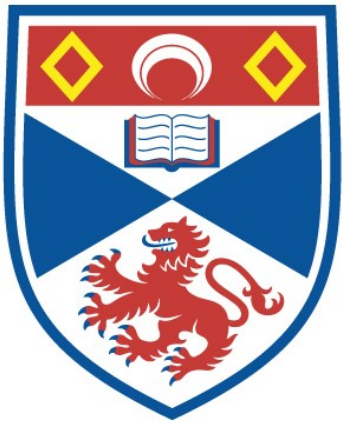
THE MECHANISM OF ACTIVATION OF THE ADENOVIRUS TYPE 2 PROTEASE

Gonçalo José Martins Cabrita

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



1997

The Mechanism of Activation of the Adenovirus Type 2 Protease

by

Gonçalo José Martins Cabrita

Division of Cell and Molecular Biology

University of St. Andrews, St. Andrews

A thesis presented for the degree of Doctor of Philosophy

at the University of St. Andrews, September 1997

Declarations

I, Gonçalo José Martins Cabrita, hereby certify that this thesis, which is approximately 63,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: 15/9/97 Signature of candidate:

I was admitted as a research student in February 1994 and as a candidate for the degree of Doctor of Philosophy in February 1995; the higher study for which this is a record was carried out in the University of St. Andrews between 1994 and 1997.

Date: 15/9/97 Signature of candidate:

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

Date: 15/9/97 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: 15/9/97 Signature of candidate:

Para os meus Pais, a quem tudo devo.

Acknowledgements

Words are not enough to adequately thank my supervisor, Dr. Graham Kemp, for his invaluable help, support, encouragement, friendship and infinite patience throughout all of my work, which made it a very pleasant experience to go through these years in St. Andrews.

Thanks also go to Tony Vaughan (FPLC), Paul Talbot (peptide synthesis), Stuart Annan (cloning), Dr. Sarah Jones (cloning), Dr. Helen Reddy (CE), Dr. Heather Murray (fluorescence) and Dr. Munir Iqbal (protease expression and purification) for putting up with me in Lab 6M and for all the help with several of the techniques used.

Some Senior Honours students also contributed with their work, and I am grateful to Catherine Hawkins, Victoria Reid and Brett Pollard for all their help. I also thank Dr. Carl Anderson for providing the vector plasmid, Ian Davidson for mass spectrometry and Ian Armitt for his advice with the fermenter.

Life in St. Andrews was made very enjoyable thanks to the many friends I met in here, and I am especially grateful to Luca and Giulia Del Zanna, Konstantin Spanoudakis and Fiona Proctor, Alex Gomez and Susana Ayllón-Martín, César Gonzalez, Luzia Piñon and Kumar, Giovanna Spera, Paola Avogadro, Grazia Botti, Francesca Tolve, Diego Angemi and Tatiana Toledo, Lewis Murray, Joana Desterro, Miguel Ferreira and Teresa Marques, Andrea Teti, Steffen Hühn, Dominic Jacobs, Steven Ogilvie, Catherine Heatlie and Catriona Davidson for being there and sharing their friendship with me. A particular thanks also goes to Christine Kolokotroni, for having an unbelievable patience with my late correspondence!

In Portugal, quite a few friends kept me going, and I want to thank Benjamim Pinto, Filipe and Ana Teles, Paulo Teodora and Ana Oliveira, João Pedro Carvalho, Nuno Alemão and Susana Sousa, and José Miguel and Cláudia Fernandes for the meaning they gave to the word 'friend'. And I am deeply indebted to my teacher Dra. Teresa Andrade for her friendship and for having developed my interest in chemistry.

Last but not least, I want to express my appreciation to my parents, brothers and sister for everything: love, encouragement and patience.

This work was supported by Junta Nacional de Investigação Científica e Tecnológica, programme PRAXIS XXI, Portugal, to whom I am indebted.

Abbreviations

β-MeOH - β-mercaptoethanol

A - adenine

Boc - benzyloxycarbonyl

bp - base pairs

C - cytosine

CAPS - 3-[cyclohexylamino]-1-propanesulfonic acid

CE - capillary electrophoresis

CM - carboxymethyl

dATP - deoxyadenosine triphosphate

DBP - DNA-binding protein

dCTP - deoxycytidine triphosphate

DEAE - diethylaminoethyl

dGTP - deoxyguanosine triphosphate

DMF - dimethylformamide

DMSO - dimethyl sulphoxide

DNA - deoxyribonucleic acid

DTT - dithiothreitol

dTTP - deoxythiamine triphