and 15 showed adenoviruses in isolates from urine in approximately 70% of patients. In contrast, under the same study conditions, only 20% of haemorrhagic cystitis in the U.S.A. could be linked to adenovirus infection. Following renal transplantation and its accompanying immunosuppression three cases have been reported of Ad 11 acute haemorrhagic cystitis.[45] In one of these the infection was introduced with the new kidney, but this is not a typical complication of renal transplants and as previously described adenoviruses are opportunistic on hosts with suppressed immunity. Coincidentally two other group B serotypes 34 and 35 were first isolated from renal transplant patients. Type 34 from urine and 35 from kidney and lung tissue.[55] Whilst neither gave rise to haemorrhagic cystitis, Ad 35 undoubtedly contributed to the pneumonia which killed the patient.

1.4.4 Diseases of the Central Nervous System

Although adenoviruses are not commonly found in the cerebrospinal fluid, some cases in which serotypes have been isolated from the brain and cerebrospinal fluid have been reported (types 3, 5, 6, 7, 7A, 12 and 32). For instance, a patient receiving chemotherapy for the treatment of a malignant lymphoma, and consequently immunosuppressed, had an adenovirus type 32 isolated from the brain at post-mortem.[56] Other instances of meningoencephalitis have been reported in children, these have been particularly associated with epidemic adenovirus type 7 pneumonia. In addition an isolated case of sudden deafness was linked to an adenovirus type 3 infection of the nasopharynx.[45]

1.4.5 Diarrhoea and Intestinal Diseases

Adenoviruses and diarrhoea have had a complicated history which has recently been clarified. Since it had been well established that many adenovirus serotypes replicate efficiently in the intestine, research was undertaken in order to identify the connection between adenovirus infection and diarrhoea.[57,58] Attempts to correlate adenovirus growth from stool with clinical illness did not prove to be conclusive as just as many adenovirus isolates were found in the stools of controls as those from subjects with diarrhoea. So although young children with systemic adenovirus infections can often develop diarrhoea these instances are probably no more common than that caused by any other systemic infection.

Following the discovery of "non-cultivable" adenoviruses by electron microscopy of stool smears from children suffering gastro-enteritis, research has re-examined the possibility that these were the cause of infantile gastro-enteritis.[58] Two new, different serotypes of the human adenovirus, known as the enteric adenoviruses (EAd), have been strongly associated with diarrhoea and acute gastrointestinal disease (serotypes 40 and 41). Of 27 hospitalised young children with acute gastrointestinal disease, 14 were related to enteric adenoviruses and it was suggested that the respiratory symptoms also observed were a major facet of the clinical manifestations. The incidence of adenovirus related gastro-enteritis is approximately 12% of all the intestinal infections in young children. Over some population groups 50% of children aged 4 express antibodies to these enteric adenoviruses.[59]

Intussusception, another intestinal syndrome, has been associated in some cases to adenovirus infection.[45,60] Adenoviruses have been isolated from the infected lymph nodes removed at surgery and also from stool cultures. Two

explanations have been proposed; i) the telescoping bowel characteristic may be caused by mesenteric adenitis acting as a lead point to the obstruction and ii) that the hyper irritability of the small intestine might be caused by adenovirus infection and thus may lead to the intussusception. However, most of the patients with intussusception show no signs of adenovirus infection. Therefore, adenoviruses are probably one of many factors which might lead to this syndrome.[61]

Interestingly A-gliadin, a major component of α-gliadin which is known to activate coeliac disease, possesses considerable structural homology to the adenovirus type 12 E1B protein. Coupled with the finding that most patients with coeliac disease had indeed suffered prior Ad 12 infection, in contrast to the controls, these findings suggest that the type 12 viral protein E1B may act as a trigger inducing the pathogenesis of coeliac disease.[62]

1.4.6 Immunocompromised Patients

Adenovirus infections amongst immunocompromised patients, whilst not as common as the various herpes virus infections, have contributed to the mortality of these patients.

From 1967 to 1978 at the UCLA Centre for the Health Sciences 15 immunocompromised patients with typical endemic adenovirus infection were studied.[50] Bearing in mind that most of the population are infected with many different serotypes of adenovirus and have never manifested any clinical symptoms. Twelve of these patients suffered from pneumonia, 8 had hepatic abnormalities and 9 died of hepatic or respiratory failure. In only three instances was there evidence of other types of infection. In children the serotypes responsible were 1, 2, 5 and 6 and in adults, type 4. It is obvious

from these data that, whilst immunocompromised patients are no more susceptible to adenovirus infection, the consequences of such infections are much more serious and in the majority of cases fatal. Over recent years the obvious group to suffer from immunity suppression are AIDS patients, indeed, 12% of the AIDS patients at UCLA displayed adenoviruses in their urine. As mentioned previously (Page 14) restriction endonuclease digestion of viral DNA can be used to group serotypes by their common cleavage fragments. This technique was used to identify the serotypes isolated from AIDS patients urine and showed that more than 85% of the viral DNA fragments isolated were attributable to Ad 35. Also there appears to be some recombination across serotypes as observed from the haemagglutination patterns of the AIDS urine isolates.[63] These recombinations are poorly understood and only occur in AIDS patients to date.

Another group who suffer immunity suppresion are patients who have undergone bone marrow transplants. These patients have also had similar group B adenoviruses isolated from their urine but this is not as common as in AIDS patients. AIDS patients have also had many group D adenoviruses isolated from stool and rectal swabs.[64]

1.4.7 Latency and Malignancy

The latency of adenoviruses has been well established, indeed, it was a consequence of this property that led to the discovery of adenoviruses in adenoidal tissue.[1,2] Adenoviruses were shown to persist in healthy hosts up to 2 years after initial infection. B lymphocytes in particular have been suggested as helping to maintain latency by several groups who have shown low levels of viral DNA incorporated in peripheral blood lymphocytes.[65] A confirmation of latency in tonsil tissue was made by the demonstration that adenovirus

nucleic acid sequences were present in this tissue but without infectious virus.[66] However, the ultimate example of latency which would be the stable integration of sequences of viral DNA in covalent linkage with the host chromosome has not been observed. It has been clearly established that some serotypes are highly oncogenic in animals, particularly types 12, 18 and 31, but attempts to link human tumours to adenovirus infection have been unsuccessful.[67] However, subject to confirmation is the finding by one research group of Ad nucleic acid in neurogenic tumours.

Site of infection/ disease	Serotypes
RESPIRATORY TRACT	
Acute febrile pharyngitis	1,2,3,5,7
Adenopharyngoconjunctival fever	3,7
Acute respiratory disease of military recruits	4,7,14,21
Pneumonia	1,2,3,7
Pertussis-like syndrome	5
EYE	
Follicular conjunctivitis	3,4,11
Epidemic keratoconjunctivitis	8,19,37
KIDNEY	
Acute haemorrhagic cystitis	11,21
CENTRAL NERVOUS SYSTEM	
Meningitis	3,7
Subacute focal encephalitis	
GUT	
Diarrhoea	40,41
Intussusception	1,2,5
Coeliac disease (?)	
OTHER	
Rubelliform rash	
Acute febrile polyarthritis	

Table 2: *The clinical syndromes produced by adenoviruses and the most common serotypes associated with them.*

Disease	Individuals most at risk	Principal serotypes
Acute febrile pharyngitis	Infants, young children	1,2,3,5,6,7
Pharyngoconjunctival fever	School-age children	3,7,14
Acute respiratory disease	Military recruits	3,4,7,14,21
Pneumonia	Infants, young children	1,2,3,7
Pneumonia	Military recruits	4,7
Epidemic keratoconjunctivitis	Any age group	8,11,19,37
Pertussis-like syndrome	Infants, young children	5
Acute haemorrhagic cystitis	Infants, young children	11,21
Gastroenteritis	Infants, young children	40,41
Hepatitis	Infants and children with liver transplants	1,2,5
Persistence of virus in urinary tract	AIDS, other immunosuppression and bone marrow transplant recipients	34,35

Table 3: *Illnesses associated with adenovirus infections, the groups most at risk and the most common serotypes responsible.*

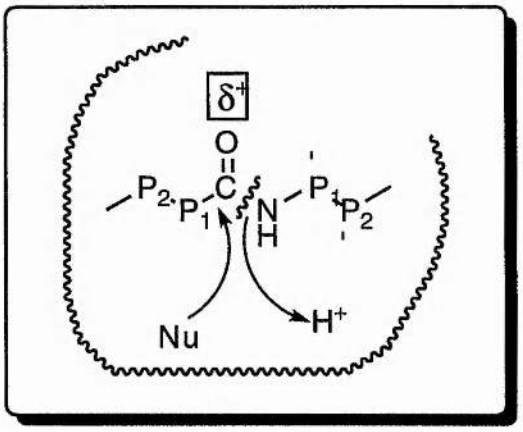
1.5 THE ADENOVIRUS SEROTYPE 2 L3 23 kDa PROTEINASE:

The 23 kDa proteinase is a virally encoded enzyme and has been the subject of much discussion and research effort since its putative identification and the observation of its varied functions.[43,44,68-78] Many viruses encode for their own proteinases, this "self-encoding" is not uncommon in DNA viruses, occurs in all retroviruses and in many RNA viruses.[79] Proteolytic processing in the life cycle of a large number of viruses is an absolute requirement for successful replication, maturation and spread of new infectious virus

particles.[79] Proteolytic enzymes are normally thought of as serving degradative roles but proteolysis is involved in a productive manner in several important biosynthetic processes such as blood coagulation, fertilisation and the production of peptide hormones.[79] However, these proteolytic enzymes are subject to strict regulation mechanisms as over-enthusiastic digestion would prove disastrous to the organism. This requirement for regulation and specificity for the types of peptide bonds to be hydrolysed is nowhere more evident than in the polypeptide processing in viruses.[79] Cleavages catalysed by virus encoded proteinases are particularly complex and occur by a variety of mechanisms serving diverse functions. One of their major functions is to facilitate the assembly of new virus particles principally by proteolytic processing of precursor polypeptides into the requisite forms for capsid formation and DNA encapsidation. Other functions include a role in the uncoating of virus particles in the newly infected cells and contributing to the cytopathic effect by degrading structural proteins within the host cell.[80] Therefore, the extremely high substrate selectivity and specificity of peptide bond hydrolysis is inherently necessary for the successful propagation of viral infection. This strict specificity also removes viral proteinases from regulation by host cell inhibitors. It is this characteristic, coupled with their pivotal role in infection that sets up virally encoded proteinases as potential therapeutic targets. The rationale being that a synthetic compound which specifically inhibits the action of the viral proteinase would, by definition, be unlikely to interfere with similar host enzymes offering an extremely attractive opportunity for the development of anti-viral drugs. Obviously before potential inhibitors or therapeutic drugs can be designed or synthesised the precise chemical mechanism of the enzyme must be elucidated. In order to determine the chemical mechanism, the residues which are actively involved in the reaction must be known as does their spatial arrangement and their electronic properties. As previously mentioned the L3 23 kDa proteinase

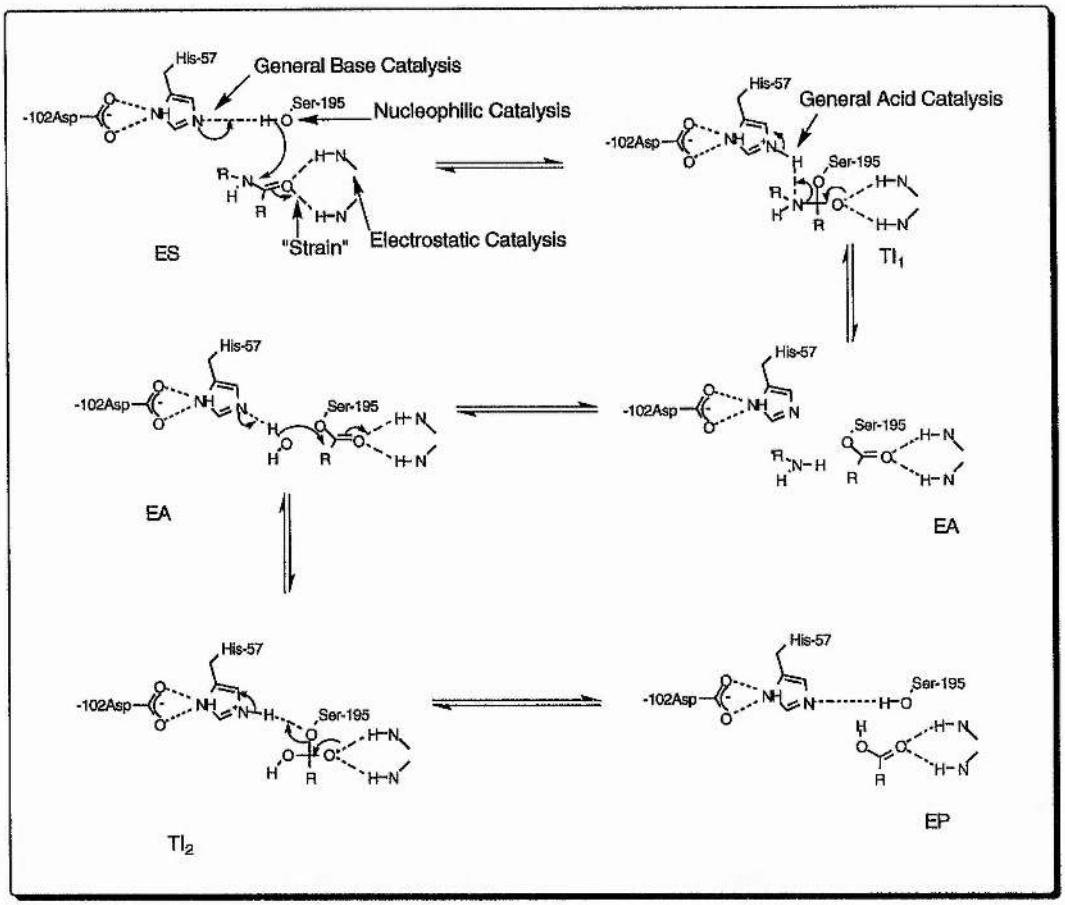
amino acid sequence displays little homology with any known enzymes and as such probably represents a new class of enzyme or at least presents a novel mechanistic problem.

At the present time there are four classes of proteinase which are defined by the residues involved in the proteolysis. Serine proteinases and cysteine (thiol) proteinases are so-called because the active site nucleophile is a serine and a cysteine residue, respectively. The nomenclature for the other two classes refers to the group in the acitve site which polarises the carbonyl of the substrate in order to facilitate nucleophilic attack and these are the Acid proteinases and Metalloproteinases [See Scheme 2, Page 26].



Scheme 2: *Generalised mechanism of enzyme catalysed peptide bond hydrolysis. Where for serine and cysteine proteinases, Nu = oxygen and sulphur respectively and for acid and metallo- proteinases, δ^+ = carboxylate and a metal ion respectively.*

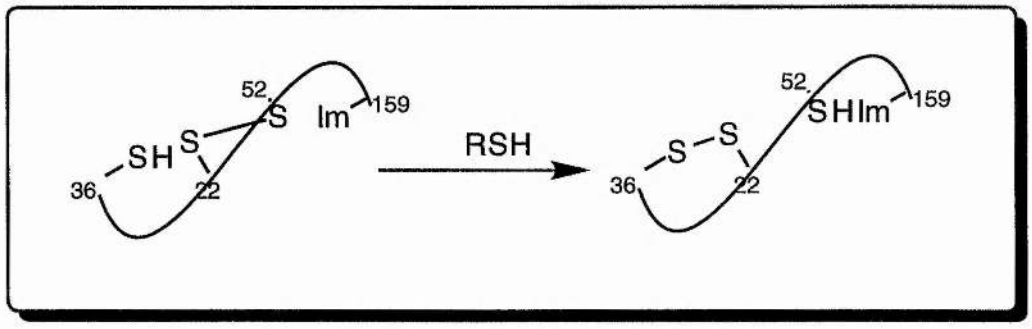
Examples of enzymes which belong to these four classes are; chymotrypsin (serine) [See Scheme 3, Page 27], papain (cysteine), pepsin (acid) and carboxypeptidase A (metalloproteinase, Mn^+).



Scheme 3: *Detailed scheme of enzymatic peptide hydrolysis, in this case the enzyme is the serine proteinase chymotrypsin.*

Attempts to classify the enzyme as either a serine or a cysteine proteinase have recently been clarified by site directed mutagenesis.[74,75] Site directed mutagenesis on its own is not sufficient to prove one hypothesis or the other since by its very application one alters the primary structure of the protein and therefore probably the secondary structure and possibly the tertiary structure of that protein. However, the results of the SDM experiments together with an increasing plethora of evidence strongly suggesting that the enzyme is a cysteine type proteinase makes the case for cysteine type catalysis very strong indeed. The evidence that had been presented classifying the L3 23 kDa proteinase as a serine proteinase was based on inhibitor studies which showed that diisopropylfluorophosphate, a compound which reacts covalently

with the active site of serine proteinases, blocked the activity of this enzyme.[68] However, on it own this is not sufficient confirmatory evidence since it is known that other cysteine proteinases, notably carboxypeptidase B and the neutral cytosolic proteinase that cleaves the B chain of insulin, are inhibited by both cysteine proteinase inhibitors and metallo proteinase inhibitors.[81] Further another group of cysteine proteinases, lysosomal catheptic carboxypeptidase A and prolylendopeptidase, are inhibited by both cysteine proteinase inhibitors and serine proteinase inhibitors.[81] Therefore classification on this criterion alone can be erroneous. However, the grounds for this classification lay not only in the inhibition by DFP but with the additional information that mercaptoethanol had a negative effect on catalysis.[68] This certainly is uncharacteristic of cysteine proteinases as very many of the previously studied cysteine proteinases, the papains and cathepsins, are activated by reducing agents [See Scheme 4, Page 28].[83] This apparent contradiction will be further discussed as part of the novel mechanism of this enzyme.



Scheme 4: *Conversion of the catalytically inactive propapain to the active enzyme, papain, by the action of a low molecular weight thiol.*[83]

1.5.1 Summary of Some Characteristics of Cysteine Proteinases

Cysteine proteinases, previously known as thiol proteinases, are that group of endopeptidases whose members rely on the thiol group of a cysteine residue for catalytic activity [See Scheme 2, Page 26].[80] The cysteine proteinases

which have been extensively studied all appear to contain a thiol-imidazole interactive system within the active site and this is probably a common motif throughout the group. This system most probably functions as the thiolate/imidazolium (S^-/ImH^+) ion pair, the thiolate anion being transiently acylated during catalysis, assisted by general acid catalysis from the imidazolium ion [See Scheme 6, Page 44]. The ion pair produced by the protonic dissociation with a pK_a 3-4 could be a common feature of the group and localised low pH environments within enzymes are common. A variable feature between members of the group is another pH dependant process which endows or enhances the ion pair catalytic competence. This second pH dependant event which is probably involved in the control of the ion pair geometry is required to separate the components of this intimate, mutually-solvated, ion pair at some point of the acylation process of catalysis to permit them to act as nucleophile and general acid catalyst. Interestingly in the cysteine proteinases papain and ficin the residue on the N-terminal side to the active site histidine is aspartic acid. In cathepsin B, streptococcal proteinase and calpain I it is glycine and in cathepsin H it is asparagine. It is further interesting to note that the residue adjacent to the catalytic cysteine residue in streptococcal proteinase is histidine and in calpain I is aspartate. The presence of these charges at or around the active site are probably responsible for the variation in the second pH dependant event and contribute to the minor reactivity differences of cysteine proteinase S^-/ImH^+ catalytic systems.

Cysteine proteinases have been isolated from a wide variety of biological sources spanning plants, animals (including humans, rats, cattle and poultry) and bacteria. Almost all cysteine proteinases are small enzymes in the range 20-35 kDa. The molecular weights of papain and actinidin are 23.35 and 23.5 kDa respectively.[81,83] The calpains ~110 kDa and clostripain ~55 kDa are two examples of cysteine proteinases that are significantly larger. Both enzymes

are composed of two subunits, one light and one heavy. The calpains are heterodimers, both subunits contain potential Ca^{2+} binding sites and it is the heavy subunit that contains the active site. Clostripain is composed of two chains linked by non-covalent forces, it is interesting to note that the lighter (15.4 kDa) of the chains is terminated by an arginine residue. It appears probable that the chains originate from a single chain precursor given that clostripain has a marked specificity for cleavage at arginine residues. Isoelectric points of cysteine proteinases span a range from 3 to almost 12. Most are acidic to neutral including actinidin (pI 3.1) fruit bromelain, clostripain and cathepsin B (pI 4.5-5.5) the other cathepsins (pI 6-7). Ficin, stem bromelain and calotropin DI are all basic pI >8.75 and the most basic is the papaya proteinase W (pI 11.7).

In biological systems where cysteine proteinases occur naturally *i.e.* manufactured by the host as in the examples above (papain, actinidin, the calpains, clostripain, fruit bromelain, the cathepsins, ficin, stem bromelain, calotropin DI and papaya proteinase W), their principle function is the regulation of intracellular and extracellular protein turnover.[83] Cysteine proteinases are responsible for over 50% of intracellular proteolysis, in particular cathepsin L alone was found to be responsible for 50-60% of the proteolytic activity of hepatic lysosomes at pH 6.0.[81] Cathepsins are distributed throughout many tissues (liver, spleen, kidneys, lungs, skeletal muscle and brain) and are associated with various inflammatory reactions and diseases where their activity increases dramatically. In developing muscular dystrophy there is an increase in the number of cysteine proteinase containing lysosomes near the degenerating myofibrils of the muscle and increased cathepsin B activity has been reported in the macrophages of patients with rheumatoid arthritis.[84] Cathepsin B and cathepsin B-like activity has been observed in various forms of cancer cells.[85] A cathepsin B-like enzyme is

selectively secreted by malignant human breast tumours and a similar enzyme which is immunologically related to cathepsin B is found in the ascitic fluid from patients with neoplasia.[81] In the plasma membranes and nuclei of neoplastic cervical cells a heat and alkali stable cathepsin B-like enzyme has been detected. Also reported is an increase in the cathepsin B activity in a B16 melanoma variant and it is probably the case that cathepsin B release from tumour cells aids invasion and extravasation of tumours.[86]

The preceding group of cysteine proteinases are known as the "Low Molecular Weight" lysosomal proteinases, a second group of cysteine proteinases are those referred to previously as the calpains (Page 29). These enzymes are dependent on specific cellular Ca^{2+} concentrations for their activity. Within this second subgroup there are a further 2 subdivisions. Calpain I and calpain II. The calpains occur in almost all of the tissue cytoplasms studied to date and in many of the corresponding cell membranes. Their calcium dependency is poorly understood but it is known that calpain I is activated by micromolar concentrations of Ca^{2+} whilst calpain II requires millimolar concentrations for activity.[83] Endogenous, specific protein inhibitors, calpastatins, occur as widely as the calpains themselves, and almost certainly constitute an irreversible intracellular regulatory system. It has been shown that calpastatin inhibits calpains I and II in the presence of Ca^{2+}. The calpain-calpastatin complex does not associate in the absence of Ca^{2+} and dissociates if Ca^{2+} is removed from an existing complex. This suggests, therefore, that the calpains undergo a conformational change in the presence of Ca^{2+} resulting in the active form of the enzyme and it is this form of the enzyme to which the calpastatin binds. A calcium dependent proteinase regulator (CDPR) has been partially purified from bovine brain which stimulates calpain activity by up to 25-fold but does not affect calcium sensitivity.[87] The role of CDPR in the regulation of calpain activity is not understood. The biological role of calpains

differs slightly from the lysosomal cysteine proteinases in that they are not major contributors to intracellular proteolysis. Their major functions include the regulation and activation of key metabolic enzymes, the modification of receptors by virtue of fluctuations in cell Ca^{2+} concentration and the degradation of cytoskeletal proteins and myofibrillar structures. It has been demonstrated that both calpains are responsible for the activation of protein kinase C and that calpain II modifies pyruvate kinase.[88] Calpains are also involved in the degradation of smooth muscle tissue by their hydrolysis of myofibrils. Calpains have also been found in the retina where their function is to rapidly hydrolyse microtubule proteins along with actin and vimertin, two important cytoskeletal proteins in the eye.[83] Proteolytic modification of receptors by calpains has been observed in a range of systems; the oestrogen receptor in the bovine uterus, the progesterone receptor of chick oviducts, the membrane neurotransmitter for glutamate in the rat hippocampus, the receptor for epidermal growth factor and others.[89] Although the full physiological significance of all these processes is not fully understood it appears that calpains are not only involved in the regulation of these receptors and their efficient function but also in some cases the production of receptor fragments required for further biological processes.[90] It also seems likely that calpains play a role in muscular degradation in muscular dystrophy as well as lysosomal proteinases by the degradation of the Z band, troponin I and C and the myosin heavy chains in muscle fibres.[81,83]

In addition to the hydrolysis of peptide bonds, cysteine proteinases catalyse the hydrolysis of esters, anilides, *p*-nitroanilides and *p*-nitrophenyl esters and these compounds are widely used mechanistic tools.[83] The specificity of proteinases is described in terms of the Schechter and Berger model, in this model a polypeptide substrate is represented as:

$$P_4\text{-}P_3\text{-}P_2\text{-}P_1\text{-}\downarrow\text{-}P'_1\text{-}P'_2....$$

where P represents an amino acid residue which binds to a particular region in the active centre (the enzyme subsites; $S_4\text{-}S_3\text{-}S_2\text{-}S_1\text{--}S'_1\text{-}S'_2....$) and the long bond indicates the scissile bond of which P_1 provides the carboxyl moiety and P'_1 the amino moiety.[91]

CHAPTER TWO

RESULTS AND DISCUSSION

2.0 RESULTS AND DISCUSSION

2.1 INTRODUCTION

The design and synthesis of potential inhibitors and chromophoric inhibitors of the adenovirus proteinase is important as they can provide valuable tools for the study of the chemical mechanism by which the enzyme catalyses the various protein cleavages observed *in vivo*. From this information potential therapeutic drugs may then be designed and synthesised to block the enzyme and hence stop the virus replicating, effectively "killing" it.

2.2 IDENTIFICATION OF THE ENZYME

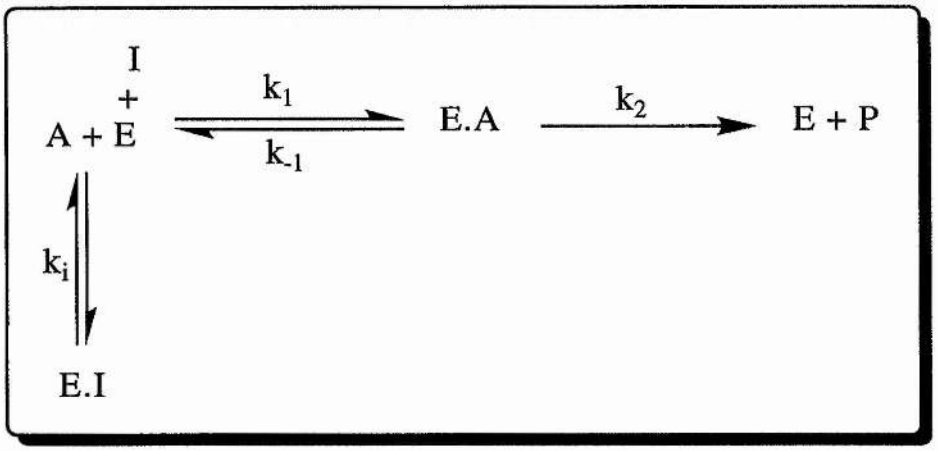
Over recent years this enzyme has been the subject of extensive research by virtue of its novel catalytic mechanism and structure.[43,44,68-70] Although it had been widely accepted that the L3 23 kDa viral polypeptide was the proteinase responsible for the viral polypeptide cleavages it was not until recently that this was confirmed.[74,75] The L3 23 kDa polyprotein had been assumed to be the proteinase on the basis that a mutation of the virus, Ad ts1, displayed no proteolysis.[73,92] This mutation had been mapped to the gene coding for the L3 23 kDa protein *i.e.* when the gene does not make the protein no proteolysis is observed. In addition to this, all previous studies on the proteinase had been carried out using unfractionated cell extracts and without purification of the proteinase. This left the possibility, therefore, that although the L3 23 kDa protein was required for proteolysis it may not itself be the proteinase. Kemp and Webster demonstrated unequivocally that this was the proteinase by expressing the proteinase in *E. coli*, purifying to homogeneity and showing that it cleaved wild-type substrates.[73] They further showed that the proteinase is active in its monomeric form. Underpinning this work was

the discovery of a virally encoded protein fragment on which the proteinase is dependent for activity.[71] Previously it had been demonstrated that pure proteinase was inactive whereas proteinase extracted from disrupted cells was active suggesting that a cofactor was required for activation. Following this Webster *et al.*, demonstrated that this cofactor was in fact a cleavage fragment of the viral polypeptide pVI.[71] This cofactor is the 11 amino acid sequence from the C-terminus of pVI, denoted pVI-CT, and contains a cysteine residue essential for activation. Further they demonstrated that it is the oxidised disulphide linked dimer of this peptide which is the activating factor and not the reduced monomer. Thus Kemp *et al.* were able to confirm that the L3 23 kDa protein is the proteinase and that it is activated by a precursor protein cleavage fragment. This also explains the control mechanism of the proteinase, the virion polypeptides are synthesised late in infection and the cytopathic effect due to the action of the Ad 2 proteinase occurs late in infection. As has been demonstrated in other viral proteinase systems, notably the HIV-1 proteinase, premature activation of the enzyme results in premature cleavage of precursor viral polypeptides which do not assemble coherently leading to premature cell death with a subsequent lack of efficient infectivity.[79,80,93] This disulphide linked dimer control mechanism represents a new type of activation mechanism suggesting a disulphide/-SH interchange between pVI-CT and thiols in the active site to expose the catalytic thiol and allow it, therefore, to adopt the proposed S^-/ImH^+ catalytic system.

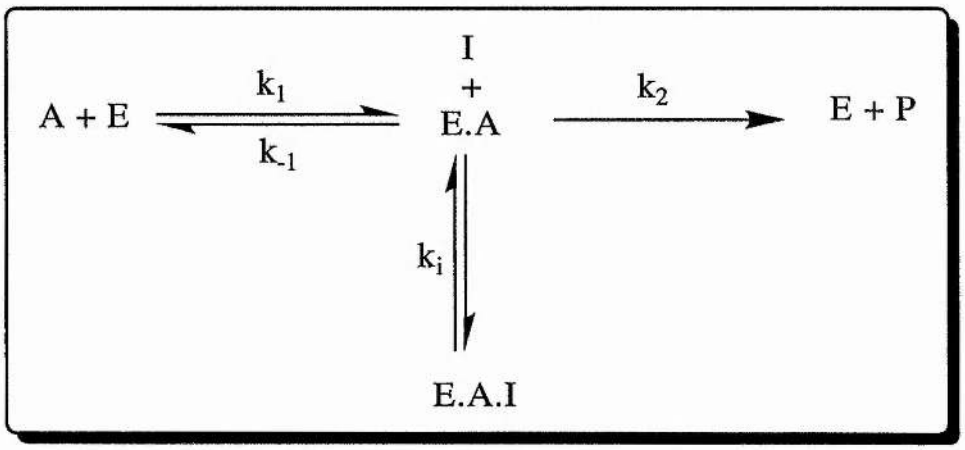
2.2.1 Modes of Inhibition

The four types of inhibition are, competitive, uncompetitive, noncompetitive and mixed inhibition.

When an inhibitor molecule I binds reversibly to the active site of the enzyme and therefore prevents a substrate molecule A binding and when A binds to the active site preventing I binding then I acts as a competitive inhibitor, *i.e.* I competes with A for a place in the active site.

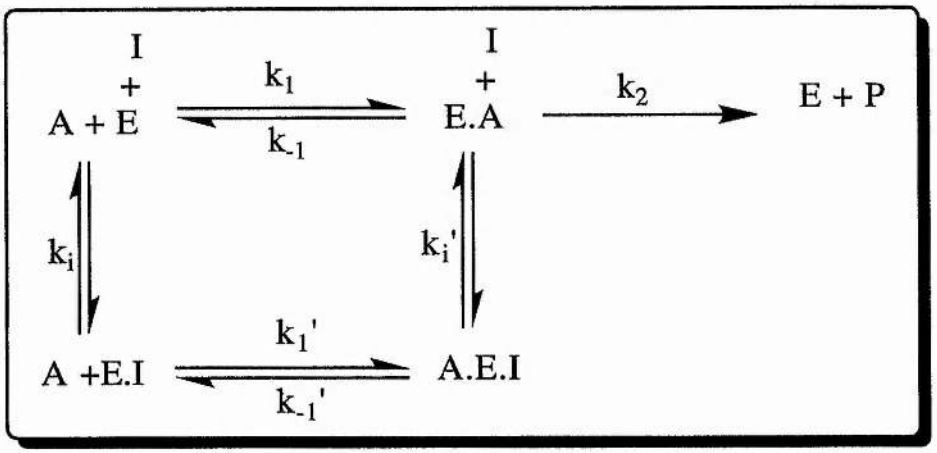


When an inhibitor molecule I binds reversibly to the enzyme-substrate, enzyme-intermediate or enzyme-product complex but not to the free enzyme then I acts as an uncompetitive inhibitor.

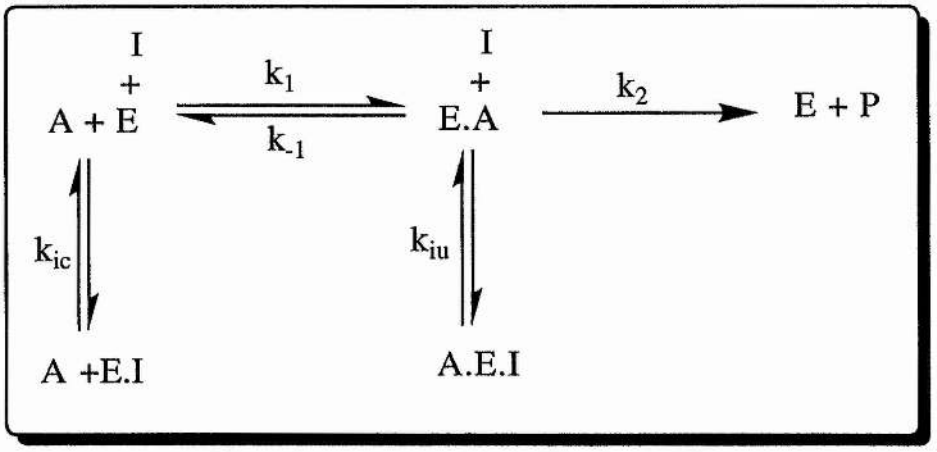


When an inhibitor molecule I can bind reversibly to the free enzyme or the enzyme-substrate complex and when the dissociation constant of A from the enzyme-substrate complex is unaffected whether an inhibitor molecule is

bound or not, then I acts as a noncompetitive inhibitor. This type of inhibition can occur when A and I bind simultaneously at different sites on the enzyme.



When I acts as a mixed inhibitor it displays an inhibiton pattern which is a mixture between that for competitive and uncompetitive inhibition.



2.3 DESIGN OF POTENTIAL INHIBITORS OF THE ENZYME

Before designing our inhibitors it was necessary to obtain an accurate picture of the active site, *i.e.* identify i) which cysteine residue is responsible for catalysis, ii) what other residues are involved and iii) what is their steric and electronic relationship to each other. Extensive site directed mutagenesis

(SDM) experiments have been performed in order to i) show further that this is a cysteine proteinase and ii) identify which of the cysteines is the active site cysteine.[74,75] Proteinase amino acid sequences from 12 adenovirus serotypes all display conserved residues at serine 95, serine 182, cysteine 104 and cysteine 122 and to a lesser extent serine 160 and cysteine 126/127 [See Figure 4, Page 39].[71,75] It had been proposed that serine 160, histidine 54 (the only conserved histidine) and aspartate 102 made up the catalytic triad, therefore, SDM experiments were performed in which all of the previously mentioned serine and cysteine residues were mutated, in turn, to alanine.[74,75] In addition the 2 conserved cysteines 104 and 122 were also changed to serines. The results clearly showed that mutations S95A, S182A and S160A all resulted in active enzyme with little or no effect on relative activity. The mutations C104A, C122A and C104S and C122S resulted in a proteinase with less than 5% of the wild-type activity. These results taken with the previous evidence confirmed that one of the conserved cysteines contributes the active site thiol. The absence of a thiol group at 104 or 122 interferes with the enzyme reaction mechanism either by preventing activation by the pVI-CT peptide or by abolishing the active site. Mutants containing C126A, C127A and the double mutant C126A/127A all resulted in active proteinase strongly suggesting that these semi-conserved residues do not contribute to the active centre or the activation mechanism. Attempts to determine which of the cysteine is the catalytic nucleophile and which is responsible for the disulphide activation were inconclusive although, interestingly Rancourt *et al.* observed an active proteinase with the mutation C122G.[74] Activation and catalysis, therefore, appear to be inextricably linked. The fact that the proteinase partially digests pTP in the absence of pVI-CT and requires a 200 fold excess of pVI-CT for activation suggests that the activating peptide is not directly involved in the catalytic cycle. It is more likely, therefore, to contribute to setting up the appropriate redox state of one or more of the essential cysteine residues. The

SDM experiments performed by Rancourt *et al.* using glycine agree in general with those of Webster *et al.* using alanine, with the notable exception that C122G retains 81% activity.[74] Activation by pVI-CT may promote a conformational change which in the C122G mutant is permitted by the small size of the glycine and hence some activity is observed. In addition, the mutant H54A resulted in an inactive proteinase supporting the proposal that H54 is required for catalysis.[74] In most cysteine proteinases the active site is composed of cysteine, histidine and aspartate although in papain only cysteine and histidine are directly involved.[83] Rancourt *et al.* failed to identify a third member for the active site, they also synthesised the serine mutant C104S and reported a 31% relative activity which further establishes a link with trypsin and poliovirus 3C proteinases which accept serine or cysteine as their active site nucleophiles.[94] Their mutation C122S resulted in a complete loss of enzyme activity further supporting the proposal that C104 is the active thiol.

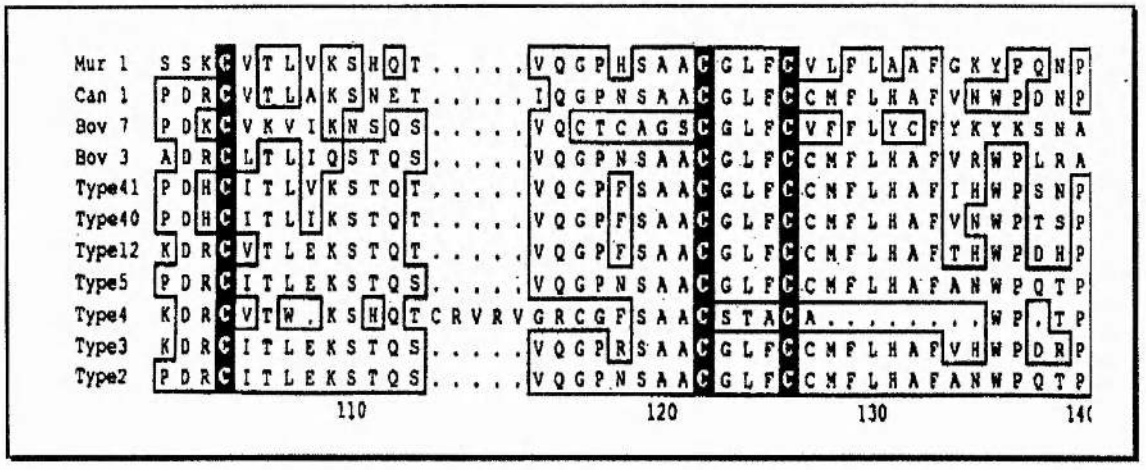


Figure 4: *Alignment of part of the proteinase sequence showing the 3 conserved cysteine residues. The figure shows residues 101-140 of the Ad2 L3 23 kDa protein aligned with the corresponding section of the 23 kDa sequence from human serotypes 3, 4, 5, 12, 40 and 41, bovine serotypes 3 and 7 plus the murine and canine serotypes 1. The 3 conserved cysteines are highlighted.*

Determining the reaction mechanism of an enzyme is central to the design of specific inhibitors of that enzyme and although the Ad 2 proteinase has been classified as a cysteine proteinase its mode of action is still somewhat uncertain. Webster *et al.* completed a comprehensive substrate specificity study which allows us to design potential inhibitors.[70] The L3 23 kDa proteinase is remarkably substrate specific and obtaining a substrate specificity profile was essential to further investigate the catalytic mechanism. It was shown that a four amino acid motif to the carboxyl terminal of the scissile bond is the minimum requirement for cleavage.[70] The major requirements in the primary structure of this tetrapeptide are at P_2 which must be a glycine and at P_4 which must be methionine, leucine or isoleucine. The restrictions on the types of scissile bonds cleaved P_1--P'_1 are somewhat less. Providing glycine occupies P'_1 almost any amino acid may occupy P_1 and be cleaved, except possibly proline. This was supported by molecular modelling which showed that the side chains of residues at P_1 point away from the intolerant sites of P_2 and P_4.[70] If any other amino acid occupies P'_1 then there are increased steric restrictions on P_1, indeed, in the natural substrates where glycine is not at the P'_1 position then in almost all cases glycine occupies the P_1 position. From these data two profiles may be drawn;

(M,L,I) X G G↓X and (M,L,I) X G X↓G

Using these profiles Webster *et al.* confirmed all known cleavages in the viral substrates and predicted unidentified cleavage sites in known substrates pVI, pVII, pVIII, pTP, IIIa as well as the penton, hexon, 52 kDa, early 21 kDa, 145 kDa and D172 proteins. This group has designed and tested a range of potential inhibitors based on the four amino acid motif in an attempt to further understand the catalytic mechanism of this enzyme [See Figure 5, Page 41].[76] The tetrapeptide derivatives were of the form;

R-LAG-(NH_2-CH_2)-R'

where R is the nitrogen protecting group Cbz or t-Boc and R' represents the downstream residue over the scissile bond. Various derivatives were synthesised as R'; ester, acid, amide, nitrile and dimethylacetal.

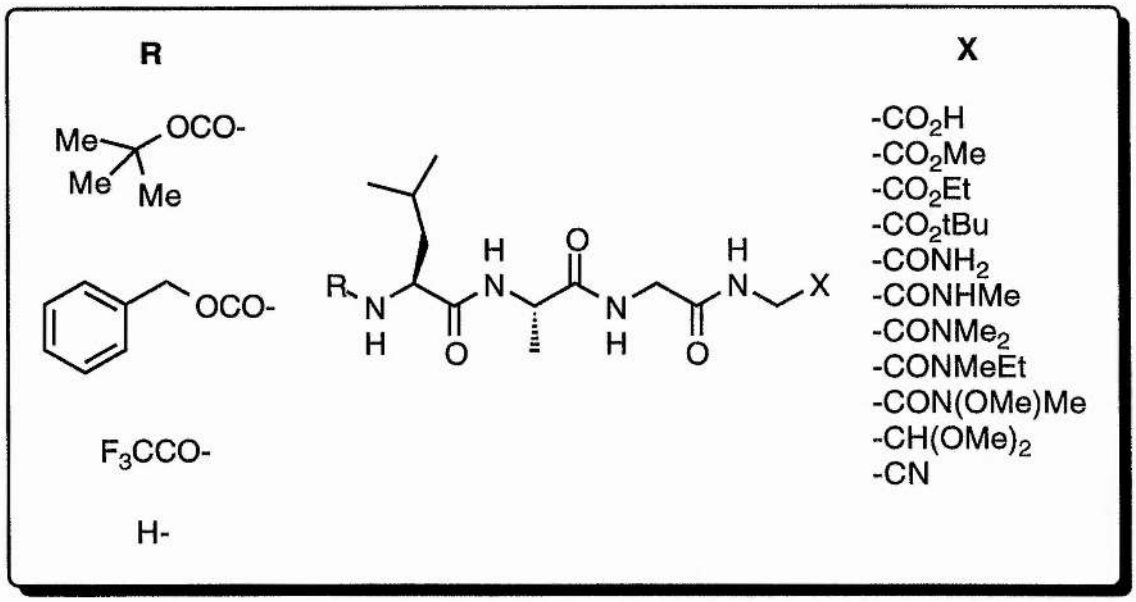
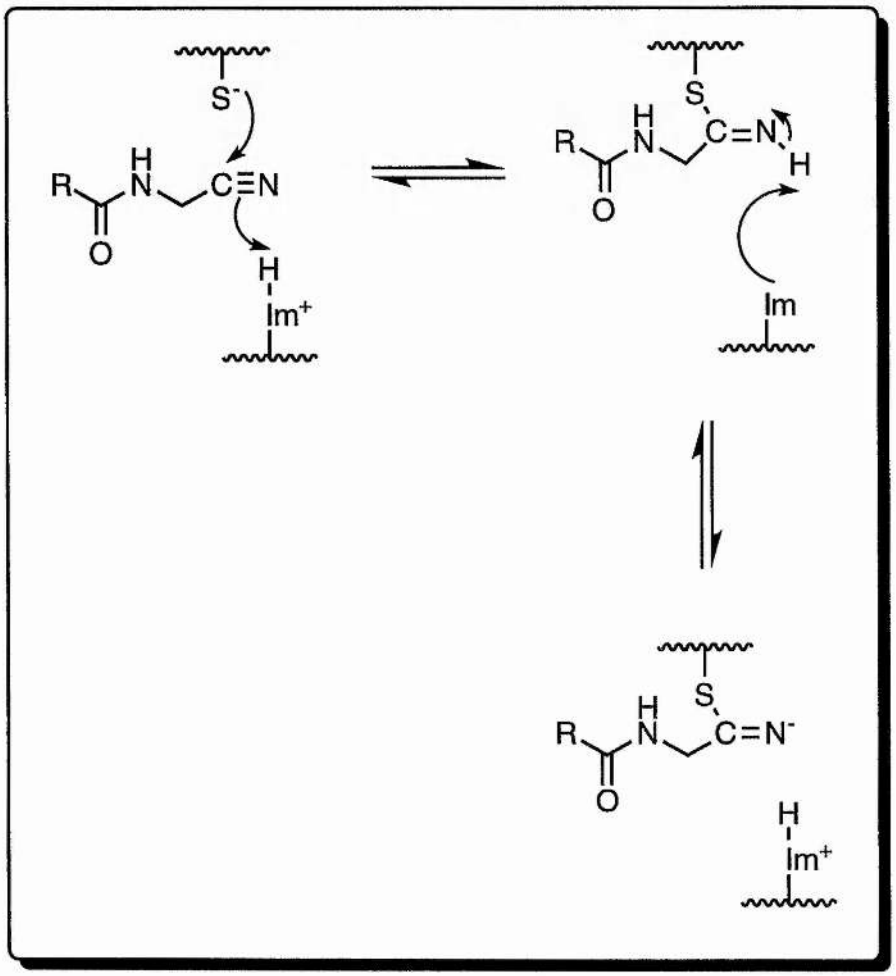


Figure 5: *Potential inhibitors synthesised and tested by our group.*

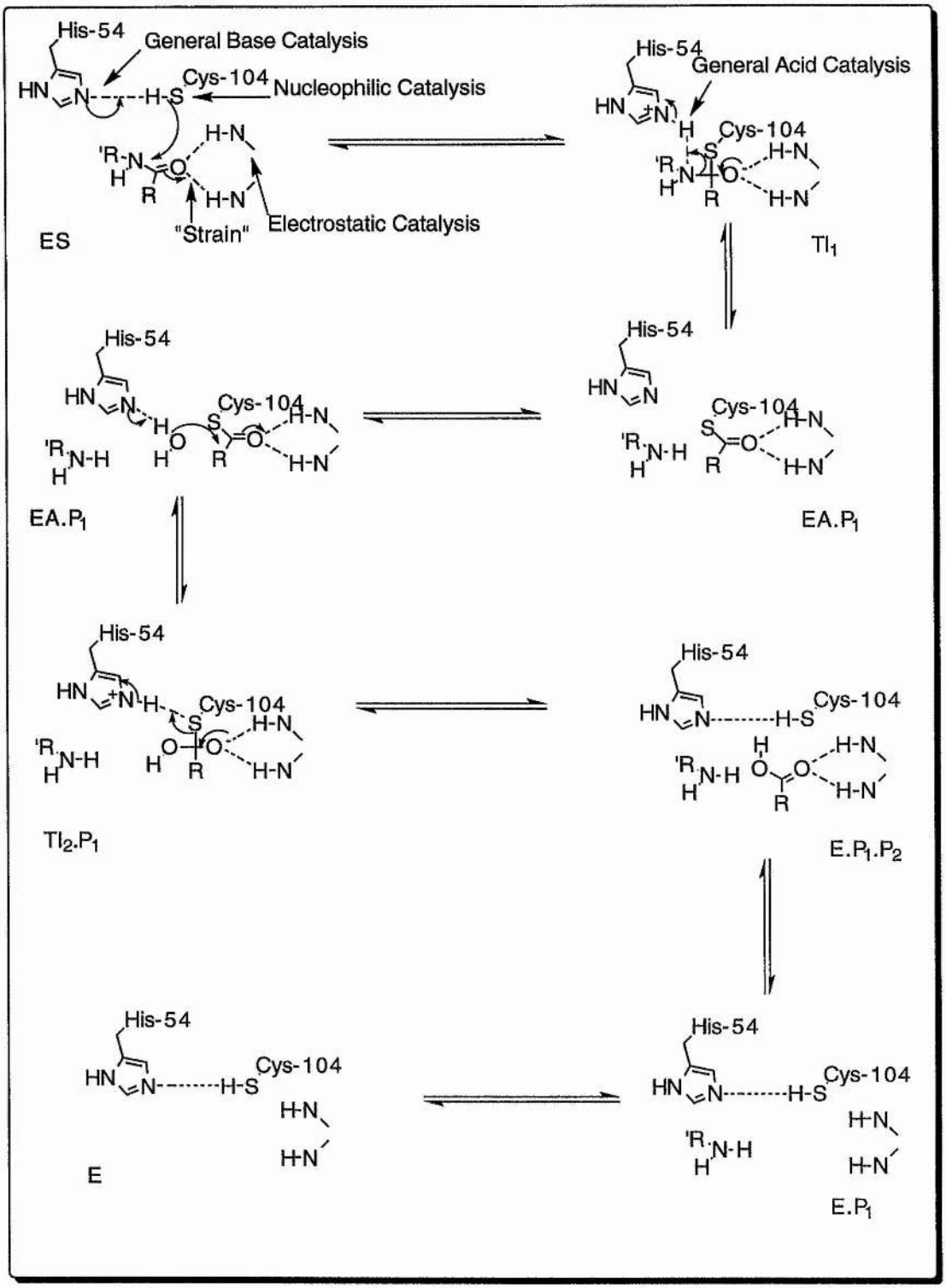
All the esters were substrates including the tertiary butyl ester, this was the only ester with a t-Boc N-protection and acted as a substrate at two sites. The fact that this enzyme hydrolyses t-butyl esters in itself is interesting since the usual manner in which these sterically hindered esters are hydrolysed is under acidic conditions. It was observed that a valine residue at P'_1 resulted in very slow cleavage and it was postulated that there is some form of tertiary structure *in vivo* which pulls the hydrophobic residue away from the scissile bond and allows efficient cleavage. The hydrolysis of the t-butyl ester may well be similar in that it seems probable that some hydrophobic interaction is set up which pulls or pushes the bulky group away from the carbonyl which

facilitates nucleophilic attack by the sulphur ion. The acids and amides did not act as inhibitors at all and were not pursued. The dimethylacetals and nitriles acted as moderate and good inhibitors respectively. The reciprocal plots of rate versus inhibitor concentration for the dimethylacetals indicated that more than one molecule of inhibitor can bind to the enzyme. The double reciprocal plots of initial rate versus substrate concentration were linear and intersected such that the apparent K_m was not affected *i.e.* acetal acts as a noncompetitive inhibitor. The nitrile derivatives all performed as good inhibitors and also acted noncompetitively [See Scheme 5, Page 42].



Scheme 5: *The nitrile as a reversible inhibitor and possible thioimidate adduct formation.*

It is interesting that these compounds act as noncompetitive inhibitors since they closely resemble the wild-type P_4-P_1 motif and would therefore be expected to act as competitive inhibitors. However, if one considers the order in which events take place at the active site and assume that after cleavage the N-terminal product is released first then another inhibitor molecule may enter and trap the C-terminal product on the enzyme. Coupled with its ability to bind to the same region in the absence of the C-terminal product then the inhibitor would act in a noncompetitive manner and be best described as a product analogue. If the C-terminal product desorbs first then the remaining N-terminal fragment would block the P_4-P_1 subsites against another inhibitor molecule entering. Therefore it seems probable that in contrast to many cysteine proteinases, including papain, this enzyme releases the N-terminal product first [See Scheme 6, Page 44]. All the compounds were expected to bind, the fact that they did not and that those that did all possess a form of "good leaving group" suggests a covalently bonded intermediate.



Scheme 6: *Alternative hydrolysis mechanism of the Ad L3 23 kDa proteinase.*

This is in accord with similar observations in other cysteine proteinases, the nitrile Ac-Phe-$NHCH_2CN$ inhibits papain forming a thioimidate adduct.[95] Obviously the chemical mechanism of this enzyme is not simple, one of the problems faced when studying the kinetics of such a system is that the assay systems are not continuous and therefore depend on removing aliquots at specific time intervals from the reaction mixture and analysing them separately. These systems are inherently subject to large degrees of experimental error. A continuous assay system abolishes these errors where a reaction mixture is analysed *in situ* without the need to remove aliquots.

An additional objective of our work was the development of a UV continuous assay system using the chromophore *p*-nitroaniline. However, the first step was to synthesise the ester analogues of the *p*-nitroanilides and to use them as inhibitor probes to assess whether or not these types of compounds will be turned-over or not. In these systems where R'= *p*-nitroaniline, appearance of *p*-nitroaniline, the cleavage product, is monitored, hence initial rates, v_{max}'s and K_m's may be determined directly. Compounds synthesised and tested were of the form;

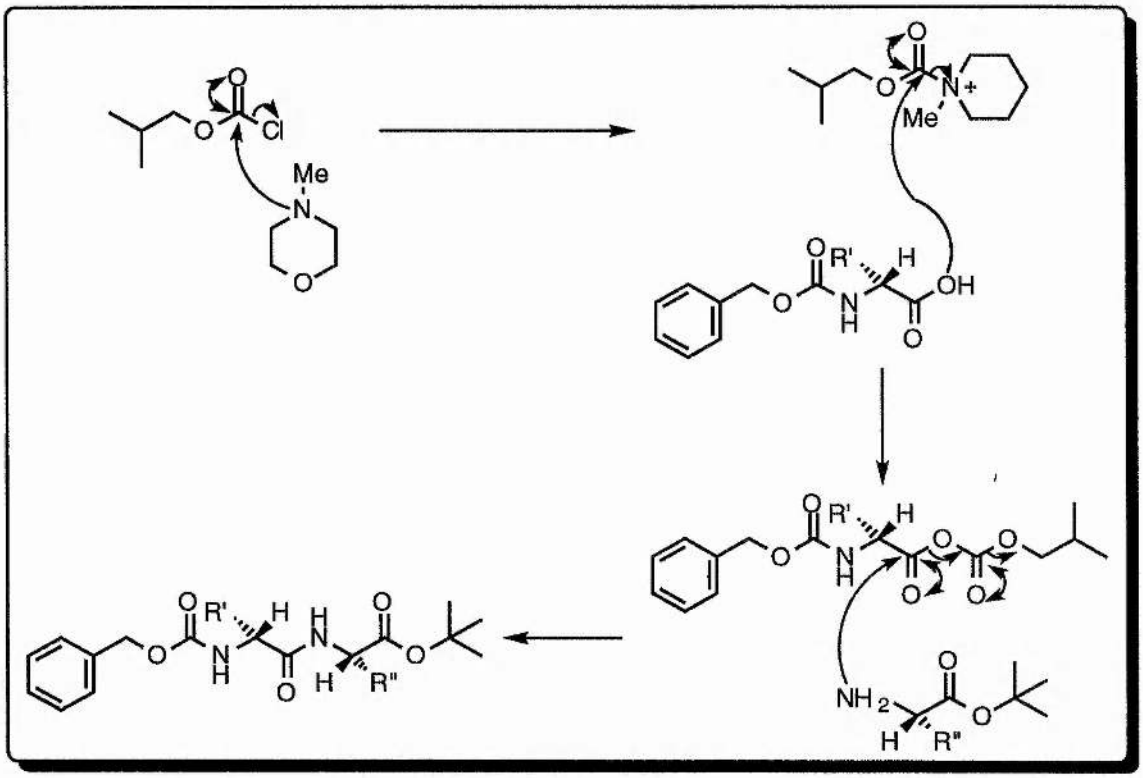
R-LAXG-R'

where R= Cbz; X= G, (2*S*)-A, (2*R*)-A, (2*S*)-V; R'= OPr^i, *p*-nitroaniline.

2.4 SOLUTION PHASE PEPTIDE SYNTHESIS

There are numerous methods for synthesising peptides in the solution phase, including the use of acid chlorides or carbodiimides for peptide couplings.

We synthesised peptide fragments and final peptides using the mixed anhydride method of peptide coupling.[96] This method was chosen for its advantages in reaction time, yield, ease of purification and quality of the products. The mixed anhydride method has been studied in detail to increase yield and minimise the possibility of racemisation. The reaction uses alkyl chloroformates as peptide forming reagents. A suitably protected amino acid reacts with the alkyl chloroformate in the presence of a base, to form an anhydride. A second suitably protected amino acid is then added to the reaction mixture which then reacts with the anhydride to form the peptide linkage [See Scheme 7, Page 46].



Scheme 7: *Reaction mechanism of the mixed anhydride peptide coupling.*

It is believed that the first step of the reaction is the attack of the tertiary base on the alkyl chloroformate to form a quaternary intermediate which in turn reacts with the carboxylic acid. An important consideration is that of

racemisation, which can be minimised depending on the steric bulk of the base and the choice of solvent. It has been found that methylamine reacts the fastest, but results in extensive racemisation. Base, solvent, temperature and alkyl chloroformate variations have been examined, it was found that the best results were obtained using a combination of N-methylmorpholine (NMM), isobutylchloroformate and dry tetrahydrofuran (THF). This combination of reagents was used in the course of this work. However, when there was a problem of solubility of one or more of the starting materials, a minimum volume of dry dimethylformamide (DMF) in THF was used to aid the solubility of the reactants.

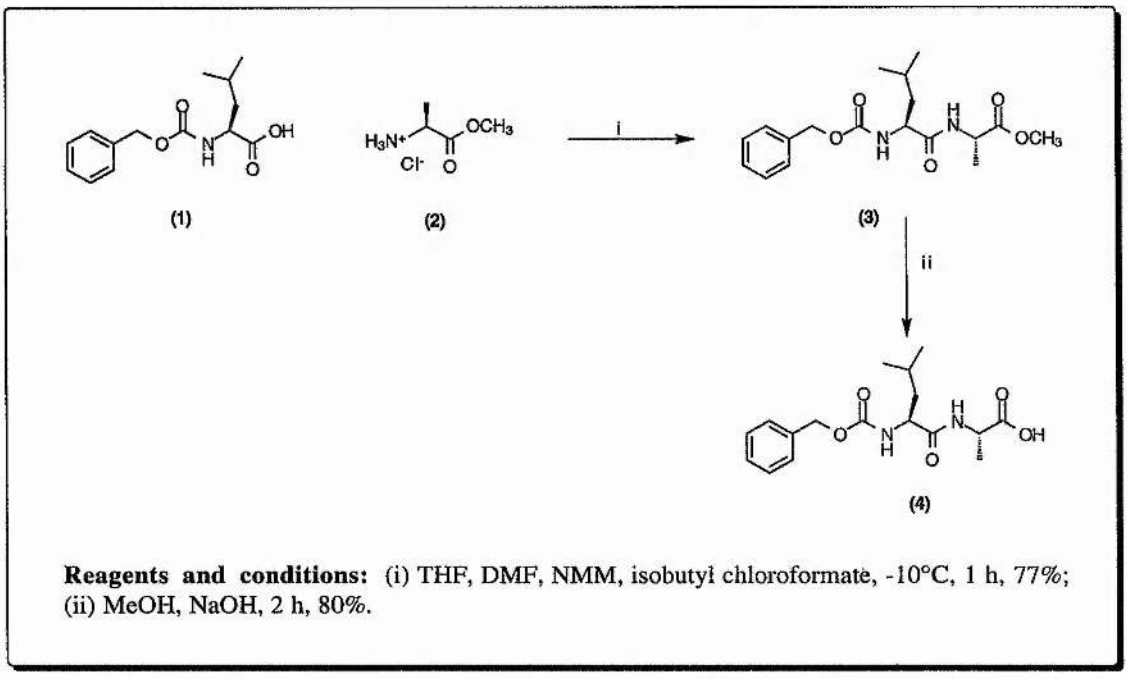
2.5 SYNTHESIS OF THE TETRAPEPTIDE ESTERS

We chose to synthesise these compounds from the two appropriately protected dipeptides which would be prepared from the monoprotected amino acids. This method was chosen rather than the sequential adding of residues as dipeptides tend to be easier to handle in terms of recrystallisation, chromatography and solubility. Also the yield from coupling two dipeptides is significantly higher than that of coupling a tripeptide and a monprotected amino acid. This approach gave us the common dipeptide precursor on the N-terminus; Cbz-Leu-Ala-OH, to which we coupled the varying other half; NH-X-Gly-Pr^{i}, to give the tetrapeptides desired.

2.5.1 Synthesis of the Dipeptide Precursor Cbz-Leu-Ala-OH (4)

Carboxybenzyl (Cbz) protection of (2*S*)-leucine was achieved using benzyl chloroformate and aqueous sodium hydroxide under Schotten-Baumann[97] conditions to give Cbz-(2*S*)-leucine (**1**) in 96% yield {υ_{max} 1720 cm^{-1} urethane CO}. Methyl-(2*S*)-alaninate hydrochloride (**2**) was prepared from

(2*S*)-alanine, thionyl chloride and methanol in 62% yield {m.p. 110-2 °C (lit.[98], 109-110 °C)}. The two monoprotected amino acids (**1**) and (**2**) were coupled using mixed anhydride methodology to give methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alaninate (**3**) in 77% yield after recrystallisation from ethyl acetate/ light petroleum {m.p. 91-2 °C (lit.[99], 92-3 °C)}. Treatment of the dipeptide (**3**) with 1 mol dm^{-3} sodium hydroxide at 35 °C for 3 h, followed by acidification gave carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanine (**4**) in 80% yield {m.p. 149-150 °C (lit.[100], 150-1 °C)} [See Scheme 8, Page 48].



Reagents and conditions: (i) THF, DMF, NMM, isobutyl chloroformate, -10°C, 1 h, 77%; (ii) MeOH, NaOH, 2 h, 80%.

Scheme 8: *Synthesis of dipeptide precursor (4).*

2.5.2 Synthesis of the Various Carboxyl Terminal Dipeptides H-X-Gly-OPr^{i}

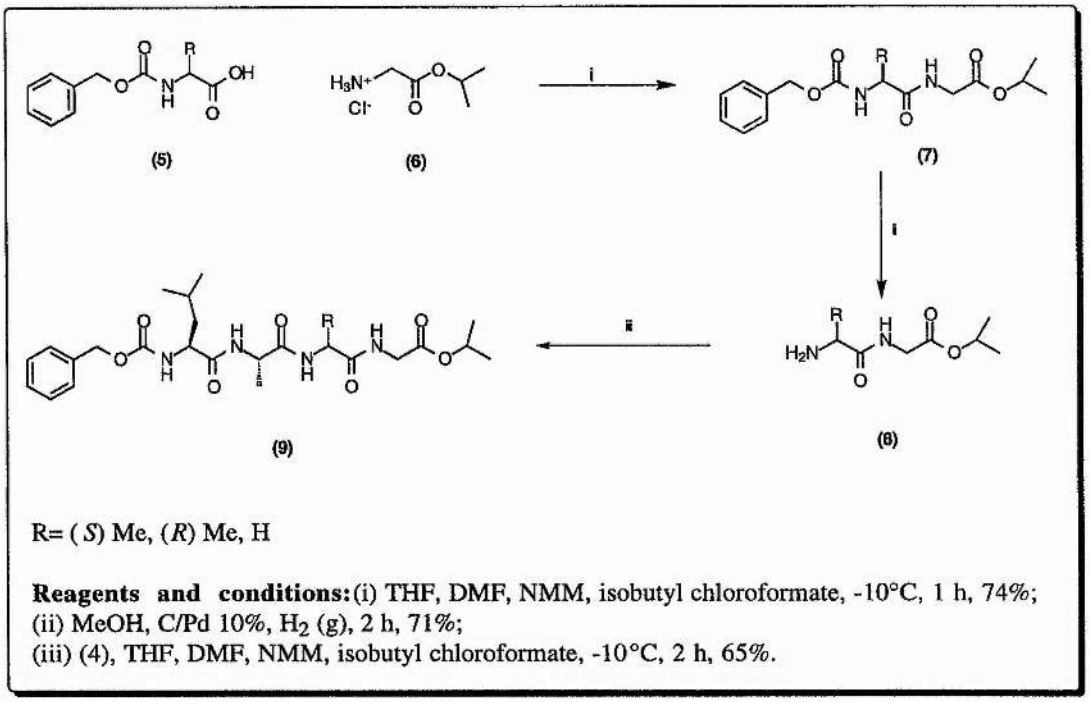
Carboxybenzyl (Cbz) protection of glycine was achieved in a manner identical to (**1**) to give carboxybenzyl glycine (**5**) in 55% yield {m.p. 119 °C, lit.[101], 120 °C}. Isopropyl glycinate hydrochloride (**6**) was prepared in a manner identical to (**2**) in 86% yield {m.p. 95-7 °C, lit.[102], 96 °C}. The two

monoprotected amino acids **(5)** and **(6)** were coupled using mixed anhydride methodology in a manner identical to that for **(3)** to give isopropyl carboxybenzyl glycyl glycinate **(7)** in 74% yield {m.p. 92-4 °C}. Removal of the nitrogen protecting group from the dipeptide **(7)** was effected by catalytic hydrogenolysis (H_2, Pd/C) over 2 h, followed by removal of the by-products under reduced pressure to give isopropyl glycyl glycinate **(8)** in 71% yield {m.p. 120 °C (decomp.)}.[103] Two further dipeptides, isopropyl-(2*S*)-alanyl glycinate and isopropyl-(2*R*)-alanyl glycinate, were synthesised in a manner identical to that for **(8)** starting with carboxybenzyl-(2*S*)-alanine and carboxybenzyl-(2*R*)-alanine respectively.

2.5.3 Coupling of the Dipeptide Fragments

The three tetrapeptide esters; isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycyl glycinate **(9)**, isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanyl glycinate **(10)** and isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanyl glycinate **(11)** were all prepared in the same manner. Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycyl glycinate **(9)** was prepared by the coupling of **(4)** and **(8)** using mixed anyhdride methodology in 65% yield after recrystallisation from methanol {Found: C, 58.75; H, 7.6; N, 11.45. Calc. for $C_{24}H_{36}N_4O_7$: C, 58.5; H, 7.4; N, 11.4%}. Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanyl glycinate **(10)** was prepared in the same manner from **(4)** and isopropyl-(2*S*)-alanyl glycinate in 74% yield after recrystallisation from methanol/ water {Found: C, 59.1; H, 7.55; N, 10.85. Calc. for $C_{25}H_{38}N_4O_7$: C, 59.25; H, 7.55; N, 11.05%}. The (2*R*) analogue of **(10)**, isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanyl glycinate **(11)** was prepared from **(4)** and isopropyl-(2*R*)-alanyl glycinate in 46% yield after recrystallisation from methanol/ water {Found: C, 59.25; H, 7.45; N,

10.75. Calc for $C_{25}H_{38}N_4O_7$: C, 59.25; H, 7.55; N, 11.05%} [See Scheme 9, Page 50].



Scheme 9: *Synthesis of tetrapeptide esters.*

2.6 SYNTHESIS OF THE TETRAPEPTIDE *p*-NITROANILIDES

The synthesis of this series of compounds was simplified considerably by building the tetrapeptides sequentially. This was because of the reactivity of the *p*-nitroanilide group which is both acid and base labile. Glycine-*p*-nitroanilide was, therefore, coupled to a series of tripeptide free acids. The common precursor to all the tripeptides was the dipeptide Cbz-Leu-Ala-OH (**4**) and, as for the tetrapeptide esters, the residues which varied at P_2 where Gly, (2*S*)-Ala and (2*R*)-Ala. The attempted synthesis of the (2*S*)-Val analogue was not successful. It was hoped that using these compounds we could have developed a continuous UV assay system for this enzyme using the *p*-nitroaniline chromophore.

2.6.1 Synthesis of the Tripeptide Cores

The three tripeptides; carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycine **(13)**, carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanine **(15)** and carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanine **(17)**, were all prepared in the same manner as **(3)** [See Figure 6, Page 52]. The two monoprotected species **(4)** and **(6)** were coupled using mixed anhydride methodology to give isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycinate **(12)** in 55% yield after purification by silica chromatography {m.p. 137 °C}. Treatment of **(12)** with 1 mol dm^{-3} sodium hydroxide at 35 °C for 2 h, followed by acidification gave carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycine **(13)** in 83% yield {m.p. 149 °C}. Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alaninate **(14)** was prepared in a manner identical to **(12)** from **(4)** and **(2)** in 68% yield after purification by silica chromatography {m.p. 174-6 °C}. Treatment of **(14)** with 1 mol dm^{-3} sodium hydroxide at 35 °C for 2 h, followed by acidification gave carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanine **(15)** in 85% yield {m.p. 183-4 °C}. Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alaninate **(16)** was prepared as for **(12)** from **(4)** and methyl-(2*R*)-alaninate hydrochloride in 66% yield after purification by silica chromatography {m.p. 164 °C}. Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanine **(17)** was prepared in a manner identical to that for **(13)** in 54% yield {m.p. 170 °C}. A further tripeptide synthesised was methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-valinate **(18)** in the same manner as **(12)** but the base hydrolysis of the methyl ester was unsuccessful using a variety of sodium hydroxide concentrations, molar ratios and reaction temperatures, so this derivative was not pursued.

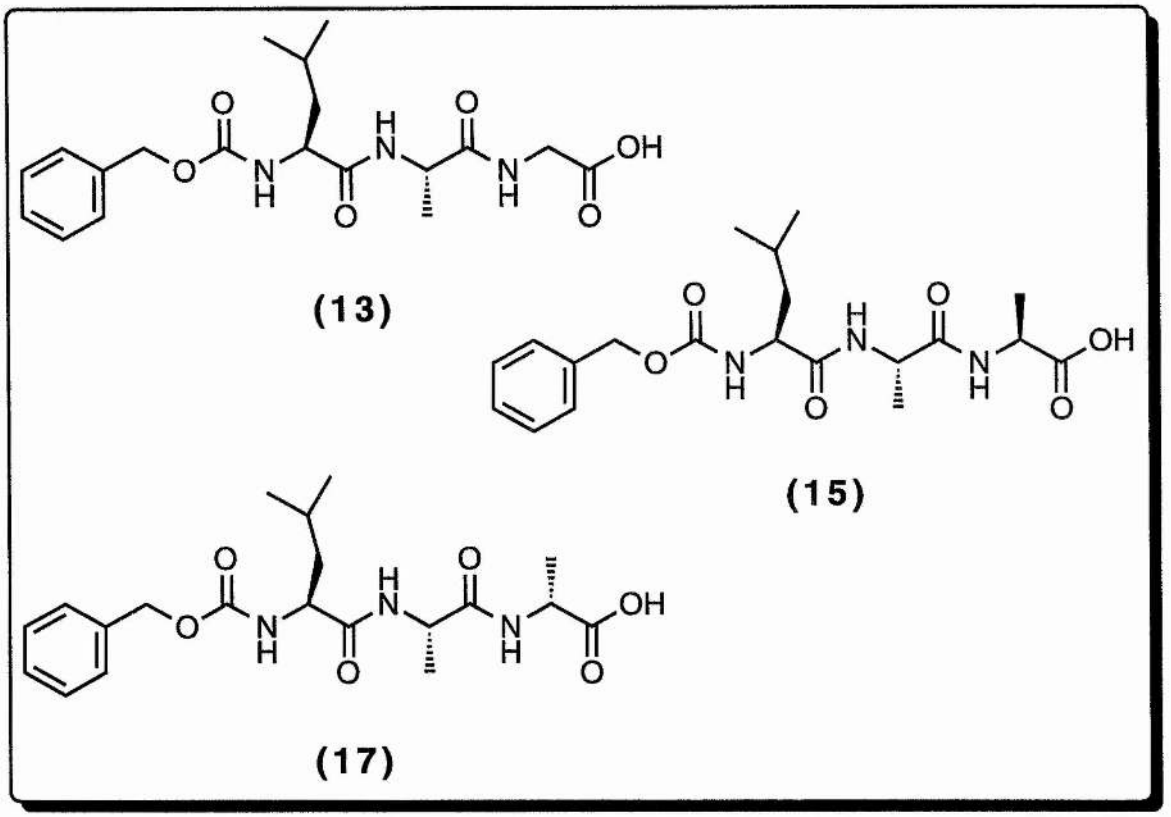
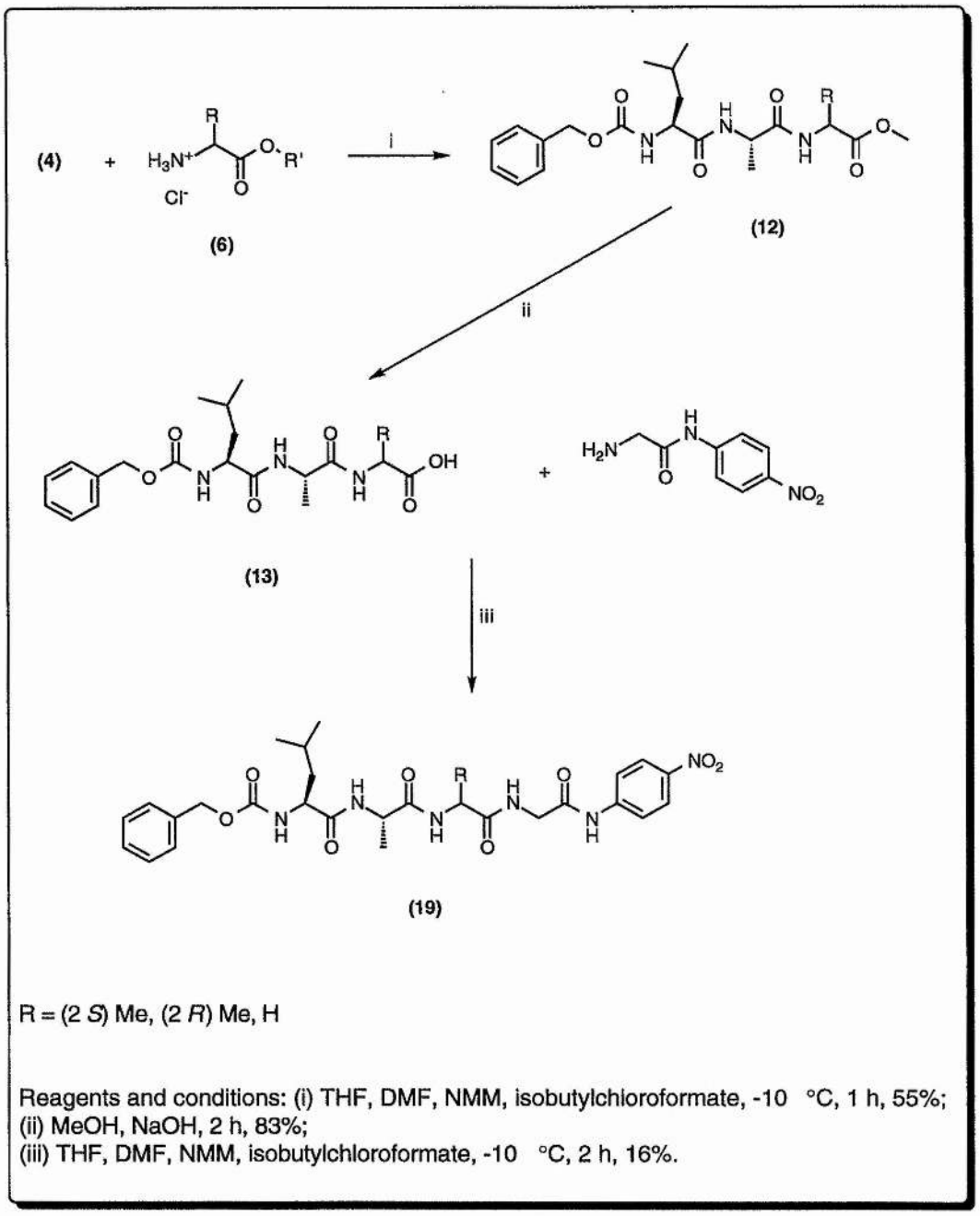


Figure 6: *The three tripeptide cores (13), (15) and (17).*

2.6.2 Synthesis of the Tetrapeptide *p*-Nitroanilide Derivatives

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycyl glycine-*p*-nitroanilide (**19**) was prepared from (**13**) and glycine-*p*-nitroanilide using mixed anhydride methodology. Although the δ_H NMR showed that reaction had occured {δ 7.83 (2H, d, p-nitroanilide protons 2 & 6) and δ 8.22 (2H, d, p-nitroanilide protons 3 & 5)}, the yield of the reaction was very low (16%). Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanyl glycine-*p*-nitroanilide (**20**) was prepared in a manner identical to that for (**19**) from (**17**) and glycine-*p*-nitroanilide in a moderate yield of 42% (Found: C, 57.2; H, 6.1; N, 14.05. Calc for $C_{28}H_{36}N_6O_8$: C, 57.55; H, 6.2; N, 14.35%) [See Scheme 10, Page 53]. Due to the lack of time the synthesis of the other tetrapeptide *p*-nitroanilides, carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanyl glycine-*p*-

nitroanilide and carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-valyl glycine-*p*-nitroanilide, was not completed.



Scheme 10: *Synthetic route to the tetrapeptide p-nitroanilides*

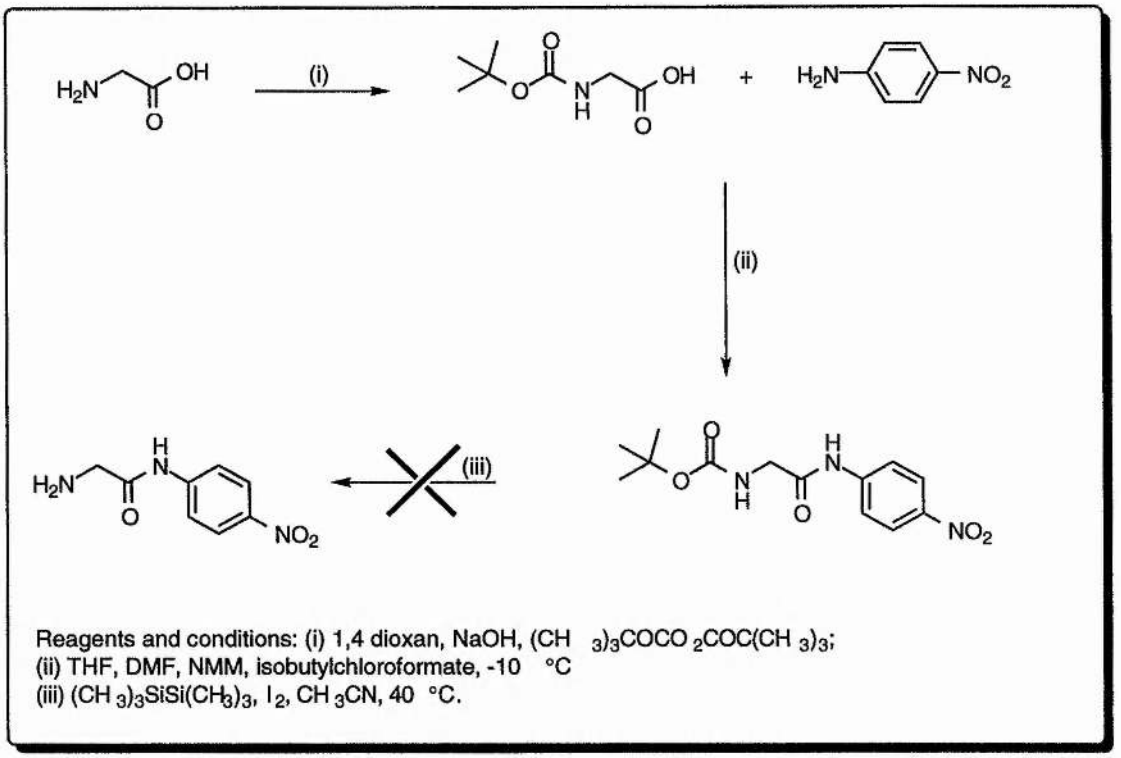
2.6.3 Proposed Synthetic Route to Glycine-*p*-Nitroanilide

The synthetic problem associated with amino acid *p*-nitroanilide derivatives is the electron deficient amide that is susceptible to nucleophilic attack on the carbonyl. During a standard synthesis of peptides which involves protections using aqueous base, deprotections using aqueous base and couplings which involve three separate nucleophilic attacks on electron deficient carbonyls it is impossible that this labile group should not be hydrolysed. Glycine-*p*-nitroanilide is a relatively expensive compound to buy, *p*-nitroaniline, however, is not and the possibility of synthesisng glycine-*p*-nitroanilide was investigated.

A literature method for the deprotection of N-^{t}Boc groups which does not require the relatively harsh conditions of aqueous base such as sodium hydroxide, but uses trimethyl silyl iodide would be a possible solution.[104,105] According to the literature this method proceeds under very mild conditions with very high yields and, in the example quoted ^{t}Boc-Gly-OBzl, the reaction is completed within 20 min without hydrolysing the ester.[105] This appeared to be an attractive alternative and was, therefore, pursued further.

^{t}Butoxycarbonyl glycine was prepared from di-^{t}butyldicarbonate, glycine and aqueous sodium hydroxide in 1,4 dioxane in good yield and reacted with *p*-nitroaniline again using mixed anhydride methodology. This method was chosen for its simplicity and speed of reaction although the aromatic amine is not a good nucleophile and an extra equivalent of NMM was added to assist in the removal of the amine proton. The reaction went with moderate yield and the product was purified by recrystallisation from methanol and carried on to the next step. The literature method was followed for the removal of the tertiary butoxycarbonyl group [See Scheme 11, Page 55].[105] There was no

evidence in the ^1H NMR that reaction had occured and following several further unsuccessful attempts this synthetic route was abandoned. It is unclear why this was not successful.



Scheme 11: *Synthetic strategy towards glycine-p-nitroanilide*

2.7 BIOLOGICAL TESTING OF THE TETRAPEPTIDE ESTERS

The series of three tetrapeptide esters were tested in an enzyme assay as potential inhibitors. The fact that the enzyme cleaves esters as substrates was also established by ourselves during the design and synthesis of potential inhibitors. To date the esters tested were of the form Cbz-Leu-Ala-Gly-Gly-OR, where R= Me, Et, ^{t}Bu.[76] The compounds in our study are all isopropyl esters and the amino acid sequence has been varied at P_2. Kemp *et al.* reported that they observed no cleavage of synthetic viral peptides when the residue at P_2 was (2*S*)-alanine or larger and there seems to be a very strict steric specificity at this residue.[70] We have, therefore, synthesised the "wild-type"

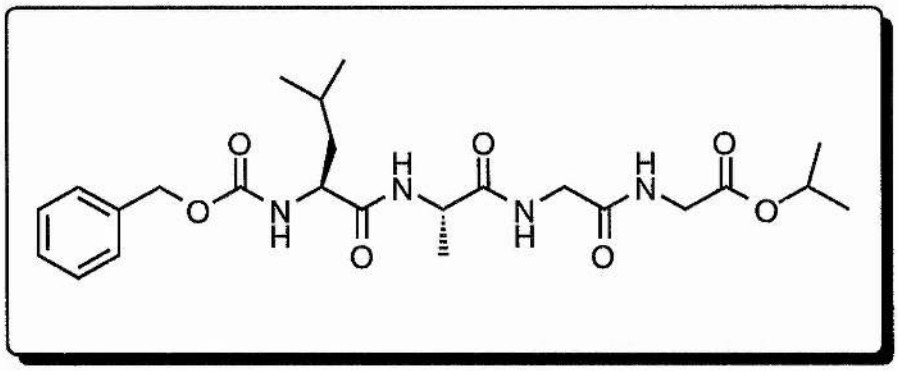
LAGG, LA-(2*S*)-AG and LA-(2*R*)-AG ester derivatives in order to probe the steric properties around this P_2 site since it may be that there is a gap on the other side to the (2*S*)- stereochemistry and that the (2*R*)-alanine derivative may be turned over.

2.7.1 Peptide Assay for Recombinant Adenovirus 2 L3 23kDa Enzyme

The assay used is a modification of the one developed by Kemp *et al.* during their work to identify all of the viral peptide cleavage sites confirming their work on substrate specificity [See Appendix 1, Page 92].[69,70]

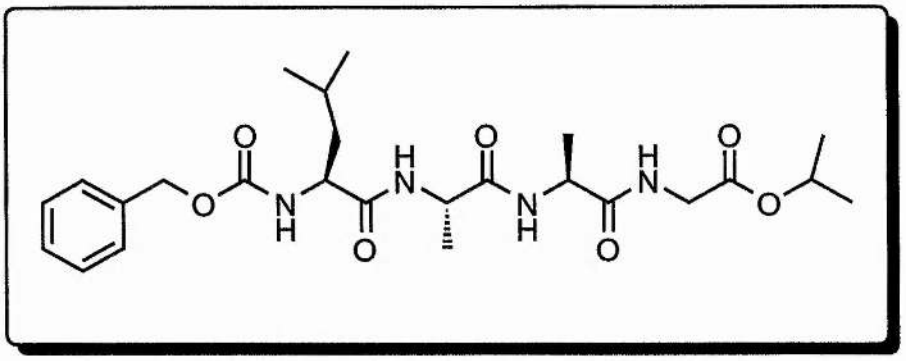
2.7.2 Results

Cbz-(2*S*)-Leu-(2*S*)-Ala-Gly-Gly-OPr^i (**9**);



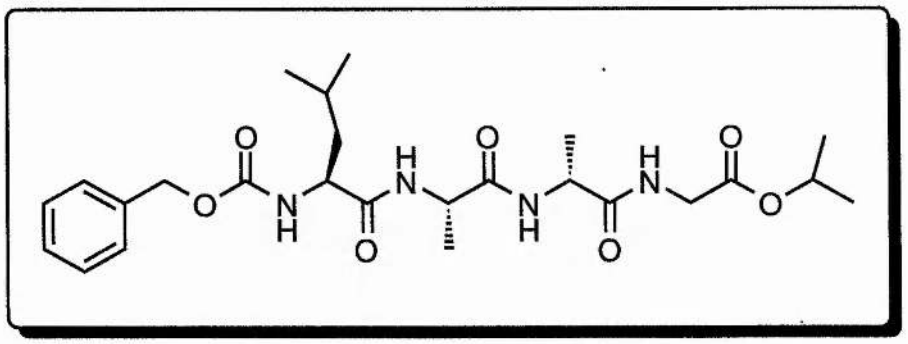
(**9**) was assayed at concentrations of 0.006 mM and 0.031 mM in DMSO and at both concentrations there was no observable inhibition of the proteinase. We would not expect this compound to inhibit since it is based on the tetrapeptide motif and therefore mimics the "wild-type" peptide substrates. Indeed know that wild-type esters are substrates.[76]

Cbz-(2*S*)-Leu-(2*S*)-Ala-(2*S*)-Ala-Gly-OPr^{i} (**10**):



This tetrapeptide differs from the "wild-type" at P_2. The glycine residue has been replaced by (2*S*)-alanine, as reported by Kemp *et al.* this is not a natural substrate and does not fit their motif.[70] A factor which was detrimental to the testing of this compound was the fact that it required a fairly high concentration of methanol to dissolve the peptide (7.7% MeOH in the assay). **(10)** was assayed at concentrations of 0.15 mM and 0.77 mM. At the lower concentration of 0.15 mM no inhibition was observed. At the higher concentration of 0.77%, inhibition of 37.41% was observed even after the inhibition due to the high methanol concentration had been taken into consideration (20% inhibition of initial rate due to a methanol concentration of 7.7% in the assay). This is very poor and **(10)** is certainly not a potent inhibitor, the estimated K_i for this inhibition is 2.1 mM.

Cbz-(2*S*)-Leu-(2*S*)-Ala-(2*R*)-Ala-Gly-OPr^i **(11)**:



(11) was also assayed at concentrations of 0.15 mM and 0.77 mM and with the same high concentration of methanol (7.7% in the assay and 20% inhibition of the initial rate). As for **(10)**, **(11)** showed no inhibition at the lower concentration. At the higher concentration **(11)** showed 14.64% inhibition and the estimated K_i was 5.3 mM, this is considerably lower than for **(10)** and again **(11)** is not a potent inhibitor. The lower percentage inhibition and higher estimated K_i may be indicative that this (2*R*)-Ala analogue has lessened the steric restriction around the P_2 subsite and therefore, appears more like the "wild-type" substrates. Kemp *et al.* showed, using molecular modelling of proposed and actual cleavage sites in the viral polypeptide precursors, that only sequences that i) hold the sidechain of the residue at P'_1 away from the scissile bond are cleaved and ii) that even the sidechain of alanine at P_2 protrudes enough in terms of the secondary structure to prevent access to the proteinase.[70] This tends to indicate that, whilst the (2*R*) analogue has indeed lessened the steric restriction slightly, these steric restrictions are much more severe that at first presumed and indeed "claustrophobic" particularly around the P_2-S_2 site.

See Experimental, page 80 and Appendix One, page 92, for full assay details and results.

CHAPTER THREE

EXPERIMENTAL PROCEDURE

3.0 EXPERIMENTAL PROCEDURE

Elemental microanalyses were performed in the departmental microanalytical laboratory.

NMR spectra were recorded on a Bruker AM-300 (300 MHz; f.t. ^{1}H-NMR, and 74.76 MHz; ^{13}C-NMR), or a Varian gemini 200 (200 MHz; f.t. ^{1}H-NMR and 50.31 MHz; ^{13}C-NMR) spectrometers. ^{1}H-NMR spectra are described in parts per million downfield shift from TMS and are reported consecutively as position (δ_H or δ_C), relative integral, multiplicity (s -singlet, d -doublet, t -triplet, q -quartet, dd -double of doublets, sep -septet, m -multiplet, and br -broad), coupling constant (Hz) and assignment (numbering according to the IUPAC nomenclature for the compound). ^{1}H-NMR were referenced internally on ^{2}HOH (4.68 ppm), $CHCl_3$ (7.27 ppm) or DMSO (2.47 ppm). ^{13}C-NMR were referenced on CH_3OH (49.9 ppm), $C^{2}HCl_3$ (77.5 ppm), or DMSO (39.70 ppm).

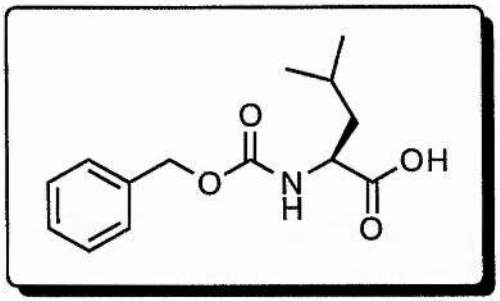
IR spectra were recorded on a Perkin-Elmer 1710 f.t. IR spectrometer. The samples were prepared as Nujol mulls, solutions in chloroform or thin films between sodium chloride discs. The frequencies (υ) as absorption maxima are given in wavenumbers (cm^{-1}) relative to a polystyrene standard. Mass spectra and accurate mass measurements were recorded on a VG 70-250 SE, a Kratos MS-50 or by the SERC service at Swansea using a VG AZB-E. Fast atom bombardment spectra were recorded using glycerol as a matrix. Major fragments were given as percentages of the base peak intensity (100%). UV spectra were recorded on Pye-Unicam SP8-500 or SP8-100 spectro-photometers.

Flash chromatography was performed according to the method of Still *et al.*[106] using Sorbsil C 60 (40-60 μm mesh) silica gel. Analytical thin layer chromatography was carried out on 0.25 mm precoated silica gel plates (Macherey-Nagel SIL g/UV254) and compounds were visualised using UV fluorescence, iodine vapour, ethanolic phosphomolybdic acid, or ninhydrin.

Melting points were taken on an Electrothermal melting point apparatus and are uncorrected. Optical measurements were measured at 23 °C on a Optical Activity AA-1000 polarimeter using 10 or 20 cm path length cells; $[\alpha]_D$ values are given in 10^{-1} deg cm^2 g^{-1}.

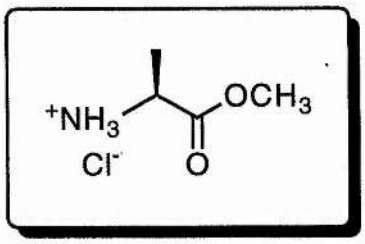
The solvents used were either distilled or of analar quality and light petrol ether refers to that portion boiling between 40 and 60 °C. Solvents were dried according to literature procedures. Ethanol and methanol were dried using magnesium turnings. Isopropanol, isopentanol, DMF, toluene, CH_2Cl_2, acetonitrile, diisopropylamine, triethylamine and pyridine were distilled over CaH_2. THF and diethyl ether were dried over sodium/ benzophenone and distilled under nitrogen. Thionyl chloride was distilled over sulphur, and the initial fractions were always discarded. BuLi was titrated according to the method of Lipton.[107]

Carboxybenzyl-(2*S*)-leucine (1)



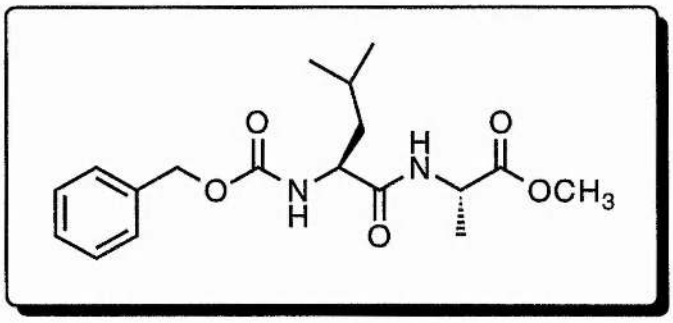
To a vigorously stirred solution of (2*S*)-leucine (3.03 g, 23 mmol) in water (80 cm^3) containing sodium hydroxide (3.2 g, 80 mmol) was added benzyl chloroformate (5.4 cm^3, 37 mmol) in 5 portions over 15 min. The reaction mixture was stirred for 3 h at 0 °C and then extracted with diethyl ether (2 x 50 cm^3). The aqueous phase was acidified with concentrated hydrochloric acid and extracted with diethyl ether (3 x 50 cm^3). The ethereal extracts were washed with brine (50 cm^3), dried (Na_2SO_4) and concentrated under reduced pressure to give (**1**) as a thick yellow oil (5.8 g, 96%), υ_{max} (thin film)/cm^{-1} 3320 (NH), 3090-3040 (aromatic), 2960-2290 (aliphatic), 1720 (urethane CO) and 1710 (acid CO); δ_H (200 MHz; C^2HCl_3) 0.97 (6H, d, $CH(C\underline{H}_3)_2$), 1.52-1.79 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 4.38 (1H, t, $C\underline{H}CO_2H$), 5.11 (2H, s, $C_6H_5C\underline{H}_2$) 5.29 (1H, d, NH), 7.36 (5H, m, C_6H_5) and 9.5 (1H, br, CO_2H); δ_C (50.3 MHz; C^2HCl_3) 22.20 & 23.34 ($CH(\underline{C}H_3)_2$), 25.25 ($\underline{C}H(CH_3)_2$), 41.94 ($\underline{C}H_2CH(CH_3)_2$), 52.88 ($\underline{C}HCO_2H$), 67.69 ($C_6H_5\underline{C}H_2$), 128.62, 128.74 & 129.04 (aromatic), 136.59 (quat. aromatic), 156.71 (urethane CO) and 178.71 (CO_2H).

Methyl-(2*S*)-alaninate hydrochloride (2)



To a stirred suspension of (2*S*)-alanine (6.02 g, 67.6 mmol) in dry methanol (50 cm^3) at 0 °C was added thionyl chloride (13 cm^3, 178.2 mmol) dropwise with vigorous stirring. The reaction mixture was then refluxed for 2 h. The solution was allowed to cool and the methanol removed under reduced pressure to give a white solid which was recrystallised from dichloromethane/ diethyl ether to give the ester (**2**) as white crystals (5.80 g, 62%), m.p. 110-1 °C (lit.[98], 109-110 °C); υ_{max} (Nujol)/cm^{-1} 3430 (NH) and 1740 (ester CO); δ_H (200 MHz; ^2H_2O) 1.53 (3H, d, CH_3), 3.74 (3H, s, OCH_3) and 4.10 (1H, q, CH); δ_C (50.3 MHz; ^2H_2O) 17.95 (CH_3), 51.84 (CH), 56.43 ($CO_2\underline{C}H_3$) and 174.13 ($\underline{C}O_2CH_3$).

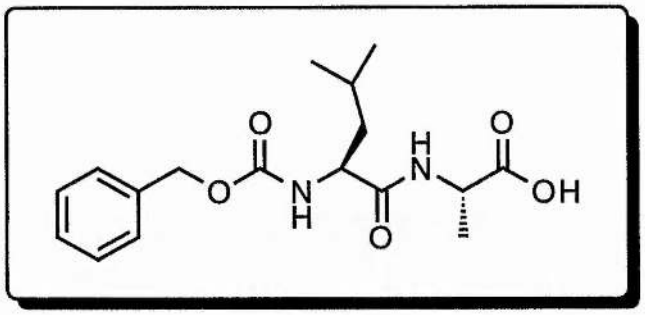
Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alaninate (3)



Carboxybenzyl-(2*S*)-leucine (**1**) (1.15 g, 4.35 mmol) was dissolved in dry THF (50 cm^3) with stirring at -15 °C. N-Methyl Morpholine (NMM) (0.48 cm^3, 4.35 mmol) was added and the mixture was then allowed to stir for 10

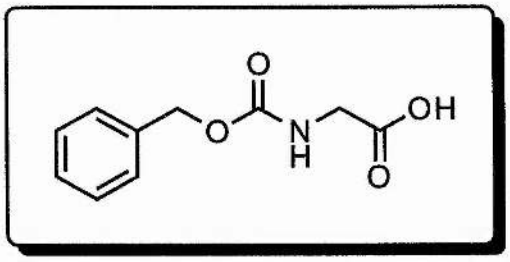
min. Isobutyl chloroformate (0.57 cm^3, 4.35 mmol) was added and a white precipitate of NMM hydrochloride was formed. Meanwhile, methyl-(2*S*)-alaninate hydrochloride **(2)** (0.613 g, 4.39 mmol) was dissolved in dry DMF (10 cm^3) and NMM (0.48 cm^3, 4.35 mmol) was added. This mixture was added to the THF mixture shortly after the addition of isobutyl chloroformate at -15 °C. The THF/DMF reaction mixture was maintained at -15 °C for 15 min and allowed to warm to room temperature. After 1 h the NMM hydrochloride was filtered and the solvents removed under reduced pressure to leave a thick yellow oil. The oil was dissolved in dichloromethane (50 cm^3) and washed successively with 10% citric acid (50 cm^3), 5% $NaHCO_3$ (50 cm^3) and saturated brine (50 cm^3), dried ($MgSO_4$) and the solvent removed under reduced pressure to give an off white solid which was recrystallised from ethyl acetate/ light petrol to give **(3)** as a white solid (1.03 g, 77%), m.p. 91-2 °C (lit.[99], 92-3 °C); υ_{max} (Nujol)/cm^{-1} 3300 (NH), 1740 (ester CO), 1690 (urethane CO) and 1650 (amide CO); δ_H (200 MHz; C^2HCl_3) 0.94 (6H, d, $CH(C\underline{H}_3)_2$), 1.39 (3H, d, $CHC\underline{H}_3$), 1.47-1.77 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.75 (3H, s, OCH_3), 4.20 (1H, m, $NHC\underline{H}CH_2$-), 4.56 (1H, m, $NHC\underline{H}CH_3$), 5.11 (2H, s, $C_6H_5C\underline{H}_2$), 5.26 (1H, d, $N\underline{H}CHCH_2$-), 6.58 (1H, d, $N\underline{H}CHCH_3$) and 7.35 (5H, m, C_6H_5-); δ_C (50.3 MHz; C^2HCl_3) 18.56 ($(NHCH\underline{C}H_3$), 22.47 & 23.41 ($CH(\underline{C}H_3)_2$), 25.07 ($\underline{C}H(CH_3)_2$), 42.10 ($\underline{C}H_2CH(CH_3)_2$), 48.48 ($\underline{C}HCH_3$), 52.94 ($\underline{C}HCH_2CH(CH_3)_2$), 53.84 ($OCH_3$), 67.49 ($C_6H_5\underline{C}H_2$), 128.51, 128.67 & 129.01 (aromatic), 136.71 (quat. aromatic), 156.77 (urethane CO), 172.54 (NHCO) and 173.71 ($\underline{C}OCH_3$).

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanine (4)



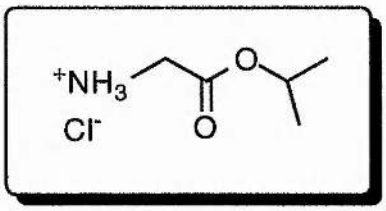
To a stirred solution of methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alaninate (**3**) (0.53 g, 1.50 mmol) in methanol (20 cm^3) was added 1 mol dm^{-3} aqueous NaOH (1.9 cm^3, 1.8 mmol). The reaction mixture was allowed to stir at room temperature, and the reaction followed to completion by tlc. Methanol was removed under reduced pressure after 2 h, and the remaining aqueous layer neutralised and acidified to pH 2 with 1 mol dm^{-3} HCl. The resultant white precipitate was filtered, dried ($MgSO_4$), recrystallised from dichloromethane/ light petrol and dried to give pure (**4**) as a white solid (0.41 g, 80%), m.p. 149-150 °C (lit.[100], 150-1 °C); υ_{max} (Nujol)/cm^{-1} 3290 (NH), 3280 (NH), 1720 (acid CO), 1710 (urethane CO) and 1700 (amide CO); δ_H (200 MHz; C^2HCl_3) 0.92 (6H, d, $CH(C\underline{H}_3)_2$), 1.40 (3H, d, $CHC\underline{H}_3$), 1.55-1.67 (3H, m, $CHC\underline{H}_2C\underline{H}(CH_3)_2$), 4.31 (1H, q, $NHC\underline{H}CH_2$), 4.54 (1H, m, $NHC\underline{H}CH_3$), 5.10 (2H, s, $C_6H_5C\underline{H}_2$), 5.70 (1H, d, urethane NH), 7.06 (1H, d, amide NH) and 7.33 (5H, s, aromatic CH); δ_C (50.3 MHz; C^2HCl_3) 18.36 ($CH\underline{C}H_3$), 22.47 & 23.39 ($CH(\underline{C}H_3)_2$), 25.06 ($\underline{C}H(CH_3)_2$), 41.80 ($\underline{C}H_2CH(CH_3)_2$), 48.68 ($\underline{C}HCH_3$), 53.85 ($\underline{C}HCH_2CH(CH_3)_2$), 67.72 ($C_6H_5\underline{C}H_2$), 128.51, 128.75 & 129.04 (aromatic CH), 136.51 (quat. aromatic) 157.01 (CO_2NH), 173.14 (CONH) and 176.15 (CO_2H).

Carboxybenzyl glycine (5)



This compound was prepared in a manner identical to that for (**1**) using glycine (3.02 g, 40 mmol) to give a white solid which was recrystallised from ethyl acetate/ light petrol to give (**5**) as white crystals (4.67 g, 55%), m.p. 118-119 °C (lit.[101], 120 °C); υ_{max} (Nujol)/cm^{-1} 3330 (NH), 3030 (OH), 2580 (aromatic CH), 1720 (urethane) and 1685 (acid); δ_H (200 MHz; d_6-DMSO) 3.69 (2H, d, NHC$\underline{H}_2$), 5.06 (2H, s, $C_6H_5C\underline{H}_2$), 7.38 (5H, m, aromatic CH) and 7.61 (1H, t, NH); δ_C (50.3 MHz; ^2H_2O) 42.40 ($NHCH_2$), 65.73 ($C_6H_5\underline{C}H_2$), 128.01, 128.10 & 128.64 (aromatic CH), 137.28 (quat. aromatic), 156.78 (CONH) and 171.88 (CO_2H).

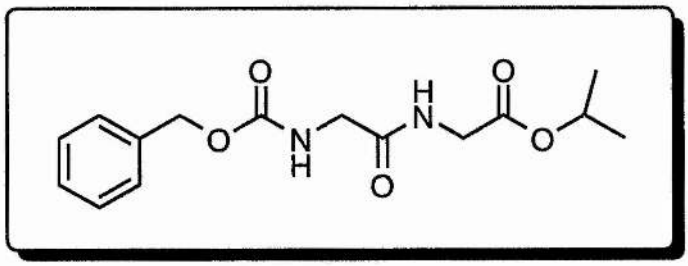
Isopropyl glycinate hydrochloride (6)



To a stirred suspension of glycine (10.03 g, 0.134 mol) in dry isopropanol (150 cm^3) at 0 °C was added thionyl chloride (26 cm^3, 0.256 mol) dropwise with vigorous stirring. The reaction mixture and was then refluxed for 2 h, the solution allowed to cool and the isopropanol removed under reduced pressure to give a white solid which was recrystallised from isopropanol/

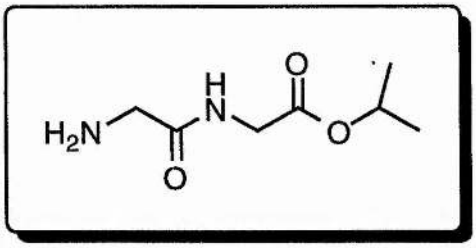
diethyl ether to give **(6)** as white crystals (17.63 g, 86%), m.p. 85 °C (lit.[102], 84-6 °C); δ_H (200 MHz; d_6-DMSO) 1.25 (6H, d, $CH(C\underline{H}_3)_2$), 3.71 (2H, s, CH_2CO), 5.0 (1H, sep, $C\underline{H}(CH_3)_2$) and 8.65 (3H, s, NH_3Cl); δ_C (50.3 MHz; d_6-DMSO) 21.74 ($CH(\underline{C}H_3)_2$), 39.85 ($\underline{C}H_2CO$), 69.65 ($\underline{C}H(CH_3)_2$) and 167.29 (CO).

Isopropyl carboxybenzyl glycyl glycinate (7)



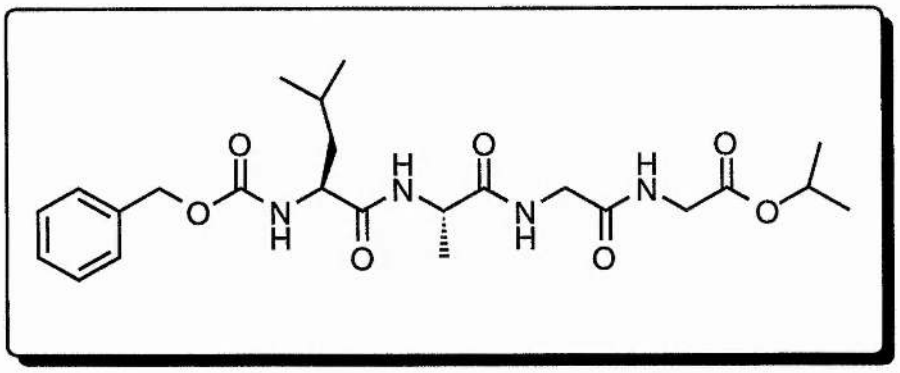
This compound was prepared in a manner identical to that for **(3)**, using **(5)** (2.00 g, 9.57 mmol) and **(6)** (1.47 g, 9.57 mmol). The crude dipeptide was purified by trituration with diethyl ether to give pure dipeptide **(7)** as a white solid (2.18 g, 74%), m.p. 92-4 °C; *m/z* (Found: $[M + H]^+$, 309.1450. $C_{15}H_{21}N_2O_5$ requires 309.1372); δ_H (200 MHz; C^2HCl_3) 1.25 (6H, d, $CH(C\underline{H}_3)_2$), 3.89 (2H, d, $NHC\underline{H}_2CO_2$), 3.95 (2H, d, $NHC\underline{H}_2CO$), 5.04 (1H, sep, $OC\underline{H}(CH_3)_2$), 5.13 (2H, s, $C_6H_5C\underline{H}_2$), 5.61 (1H, t, CONH), 6.68 (1H, t, $N\underline{H}CH_2$) and 7.36 (5H, s, C_6H_5); δ_C (50.3 MHz; C^2HCl_3) 22.10 ($(\underline{C}H_3)_2CH$), 41.90 ($NH\underline{C}H_2CO_2Pr^i$), 42.75 ($O_2CNH\underline{C}H_2CO$), 66.93 ($O\underline{C}H(CH_3)_2$), 70.05 ($C_6H_5\underline{C}H_2$), 128.66, 128.73 & 129.00 (aromatic), 136.60 (quat. aromatic), 157.15 (urethane CO), 170.00 ($CH_2\underline{C}ONH$) and 176.80 ($\underline{C}O_2CH(CH_3)_2$); *m/z* (EI) 309 ($[M + H]^+$, 3.5%), 265 (5), 201 (12.0), 159 (18.0), 141 (9.6), 108 (43.7), 91 (80.6), 79 (24.8), 65 (11.3), 56 (20.6) and 43 (100).

Isopropyl glycyl glycinate (8)



To a solution of isopropyl carboxybenzyl glycyl glycinate (**7**) (0.10 g, 0.33 mmol) in isopropanol (50 cm^3) was added 10% palladium on carbon (20 mg). The flask was sealed, flushed with argon then hydrogen and hydrogen was bubbled through the reaction mixture with vigorous stirring. The reaction was followed to completion by tlc and worked up after 2 h. The reaction mixture was filtered through a pre-washed celite pad and the filtrate was concentrated under reduced pressure to give pure (**8**) as a white solid (40 mg, 71%), m.p. 120 °C (decomp.); *m/z* (Found: $[M + H]^+$, 175.1083. $C_7H_{15}N_2O_3$ requires 175.1083); δ_H (200 MHz; ^2H_2O) 1.21 (6H, d, $CH(C\underline{H}_3)_2$), 3.32 (2H, s, $NH_2\underline{C}H_2CO$), 3.95 (2H, s, $NHC\underline{H}_2CO$) and 5.00 (1H, sep, $C\underline{H}(CH_3)_2$); δ_C (50.3 MHz; ^2H_2O) 23.66 ($CH(\underline{C}H_3)_2$), 44.36 ($NH_2\underline{C}H_2CO$), 46.48 ($NH\underline{C}H_2CO_2$), 73.67 ($\underline{C}H(CH_3)_2$), 174.15 (amide CO) and 179.07 (ester CO); *m/z* (EI) 174 (M^+, 15.2%), 159 (4.2), 145 (12.7), 132 (14.1), 115 (21.0), 105 (8.5), 87 (58.0), 75 (26.5), 58 (28.2) and 43 (100). Some ex-dimer formation *m/z* 349 in CI = $[2M + H]^+$.

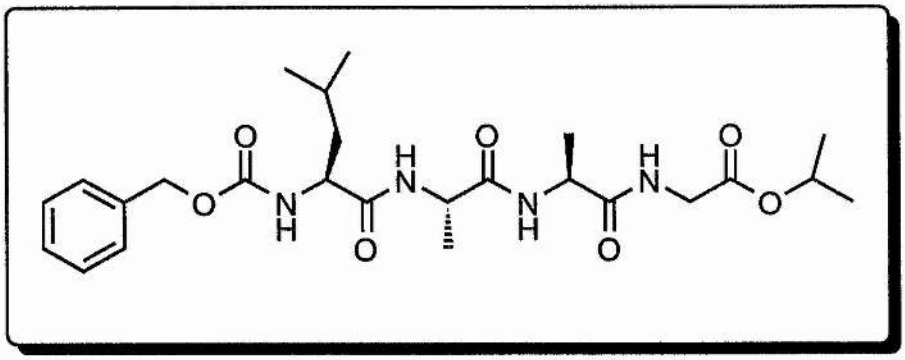
Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycyl glycinate (9)



This compound was prepared in a manner identical to that for the dipeptide **(3)**, using **(4)** (1.06 g, 3.2 mmol) and **(8)** (0.55 g, 3.2 mmol) (it was found that it was unnecessary to add an equivalent of NMM to the DMF solution of **(8)** since this is a free amine and not a hydrochloride salt). The dissolution of the amine **(8)** required gentle heat. The crude tetrapeptide was purified by crystallisation from methanol/ water to afford **(9)** as a white solid (1.02 g, 63%), m.p. 125-6 °C; (Found: C, 58.75; H, 7.6; N, 11.45. Calc. for $C_{24}H_{36}N_4O_7$: C, 58.5; H, 7.4; N, 11.4%); *m/z* (Found; $[M + H]^+$, 493.2649. Calc. for $C_{24}H_{37}N_4O_7$: 493.2584); $[\alpha]_D^{22}$ -26.8 (*c* 0.5 in CH_3OH); δ_H (200 MHz; C^2HCl_3) 0.92 (6H, d, $CH_2CH(C\underline{H}_3)_2$), 1.23 (6H, d, $OCH(C\underline{H}_3)_2$), 1.38 (3H, d, $CHC\underline{H}_3$), 1.49-1.75 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.97 (4H, d, $NHC\underline{H}_2CO$ and $NHC\underline{H}_2CO_2$), 4.16 (1H, q, $C\underline{H}CH_3$), 4.39 (1H, t, $C\underline{H}CH_2$), 5.01 (1H, sep, $C\underline{H}(CH_3)_2$), 5.09 (2H, s, $C_6H_5C\underline{H}_2$), 5.28, 6.63, 6.85 & 7.08 (all 1H, d, NH's) and 7.34 (5H, s, aromatic); δ_C (50.3 MHz; C^2HCl_3) 18.68 ($CH\underline{C}H_3$), 21.56 ($OCH(\underline{C}H_3)_2$), 21.92 ($CH_2\underline{C}H(CH_3)_2$), 22.82 & 24.56 ($CH_2CH(\underline{C}H_3)_2$), 41.43 ($\underline{C}H_2CH(CH_3)_2$), 41.99 ($NH\underline{C}H_2CO$), 42.82 ($NH\underline{C}H_2CO_2$), 48.85 ($NH\underline{C}H(CH_2)CO$), 53.45 ($NH\underline{C}H(CH_3)CO$), 66.82 ($C_6H_5\underline{C}H_2$), 69.16 ($O\underline{C}H(CH_3)_2$), 127.76, 127.98 & 128.35 (aromatic CH's), 136.21 (quat. aromatic), 156.30 (urethane CO), 168.98, 169.13 & 172.52 (amide CO's) and

173.60 (ester CO); *m/z* (FAB) 515 ([M + Na]$^+$, 100%), 493 ([M + H]$^+$, 35), 319 (5.4), 265 (5.2) and 147 (52.2).

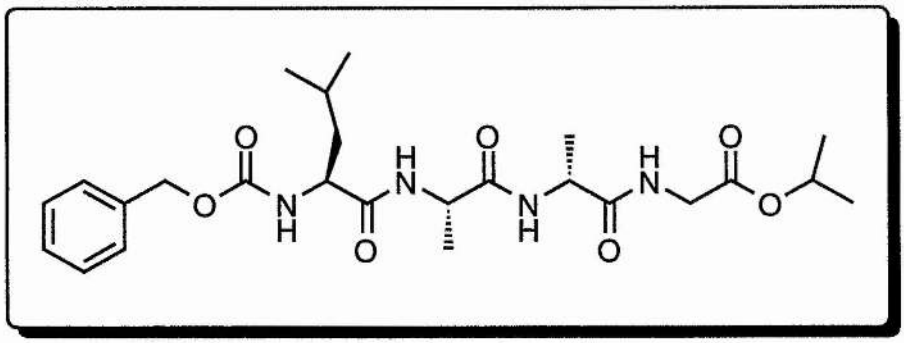
Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanyl glycinate (10)



This compound was prepared in a manner identical to that for the glycyl analogue **(9)**, using the (2*S*)-alanine dipeptide and follow **(1)**, **(7)** and **(8)**) (210 mg, 1.10 mmol) to give an off-white solid which was recrystallised from ethyl acetate/ light petrol to give the tetrapeptide **(10)** as a white solid (410 mg, 74%), m.p. 200-2 °C; (Found: C, 59.1; H, 7.55; N, 10.85. Calc. for $C_{25}H_{38}N_4O_7$: C, 59.25; H, 7.55; N, 11.05%); $[\alpha]_D^{22}$ -54.2 (*c* 0.5 in CH_3OH); δ_H (200 MHz; C^2HCl_3) 0.94 & 0.96 (6H, d, $CH_2CH(C\underline{H}_3)_2$), 1.22 (6H, d, $OCH(C\underline{H}_3)_2$), 1.38 (6H, d, ala CH_3's), 1.68 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.95 (2H, m, $NHC\underline{H}_2CO_2$), 4.30 (1H, q, $NHC\underline{H}(CH_2)CO$), 4.5 & 4.65 (1H & 1H, t & t, ala $C\underline{H}CH_3$'s), 5.05 (1H, sep, $OC\underline{H}(CH_3)_2$), 5.10 (2H, s, $C_6H_5C\underline{H}_2$), 5.5 (1H, d, urethane NH), 6.95-7.15 (3H, m, amide NH's) and 7.35 (5H, s, aromatic CH's); δ_C (50.3 MHz; C^2HCl_3) 19.20 (both ala CH_3's), 22.23 ($OCH(\underline{C}H_3)_2$), 22.51 ($CH_2\underline{C}H(CH_3)_2$), 23.45 & 25.26 ($CH_2CH(\underline{C}H_3)_2$), 42.05 ($\underline{C}H_2CH(CH_3)_2$), 42.47 ($NH\underline{C}H_2CO$), 49.20 ($NH\underline{C}H(CH_2)CO$), 49.58 ($NH\underline{C}H(CH_3)CO$), 54.26 ($NH\underline{C}H(CH_3)CO$), 67.54 ($C_6H_5\underline{C}H_2$), 69.72 ($O\underline{C}H(CH_3)_2$), 128.41, 128.71 & 129.04 (aromatic CH's), 136.68 (quat.

aromatic), 156.93 (urethane CO) and 169.73, 172.47, 173.06 & 173.09 (amides and ester CO's).

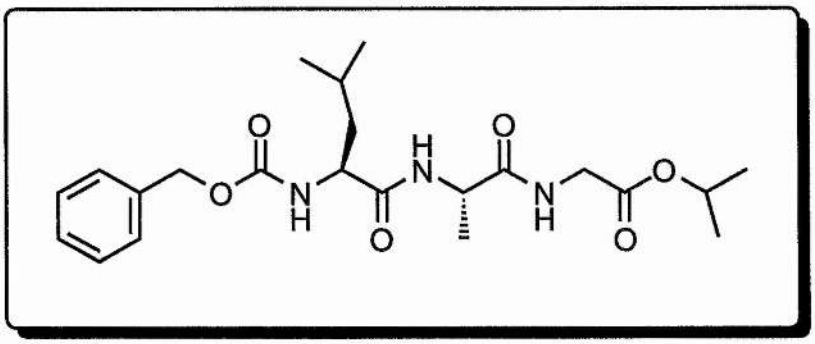
Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanyl glycinate (11)



This compound was prepared in a manner identical to that for the (2*S*) enantiomer **(10)**, using the (2*R*) alanine dipeptide (700 mg, 3.71 mmol) to give an off-white solid which was recrystallised from methanol/ water to give the tetrapeptide **(11)** as a white solid (870 mg, 46%), m.p. 185-187 °C; (Found: C, 59.25; H, 7.45; N, 10.70. Calc for $C_{25}H_{38}N_4O_7$: C, 59.25; H, 7.55; N, 11.05%); *m/z* (Found: $[M + H]^+$, 507.2827. Calc. for $C_{25}H_{39}N_4O_7$: 507.2819); $[\alpha]_D^{22}$ -12.7 (*c* 0.5 in CH_3OH); δ_H (200 MHz; C^2HCl_3) 0.89 & 0.90 (6H, m, $CH_2CH(C\underline{H}_3)_2$), 1.22 (6H, d, $OCH(C\underline{H}_3)_2$), 1.40 (6H, d, ala CH_3's), 1.59 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.95 (2H, m, $NHC\underline{H}_2CO$), 4.22 (1H, q, $NHC\underline{H}(CH_2)CO$),4.4-4.7 (2H, m, both ala CH's), 5.08 (1H, sep, $OC\underline{H}(CH_3)_2$), 5.11 (2H, s, $C_6H_5C\underline{H}_2$), 5.49 (1H, d, urethane NH), 7.05 (3H, m, amide NH's) and 7.35 (5H, s, aromatic CH's); δ_C (50.3 MHz; C^2HCl_3) 18.33 & 18.50 (both ala CH_3's), 22.23 ($OCH(\underline{C}H_3)_2$), 23.46 ($CH_2\underline{C}H(CH_3)_2$), 25.19 ($CH_2CH(\underline{C}H_3)_2$), 41.92 ($\underline{C}H_2CH(CH_3)_2$), 41.99 ($NH\underline{C}H_2CO$), 49.24 ($NH\underline{C}H(CH_2)CO$), 49.49 ($NH\underline{C}H(CH_3)CO$), 54.32 ($NH\underline{C}H(CH_3)CO$), 67.77 ($C_6H_5\underline{C}H_2$), 69.73 ($O\underline{C}H(CH_3)_2$), 128.62, 128.76 & 129.04 (aromatic CH's),

136.45 (quat. aromatic), 157.00 (urethane CO) and 169.98, 172.95, 173.3 & 173.5 (amides and ester CO's); *m/z* (CI) 507 ([M + H]$^+$, 34%), 489 (6.1), 399 (91.9), 373 (6.3), 351 (32.1), 319 (13.8), 307 (8.9), 282 (33.8), 254 (6.1), 189 (14.6), 176 (6.0), 147 (34.0), 108 (23.1) and 91 (100).

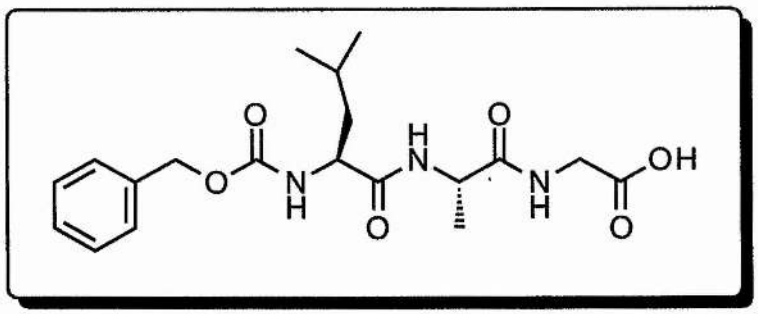
Isopropyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycinate (12)



This compound was prepared in a manner identical to that for the tetrapeptide **(9)**, using the monoprotected glycine derivative **(6)** (0.47 g, 2.99 mmol) to give the tripeptide as an off-white solid. Purification by silica chromatography using dichloromethane/ methanol (95:5) as the eluant gave the pure peptide **(12)** as a white solid (0.72 g, 55%), m.p. 137 °C; (Found; C, 60.8; H, 7.7; N, 9.6. Calc. for $C_{22}H_{33}N_3O_6$: C, 60.7; H, 7.65; N, 9.65%); *m/z* (Found: M$^+$, 435.2377. Calc. for $C_{22}H_{33}N_3O_6$: 435.2369); δ_H (200 MHz; C^2HCl_3) 0.92 (6H, d, $CH_2CH(C\underline{H}_3)_2$), 1.25 (6H, d, $OCH(C\underline{H}_3)_2$), 1.37 (3H, d, $NHCH(C\underline{H}_3)CO$), 1.60 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.96 (2H, d, $NHC\underline{H}_2CO_2$), 4.20 (1H, q, $NHC\underline{H}(CH_2)CO$), 4.58 (1H, quin, $NHC\underline{H}(CH_3)CO$), 5.09 (1H, sep, $OC\underline{H}(CH_3)_2$), 5.10 (2H, s, $C_6H_5C\underline{H}_2$), 5.36 (1H, d, urethane NH), 6.72 & 6.88 (2H, d & t, amide CO's) and 7.36 (5H, s, aromatic CH's). δ_C (50.3 MHz; C^2HCl_3) 18.71 ($CH\underline{C}H_3$), 22.22 ($OCH(\underline{C}H_3)_2$), 23.47 ($CH_2\underline{C}H(CH_3)_2$), 25.16 ($CH_2CH(\underline{C}H_3)_2$), 41.82 ($\underline{C}H_2CH(CH_3)_2$), 42.20 ($NH\underline{C}H_2CO$), 49.20 ($NH\underline{C}H(CH_2)CO$), 54.13 ($NH\underline{C}H(CH_3)CO$), 67.65 ($C_6H_5\underline{C}H_2$), 69.83

($O\underline{C}H(CH_3)_2$), 128.57, 128.72 & 129.03 (aromatic CH's), 136.58 (quat. aromatic), 148.74 (urethane CO), 169.62 (ester CO) and 172.64 & 172.72 (amide CO's); *m/z* (EI) 435 (M^+, 5.0%), 379 (15.5), 319 (10.5), 291 (23.0), 248 (33.2), 220 (36.9), 184 (8.7), 176 (58.3), 134 (17.0), 107 (8.9), 99 (5.1) and 91 (100).

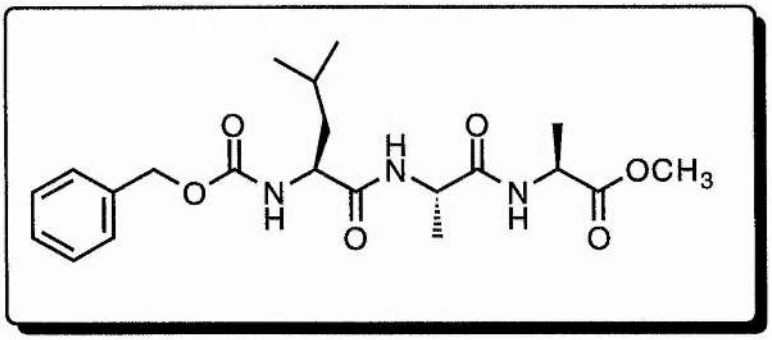
Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycine (13)



This compound was prepared in a manner identical to that for **(4)**, using the tripeptide **(12)** (110 mg, 253 mmol) to give a white precipitate which was dissolved in 1 mol dm^{-3} NaOH (5 cm^3, 5 mmol) and neutralised to pH 2 using 1 $moldm^{-3}$ HCl, the resultant white precipitate was filtered, washed with 1 mol dm^{-3} HCl and dried to give pure **(13)** as a white solid (83 mg, 83%), m.p. 149 °C; (Found: C, 58.2; H, 6.75; N, 10.65. Calc. for $C_{19}H_{27}N_3O_6$: C, 58.0; H, 6.9; N, 10.7%); *m/z* (Found: $[M + H]^+$, 394.1970. $C_{19}H_{28}N_3O_6$: 394.1978); δ_H (300 MHz; $C^2H_3O^2H$) 1.12 (6H, m, $CH_2CH(C\underline{H}_3)_2$), 1.57 (3H, d, $NHCH(C\underline{H}_3)CO$), 1.78 (2H, m, $C\underline{H}_2CH(CH_3)_2$), 1.90 (1H, m, $CH_2C\underline{H}(CH_3)_2$), 4.1 (2H, m, $NHC\underline{H}_2CO$), 4.33 (1H, t, $NHC\underline{H}(CH_2)CO$), 4.62 (1H, q, $NHC\underline{H}(CH_3)CO$), 5.29 (2H, s, $C_6H_5C\underline{H}_2$) and 7.58 (5H, m, aromatic CH's); δ_C (75.47 MHz; $C^2H_3O^2H$) 18.30 ($CHC\underline{H}_3$), 22.11 ($CH_2\underline{C}H(CH_3)_2$), 23.75 & 26.15 ($CH_2CH(\underline{C}H_3)_2$), 42.03 ($\underline{C}H_2CH(CH_3)_2$), 42.20 ($NH\underline{C}H_2CO$), 50.39 ($NH\underline{C}H(CH_2)CO$), 55.40 ($NH\underline{C}H(CH_3)CO$), 68.11 ($C_6H_5\underline{C}H_2$), 129.14, 129.31 & 129.76 (aromatic CH's), 138.43 (quat. aromatic), 159.04 (urethane CO) and

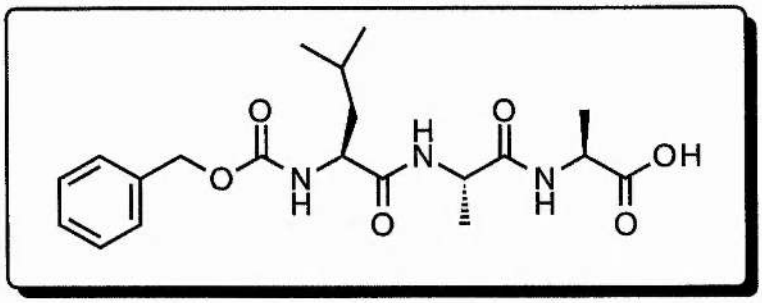
173.04, 175.34 & 175.55 (amide and acid CO's); *m/z* (CI) 394 ([M + H]$^+$, 100%), 376 (59.1), 337 (10.8), 286 (10.1) and 260 (6.1).

Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alaninate (14)



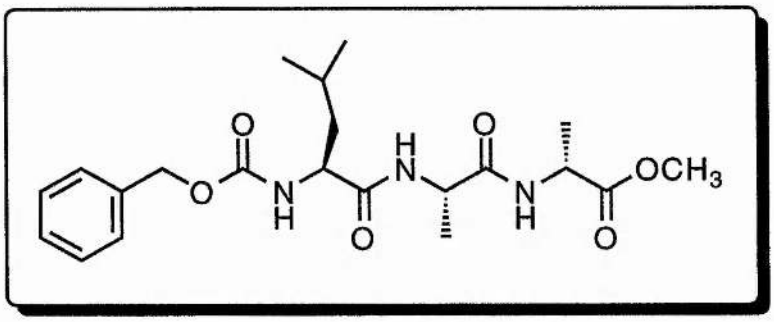
This compound was prepared in a manner identical to that for the glycyl analogue (**12**), using the monoprotected (2*S*)-alanine derivative (**2**) (660 mg, 4.73 mmol) to give the tripeptide as an off-white solid which was recrystallised from methanol/ water to give pure (**14**) as a white solid (1.36 g, 68%), m.p. 174-6 °C; (Found: C, 59.75; H, 7.5; N, 9.9. Calc for $C_{21}H_{31}N_3O_6$: C, 59.85; H, 7.4; N, 9.95%); *m/z* (Found: [M + H]$^+$, 422.2294. Calc. for $C_{21}H_{32}N_3O_6$: 422.2291); δ_H (300 MHz; C^2HCl_3) 0.91 (6H, d, $CH_2CH(C\underline{H}_3)_2$), 1.38 (6H, m, both ala CH_3's), 1.60 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.75 (3H, s, OCH_3), 4.21 (1H, m, $NHC\underline{H}(CH_2)CO$), 4.55 (2H, m, both ala CH's), 5.12 (2H, s, $C_6H_5C\underline{H}_2$), 5.40 (1H, d, urethane NH), 6.78 (1H, d, amide NH), 6.89 (1H, d, amide NH) and 7.35 (5H, s, aromatic CH's); δ_C (75.47 MHz; C^2HCl_3) 18.29 (both ala CH_3's), 22.07 ($CH_2\underline{C}H(CH_3)_2$), 23.82 & 24.86 ($CH_2CH(\underline{C}H_3)_2$), 41.79 ($\underline{C}H_2CH(CH_3)_2$), 48.28 ($NH\underline{C}H(CH_2)CO$), 48.94 ($NH\underline{C}H(CH_3)CO$), 52.63 ($NH\underline{C}H(CH_3)CO_2CH_3$), 53.76 ($CO_2\underline{C}H_3$), 67.30 ($C_6H_5\underline{C}H_2$), 128.22, 128.40 & 128.72 (aromatic CH's), 136.35 (quat. aromatic), 156.44 (urethane CO), 171.72 (ester CO) and 172.36 & 173.28 (amide CO's); *m/z* (CI) 422 ([M + H]$^+$, 100%) and 314 (6.0).

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-alanine (15)



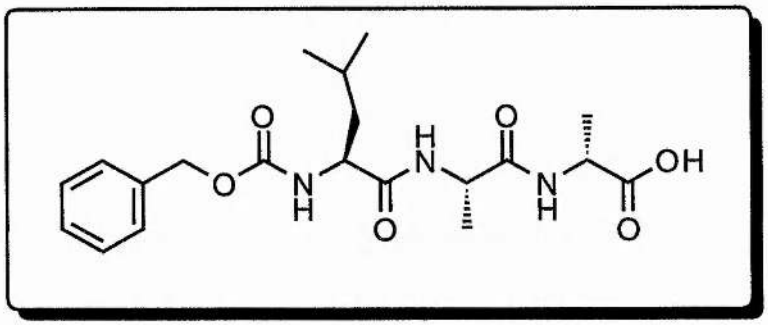
This compound was prepared in a manner identical to that for the glycine analogue **(13)**, using the tripeptide **(14)** (350 mg, 0.83 mmol) to give a white precipitate which was dissolved in 1 mol dm^{-3} NaOH (5 cm^3, 5 mmol) and neutralised to pH 2 using 1 mol dm^{-3} HCl, the resultant white precipitate was filtered, washed with 1 mol dm^{-3} HCl and dried to give pure **(15)** as a white solid (290 mg, 85%), m.p. 183-4 °C; (Found: C, 57.5; H, 6.9; N, 9.95. Calc. for $C_{20}H_{29}N_3O_6.0.5H_2O$: C, 57.7; H, 7.25; N,10.1%); *m/z* (Found: $[M + H]^+$, 408.2130. Calc. for $C_{20}H_{30}N_3O_6$: 408.2135); δ_H (300 MHz; $C^2H_3O^2H$) 1.14 (6H, m, $CH_2CH(C\underline{H}_3)_2$), 1.55 (6H, m, both ala CH_3's), 1.73 (2H, t, $NHCH(C\underline{H}_2CH(CH_3)_2)CO$), 1.89 (1H, m, $CH_2C\underline{H}(CH_3)_2$), 4.32 (1H, t, $NHC\underline{H}(CH_2CH(CH_3)_2)CO$), 4.53 (2H, m, both ala CH's), 5.28 (2H, s, $C_6H_5C\underline{H}_2$) and 7.52 (5H, m, aromatic CH's); δ_C (75.47 MHz; $C^2H_3O^2H$) 17.65 & 18.36 (both ala CH_3's), 22.14 ($CH_2\underline{C}H(CH_3)_2$), 23.72 & 26.12 ($CH_2CH(\underline{C}H_3)_2$), 42.29 ($\underline{C}H_2CH(CH_3)_2$), 49.70 ($NH\underline{C}H(CH_3)CO_2H$), 50.28 ($NH\underline{C}H(CH_2)CO$), 55.19 ($NH\underline{C}H(CH_3)CO$), 67.99 ($C_6H_5\underline{C}H_2$), 129.05, 129.28 & 129.75 (aromatic CH's), 138.44 (quat. aromatic), 158.83 (urethane CO) and 174.81 & 175.42 (amide CO's); *m/z* (CI) 408 ($[M + H]^+$, 100%), 390 (5.8), 336 (7.4), 300 (22.3), 274 (11.5), 224 (5.3), 185 (8.8), 143 (6.2) and 90 (11.7).

Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alaninate (16)



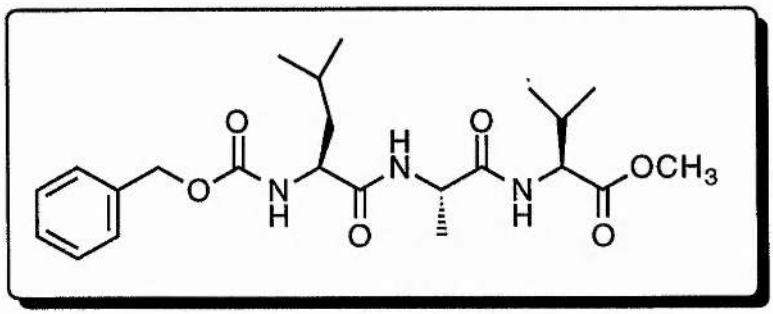
This compound was prepared in a manner identical to that for the (2*S*,2*S*,2*S*) epimer (**14**), using the monoprotected (2*R*)-alanine derivative (210 mg, 1.51 mmol) to give the tripeptide as an off-white solid which was recrystallised from methanol/ water to give pure (**16**) as a white solid (420 mg, 66%), m.p. 164 °C; *m/z* (Found: $[M + H]^+$, 422.2299. Calc. for $C_{21}H_{32}N_3O_6$: 422.2291); δ_H (200 MHz; C^2HCl_3) 0.92 (6H, d, $CH_2CH(C\underline{H}_3)_2$), 1.39 (6H, m, both ala CH_3's), 1.65 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.72 (3H, s, OCH_3), 4.18 (1H, m, $NHC\underline{H}(CH_2)CO$), 4.53 (2H, m, both ala CH's), 5.11 (2H, s, $C_6H_5C\underline{H}_2$), 5.23 (1H, d, urethane NH), 6.60 (1H, d, amide NH), 6.89 (1H, d, amide NH) and 7.36 (5H, s, aromatic CH's). δ_C (50.3 MHz; C^2HCl_3) 18.08 & 18.20 (both ala CH_3's), 21.98 ($CH_2\underline{C}H(CH_3)_2$), 23.13 & 24.95 ($CH_2CH(\underline{C}H_3)_2$), 41.49 ($\underline{C}H_2CH(CH_3)_2$), 48.28 ($NH\underline{C}H(CH_2)CO$), 52.60 ($NH\underline{C}H(CH_3)CO$), 54.13 ($NH\underline{C}H(CH_3)CO_2CH_3$), 67.53 ($C_6H_5\underline{C}H_2$), 128.38, 128.51 & 128.77 (aromatic CH's), 136.19 (quat. aromatic), 156.61 (urethane CO), 171.61 (ester CO) and 172.25 & 173.40 (amide CO's); *m/z* (CI) 422 ($[M + H]^+$, 100) and 314 (7.0).

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanine (17)



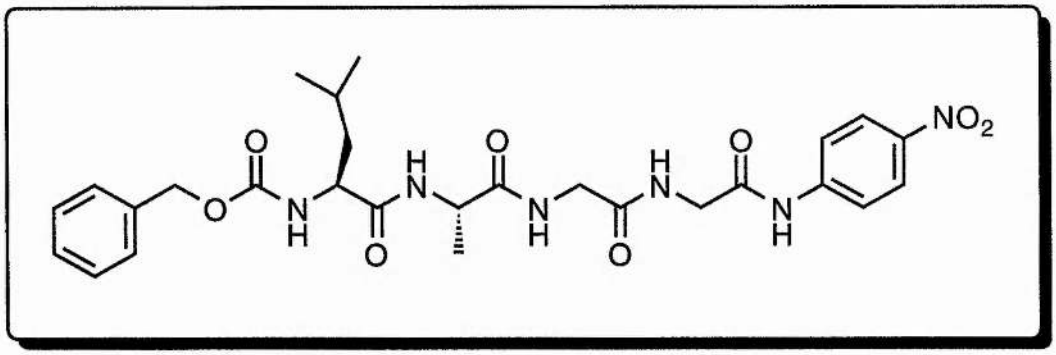
This compound was prepared in a manner identical to that for the (2*S*,2*S*,2*S*) epimer **(15)**, using the tripeptide **(16)** (440 mg, 1.05 mmol) to give a white precipitate which was dissolved in 1 mol dm^{-3} NaOH (5 cm^3, 5 mmol) and neutralised to pH 2 using 1 mol dm^{-3} HCl, the resultant white precipitate was filtered, washed with 1 mol dm^{-3} HCl and dried to give pure **(17)** as a white solid (230 mg, 54%), m.p. 170 °C; (Found: C, 58.9; H, 7.2; N, 10.2. Calc for $C_{20}H_{29}N_3O_6$: C, 58.95; H, 7.15; N, 10.3%); *m/z* (Found: $[M + H]^+$, 408.2136. Calc. for $C_{20}H_{30}N_3O_6$: 408.2135); δ_H (300 MHz; $C^2H_3O^2H$) 1.14 (6H, m, $CH_2CH(C\underline{H}_3)_2$), 1.57 & 1.62 (6H, d, both ala CH_3's), 1.76 (2H, t, $NHCH(C\underline{H}_2CH(CH_3)_2)CO$), 1.93 (1H, m, $CH_2C\underline{H}(CH_3)_2$), 4.31 (1H, t, $NHC\underline{H}(CH_2CH(CH_3)_2)CO$), 4.62 (2H, m, both ala CH's), 5.31 (2H, m, $C_6H_5C\underline{H}_2$) and 7.57 (5H, m, aromatic CH's); δ_C (75.47 MHz; $C^2H_3O^2H$) 18.04 & 18.23 (both ala CH_3's), 22.15 ($CH_2\underline{C}H(CH_3)_2$), 23.71 & 26.19 ($CH_2CH(\underline{C}H_3)_2$), 42.15 ($\underline{C}H_2CH(CH_3)_2$), 49.59 ($NH\underline{C}H(CH_3)CO_2H$), 50.38 ($NH\underline{C}H(CH_2)CO$), 55.72 ($NH\underline{C}H(CH_3)CO$), 68.22 ($C_6H_5\underline{C}H_2$), 129.24, 129.35 & 129.78 (aromatic CH's), 138.38 (quat. aromatic), 159.15 (urethane CO) and 174.60 & 175.64 (amide CO's); *m/z* (CI) 408 ($[M + H]^+$, 100%), 390 (6.0), 300 (35.3), 274 (15.1) and 147 (8.4).

Methyl carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*S*)-valinate (18)



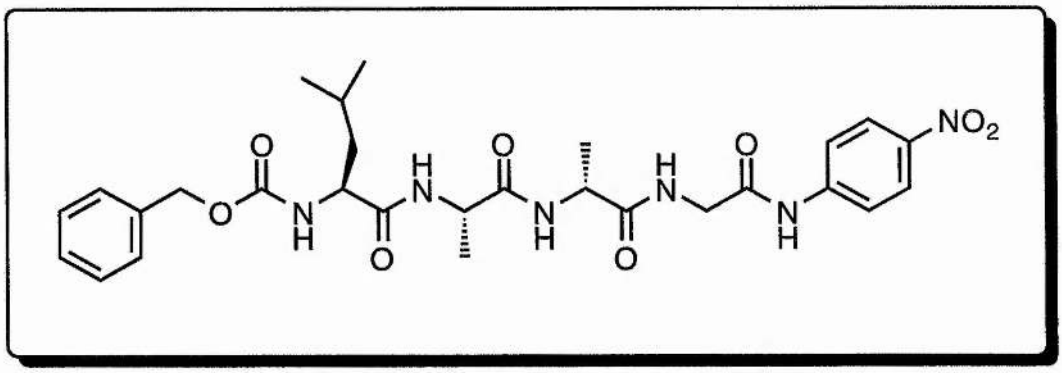
This compound was prepared in a manner identical to the for the glycine analogue **(12)**, using the monoprotected (2*S*)-valine derivative (250 mg, 1.50 mmol) to give the tripeptide as an off-white solid which was recrystallised from methanol/ water to give pure **(18)** as a white solid (500 mg, 67%), m.p. 173 °C; (Found: C, 61.3; H, 7.9; N, 9.2. Calc for $C_{23}H_{35}N_3O_6$: C, 61.45; H, 7.85; N, 9.35%); *m/z* (Found: M^+, 450.2612. Calc. for $C_{23}H_{35}N_3O_6$: 450.2604); δ_H (300 MHz; C^2HCl_3) 0.95 (12H, m, leu CH_3's & val CH_3's), 1.37 (3H, d, ala CH_3), 1.60 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 2.15 (1H, m, $CHC\underline{H}(CH_3)_2$), 3.73 (3H, s, OCH_3), 4.22 (1H, m, $NHC\underline{H}(CH_3)CO$), 4.51 (2H, m, leu & val CH's), 5.10 (2H, s, $C_6H_5C\underline{H}_2$), 5.35 (1H, d, urethane NH), 6.72 (1H, d, amide NH), 6.81 (1H, d, amide NH) and 7.35 (5H, s, aromatic CH's); δ_C (75.47 MHz; C^2HCl_3) 17.92 & 18.04 ($CHCH(\underline{C}H_3)_2$), 19.07 ($CH\underline{C}H_3$), 22.08 ($CH_2\underline{C}H(CH_3)_2$), 23.08 & 24.86 ($CH_2CH(\underline{C}H_3)_2$), 31.35 ($CH\underline{C}H(CH_3)_2$), 41.81 ($\underline{C}H_2CH(CH_3)_2$), 49.13 ($NH\underline{C}H(CH_2)CO$), 52.35 ($NH\underline{C}H(CH_3)CO$), 53.71 (NHCH(CH)CO), 57.47 (OCH_3), 67.33 ($C_6H_5\underline{C}H_2$), 128.24, 128.41 & 128.73 (aromatic CH's), 136.23 (quat. aromatic), 156.38 (urethane CO), 172.04 (ester CO) and 172.30 & 172.44 (amide CO's); *m/z* (CI) 450 (M^+, 100%), 395 (6.2), 357 (20.5), 342 (27.4) and 147 (11.1).

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl glycyl glycine-*p*-nitroanilide (19)



This compound was prepared in a manner identical to that for the tetrapeptide **(9)** using the mixed anhydride to couple the tripeptide **(13)** (65 mg, 0.165 mmol) to glycine-*p*-nitroaniline (32.2 mg, 0.165 mmol). The reaction was carried as described for the preparation of **(9)** with the exception that NMM (18 mm^3, 0.165 mmol) was added to the solution of glycine-*p*-nitroaniline. The reaction was followed to completion using tlc and compound **(19)** was obtained as a pale yellow solid (15 mg, 16%); δ_H (200 MHz; $C^2H_3O^2H$) 0.95 (6H, m, $CH_2CH(C\underline{H}_3)_2$), 1.45 (3H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), 3.88 (2H, d, $NHC\underline{H}_2CONHPhNO_2$), 3.96 (2H, s, $NHC\underline{H}_2CO$), 4.18 (2H, m, $NHC\underline{H}(CH_2)$ & $NHC\underline{H}(CH_3)$), 5.09 (2H, s, $C_6H_5C\underline{H}_2$), 7.35 (5H, m, $C_6\underline{H}_5CH_2$), 7.83 (2H, d, *p*-nitroanilide protons 2 & 6) and 8.22 (2H, d, *p*-nitroanilide protons 3 & 5).

Carboxybenzyl-(2*S*)-leucyl-(2*S*)-alanyl-(2*R*)-alanyl glycine-*p*-nitroanilide (20)



This compound was prepared in a manner identical to that for the glycyl analogue **(19)** using the (2*R*)-alanine tripeptide **(17)** (190 mg, 0.467 mmol) to give a pale yellow solid which was recrystallised from methanol/ water to give pure **(20)** as a pale yellow solid (115 mg, 42%), m.p. 188 °C (decomp); (Found: C, 57.0; H, 6.1; N, 14.05. Calc for $C_{28}H_{36}N_6O_8$: C, 57.55; H, 6.2; N, 14.35%); δ_H (300 MHz; $C^2H_3O^2H$) 0.70 (3H, d, $CH_2CH(C\underline{H}_3)_2$), 0.78 (3H, d, $CH_2CH(C\underline{H}_3)_2$), 1.43 (9H, m, $C\underline{H}_2C\underline{H}(CH_3)_2$), (2*S*)-$CHC\underline{H}_3$, (2*R*)-$CHC\underline{H}_3$), 4.09 (2H, dd, $NHC\underline{H}_2CONHPhNO_2$), 4.11 (1H, m, $NHC\underline{H}(CH_2)CO$), 4.28 (2H, m, both ala CH's), 5.06 (2H, s, $C_6H_5C\underline{H}_2$), 7.32 (5H, s, $C_6\underline{H}_5CH_2$), 8.03 (2H, d, *p*-nitroanilide protons 2 & 6) and 8.17 (2H, d, *p*-nitroanilide protons 3 & 5); δ_C (74.47 MHz; $C^2H_3O^2H$) 17.11 (both ala CH_3's), 21.88 ($CH_2\underline{C}H(CH_3)_2$), 23.65 & 25.96 ($CH_2CH(\underline{C}H_3)_2$), 42.17 ($NH\underline{C}H_2CO$), 51.36 ($NH\underline{C}H(CH_2)CO$), 51.90 ($NH\underline{C}H(CH_3)CO$), 54.97 ($NH\underline{C}H(CH_3)CO$), 68.04 ($C_6H_5\underline{C}H_2$), 120.95 (*p*-nitroanilide carbons 2 & 6), 126.00 (*p*-nitroanilide carbons 3 & 5), 129.11, 129.32 & 129.76 (Cbz CH's), 138.43 (Cbz quat. aromatic), 145.09 (*p*-nitroanilide quat. carbon 1), 146.01 (*p*-nitroanilide quat. carbon 4), 159.03 (urethane CO), 170.64 ($NHCH_2\underline{C}ONHPhNO_2$) and 176.07, 176.36 & 176.52 (amide CO's).

Peptide assay for recombinant adenovirus 2 L3 23 kDa proteinase

Substrate: 2 mg/ml LSGAGFSW in water

Assay buffer: 50 mM Tris, 10 mM EDTA, pH 8, add 7.8 µl mercaptoethanol per 50 ml immediately before use.*

Activating peptide: 10 µg/ml pVI-CT (GVQSLKRRRCF) in water

Reaction mixture:

10 µl Proteinase

10 µl activating peptide

25 µl assay buffer

Leave 5 min at room temperature

10 µl substrate

Incubate at 37 °C for required time period.

For CE analysis the reaction was stopped by adding 10 µl of CE buffer (pH 2.5 phosphate) or 1% TFA to 10 µl of sample. 80 µl of water was added and the mixture was filtered before loading onto CE.

For HPLC analysis 100 µl 0.1% TFA was added to stop the reaction. The mixture was filtered and 25 µl was applied to a C18 reverse phase column.

Buffer A: 0.1% TFA in water

Buffer B: 0.1% TFA in acetonitrile

Program: LSGAG

Wavelength: 226 nm FSD 0.2

*Not all proteinase preparations require mercaptoethanol - but the activity of most is increased by its incorporation in the assay.

The proteinase is not particularly stable and repeated freezing/ thawing results in a rapid deterioration of enzyme activity. The enzyme is best kept at -70 °C, and then thawed and kept on ice as required.

CHAPTER FOUR

REFERENCES

4.0 REFERENCES

1. *Virology (Second Edition)*, 1990, Chapter 60, 1679-1721.

2. Rowe W.P., Huebner R.J., Gilmore L.K., Parrott R.H., T.G., *Proc. Soc. Exp. Biol. Med.*, 1953, **84**, 570-573.

3. Hilleman M.R., Werner J.H., *Proc. Soc. Exp. Biol. Med.*, 1954, **85**, 183-188.

4. Huebner R.J., Chanock R.M., Schell K., Casey M.J., *Proc. Natt. Acad. Sci. USA.*, 1965, **54**, 381-388.

5. Enders J.F., Bell J.A., Dingle J.H., *et al.*, *Science*, 1956, **124**, 119-120.

6. Dingle J., Langmuir A.D., *Am. Rev. Respir. Dis.*, 1968, **97**, 1-65.

7. Jawetz E., *Br. Med. J.*, 1959, **1**, 873-878.

8. Norrby E., Bartha A., Boulanger P. *et al.*, *Inter-Virology*, 1976, **7**, 117-125.

9. Nicholson F., *Eye*, 1993, Suppl., 1-4.

10. Horne R.W., Bonner S., Waterson A.P., Wiley P., *J. Mol. Biol.*, 1959, **1**, 84-86.

11. Stewart P.L. Burnett R.M., *Jpn. J. Appl. Phys.*, 1993, **32**, 1342-1347.

12. Ginsberg H.S., Periera H.G., Valentine R.C., Wilcox W.C. *Virology*, 1966, **28**, 782-783.

13. Horwitz M.S., Maizel J.V., Scharff M.D., *J. Virol.*, 1970, **6**, 569-571.

14. van Oostrum J., Burnett R.M., *J. Virol.*, 1985, **56**, 439-448.

15. Laver W.G., Suriano J.R., Green M., *J. Virol.*, 1967, **1**, 723-728.

16. Russell W.C., Laver W.G., Sanderson P.J., *Nature*, 1968, **219**, 1127-1130.

17. Everitt E., Sundquist B., Petterson U., Philipson L., *Virology*, 1973, **62**, 130-147.

18. Robinson A.J., Younghusband H.B., Bellet A.J.D., *Virology*, 1973, **56**, 54-59.

19. Mirza M.A., Weber J., *Biochem. Biophys. Acta*, 1982, **696**, 76-86.

20. Green M., Pina M., Kimes R., Wensink P.C., MacHattie L.A., Thomas C.A., *Proc. Natl. Acad. Sci. USA*, 1967, **57**, 1302-1309.

21. Rekosh D.M.K., Russell W.C., Bellet A.J.D., Robinson A.J., *Cell*, 1977, **11**, 283-295.

22. Steenburg P.H., Maat J., van Ormondt H., Sussenbach J.S., *Nucleic Acid Res.*, 1977, **4**, 4371-4389.

23. Corden J., Engelking H.M., Pearson G.D., *Proc. Natl. Acad. Sci. USA*, 1976, **73**, 401-404.

24. Hierholzer J.C., *J. Infect. Dis.*, 1973, **128**, 541-550.

25. Trentin J.J., Yabe Y., Taylor G., *Science*, 1962, **137**, 835-849.

26. Pina M., Green M., *Proc. Natl. Acad. Sci. USA*, 1965, **54**, 547-551.

27. Garon C.F., Berry K.W., Hierholzer J.C., Rose J.A., *Virology*, 1973, **54**, 414-426.

28. Green M., *Annu. Rev. Biochem.*, 1970, **39**, 701-756.

29. Wadell G., Hammarskjold M.L., Winberg G., Varsanyi T.M., Sundell G., *Ann. NY Acad. Sci.*, 1980, **354**, 16-42.

30. Green M., Pina M., *Virology*, 1963, **20**, 199-207.

31. Chow L.T., Broker T.R., Lewis J.B., *J. Mol. Biol.*, 1979, **134**, 265-303.

32. Londberg-Holm K., Philipson L., *J. Virol.*, 1969, **4**, 323-328.

33. Hennache B., Boulanger P., *Biochem. J.*, 1977, **166**, 237-247.

34. Philipson L., Londberg-Holm K., Petterson U., *J. Virol.*, 1968, **2**, 1064-1075.

35. Smart J., Stillman B.W., *J. Biol. Chem.*, 1982, **257**, 13499-13506.

36. Horwitz M.S., *Virology*, 1971, **8**, 675-683.

37. Pina M., Green M., *Virology*, 1969, **38**, 573-586.

38. Horwitz M.S., Scharff M.D., Maizel J.V., *Virology*, 1969, **39**, 682-684.

39. Cepko C.L., Sharp P.A., *Virology*, 1983, **129**, 137-154.

40. D'Halluin J.C., Milleville M., Boulanger P.A., Martin G.R., *J.Virol.*, 1978, **26**, 344-356.

41. Ghosh-Choudhury G., Haj-Ahmed Y., Graham F.L., *Embo. J.*, 1987, **6**, 1733-1739.

42. Houde A., Weber J.M., *Gene*, 1987, **54**, 51-56.

43. Zhang Y., Schneider R.J., *J. Virol.*, 1994, **68**, 2544-2555.

44. Chen P.H., Ornelles D.A., Schenk T., *J. Virol.*, 1993, **67**, 3507-3514.

45. *Virology (Second Edition)*, 1990, Chapter 61, 1723-1740.

46. Brandt C.D., Kim H.W., Vargosdo A.J. *et al.*, *Am. J. Epidemiol.*, 1969, **90**, 484-500.

47. Bell J.A., Rowe W.P., Engler J.I., Parrott R.H., Huebner R.J., *JAMA*, 1955, **175**, 1083-1092.

48. Huebner R.J., Rowe W.P., Chanock R.M., *Annu. Rev. Microbiol.*, 1958, **12**, 49-76.

49. Sturdy P.M., Court S.D.M., Gardner P.S., *Lancet.*, 1971, **2**, 978-979.

50. Zahradnik J.M., Spencer M.J., Parker D.D., *Am. J. Med.*, 1980, **68**, 725-732.

51. Bennett F.M., Law B.B., Hamilton W., MacDonald A., *Lancet.*, 1957, **2**, 670-673.

52. Guyer B., O'Day D.M., Hierholzer J.C., Schaffer W., *J. Infect. Dis.*, 1975, **132**, 142-150.

53. Kemp M.C., Hierholzer J.C., Cabradilla C.P., Obijesti J.F., *J. Infect. Dis.*, 1983, **148**, 29-33.

54. Tal F.H., Grayston J.T., *Proc. Soc. Exp. Biol. Med.*, 1962, **109**, 881-884.

55. Hierholzer J.C., Atuk N., Gwaltney J., *J. Clin. Microbiol.*, 1975, **1**, 366-376.

56. Chou S.M., Roo R., Burrell R., Gutmann L., *J. Neuropathol. Exp. Neurol.*, 1973, **32**, 34-50.

57. Wood D.J., *Br. Med. J.*, 1988, **296**, 229-230.

58. Flewett T.H., Bryden A.S., Davies H., Morris C.A., *Lancet*, 1975, **1**, 4-5.

59. Shinozaki T., Araki K., Ushijima H., Fujii R., *J. Clin. Microbiol.*, 1987, **25**, 1679-1682.

60. Bell T.M., Steyn J.H., *Br. Med. J.*, 1962, **2**, 700-702.

61. Yunis E.J., Hashida Y., *Pediatrics*, 1973, **51**, 566-570.

62. Kagnoff M.F., Paterson V.J., Kumar P.J. *et al.*, *Gut*, 1987, **28**, 995-1001.

63. Horwitz M.S., Valderrama G., Hatcher V., Kom R., deJong P., Spigland I., *Annu. NY Acad. Sci.*, 1985, **437**, 161-174.

64. Hierholzer J.C., Wigand R., Anderson L.J., Adrain T., Gold J.W.M., *J. Infect. Dis.*, 1988, **158**, 804-813.

65. Abken H., Butzler C., Willecke K., *Anticancer Res.*, 1987, **7**, 553-558.

66. Neumann R., Genersch E., Eggers H.J., *Virus Res.*, 1987, **7**, 93-97.

67. Wold W.S.M., Mackey J.K., Rigden P., Green M., *Cancer Res.*, 1979, **39**, 3479-3484.

68. Tremblay M.L., Dery C.V., Talbot B.G., Weber J., *Biochimica et Biophysica Acta*, 1983, **743**, 239-245.

69. Webster A., Russell W.C., Kemp G.D., *J. Gen. Virol.*, 1989, **70**, 3215-3223.

70. Webster A., Russell S., Talbot P., Russell W.C., Kemp G.D., *J. Gen. Virol.*, 1989, **70**, 3225-3234.

71. Webster A., Hay R.T., Kemp G.D., *Cell*, 1993, **72**, 97-104.

72. Mangel W.F., McGrath W.J., Toledo D.L., Anderson C.W., *Nature*, 1993, **361**, 274-275.

73. Webster A., Kemp G., *J. Gen. Virol.*, 1993, **74**, 1415-1420.

74. Rancourt C., Tihanyi K., Bourbonniere M., Weber J.M., *Proc. Natl. Acad. Sci.*, 1994, **91**, 844-847.

75. Grierson A.W., Nicholson R., Talbot P., Webster A., Kemp G.D., *J. Gen. Virol.*, 1994, **75**, 2761-2764.

76. Cornish J.A., Murray H., Kemp G.D., Gani D., *Bioorg. Med. Chem. Lett.*, 1995, **5**, 25-30.

77. Anderson C.W., *Protein Exp. Purif.*, 1993, **4**, 8-15.

78. Louis N., Fender P., Barge A., Kitts P., Chroboczek J., *J. Virol.*, 1994, **68**, 4104-4106.

79. Hellen C.U.T., Wimmer E., *Experimentia*, 1992, **48**, 201-215.

80. Kay J., Dunn M., *Biochem. et Biophys. Acta*, 1990, **1048**, 1-18.

81. Lokshina L.A., Dilakyan E.A., *Molekulyarnaya Biologiya*, 1985, **20**, 1157-1175.

82. Kamphuis I.G., Drenth J., Baker E.N., *J. Mol. Biol.*, 1985, **182**, 317-329.

83. Brocklehurst K., Willenbrock F., Salih E., *Hydrolytic Enzymes*, 1987, Chapter 2, 39-158.

84. Osterson M., Morland B., Husby G., *Clin. Exp. Immunol.*, 1982, **51**, 103-109.

85. Recklies A.P., Tiltman K.J., Stoker T.A.M., Poole A.R., *Cancer Res.*, 1980, **40**, 550-556.

86. Pietras R.J., Roberts J.A., *J. Biol. Chem.*, 1981, **256**, 1151-1153.

87. De Martino G.N., Blumenthal D.K., *Biochemistry*, 1982, **21**, 4297-4303.

88. Ishiura S., *Life Sci.*, 1981, **29**, 1079-1087.

89. Puca G.A., Nola E., Sica V., Breciani F., *J. Biol. Chem.*, 1977, **252**, 1358-1366.

90. Vedeekis W., Freeman M., Schrader W., O'Malley B., *Biochemistry*, 1980, **19**, 335-349.

91. Schechter I., Berger A., *Biochem. et Biophys. Res. Commun.*, 1968, **32**, 898-902.

92. Yeh-Kai L., Akusjarvi G., Alestrom P., Petterson U., Tremblay M., Weber J., *J. Mol. Biol.*, 1983, **167**, 217-222.

93. Krausslich H.G., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 3213-3217.

94. Halm C.S., Strauss J.H., *J. Virol.*, 1990, **64**, 3069-3073.

95. Moon J.B., Coleman R.S., Hanzlik R.P., *J. Am. Chem. Soc.*, 1986, **108**, 1350-1351.

96. Anderson G.W., Zimmerman J.E., Callahan F.M., *J. Am. Chem. Soc.*, 1967, **89(19)**, 5012.

97. Schotten C., *Chem. Ber.*, 1884, **17**, 2544. and Bauman E., *Chem. Ber.*, 1886, **19**, 3218.

98. Grassmann W., Wünsel E., Reipel H., *Chem. Ber.*, 1958, **91**, 455.

99. Bergmann M., Zervas L., Fruton J.S., Scheider F., Scheleich H., *J. Biol. Chem.*, 1938, **109**, 325-346.

100. Weygand F., Stieglich W., *Chem. Ber.*, 1960, **93**, 2983-3005.

101. Boisonnas R.A., Prietner G., *Helv. Chimica. Acta*, 1953, **36**, 875-886.

102. Abderhalden E., Suzuki S., *Z. Physiol. Chem.*, 1928, **176**, 101-108.

103. Anantharamaiah G.M., Sivanandaiah K.M., *J. Chem. Soc.*, 1977, Perkin 1, 490-491.

104. Olah G.A., Subhash C., Narang B.G., Gupta B., Malhotra R., *Angew. Chem. Int. Ed. Eng.*, 1979, **18**, 612-614.

105. Schmidt A.R., *Aldrichimca Acta*, 1981, **14**, 31-38.

106. Still W.C., Kahn M., Mitra A., *J. Org. Chem.*, 1978, **43**, 2923-2925.

107. Lipton M.F., Sorenson C.M., Sadler A.C., *J. Organomet. Chem.*, 1980, **186**, 155-158.

APPENDIX ONE

APPENDIX ONE.

		[Assay]
PROTEINASE ASSAY:	10 µl Proteinase	~300 nM
	10 µl Activating Peptide	11.4 µM
	25 µl Tris EDTA pH 7.8	21 mM/3.8 mM
	10 µl Substrate (AcLRGAGRSR)	342 µM
	10 µl Inhibitor in DMSO or MeOH	Given below

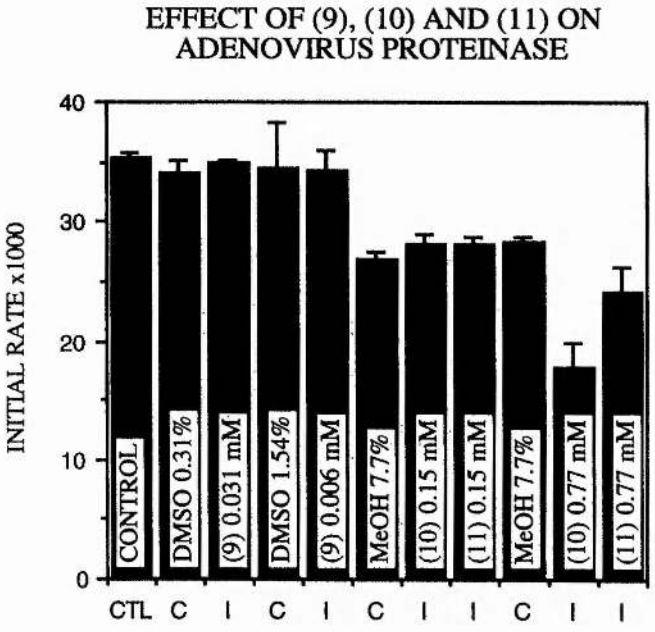
RESULTS	INITIAL RATE x1000						
	Assay 1	2	3	**Mean**	s.e.m.	% Inhibition	Estimated K_j
CONTROL	35.06	34.58	36.29	**35.31**	0.51		
DMSO 0.31%	34.81	35.51	31.65	**34.00**	1.19		
(**9**) 0.031 mM in DMSO	34.98	35.36	34.39	**34.91**	0.28	0	
DMSO 1.54%	31.35	29.70	42.13	**34.39**	3.90		
(**9**) 0.006 mM in DMSO	33.76	37.41	31.45	**34.21**	1.73	0	
MeOH 7.7%	25.53	27.05	27.96	**26.85**	0.71		
(**10**) 0.15 mM in MeOH	28.50	27.05	29.11	**28.22**	0.61	0	
(**11**) 0.15 mM in MeOH	26.58	29.63	28.01	**28.07**	0.88	0	
MeOH 7.7%	27.44	28.50	28.91	**28.28**	0.44		
(**10**) 0.77 mM in MeOH	17.60	13.94	21.56	**17.70**	2.20	37.41	2.1 mM
(**11**) 0.77 mM in MeOH	19.63	25.61	27.17	**24.14**	2.30	14.64	5.3 mM

Conclusions:

No inhibition or the proteinase was obtained with **(9)** at either concentration used.

No inhibition of the proteinase was obtained with **(10)** or **(11)** at 0.15 mM.

Some inhibition was observed with the higher concentration of **(10)** and **(11)** as tabulated. Due to solubility problems a higher concentration of **(9)** was not achieved.



C = Matched Control For Inhibitor I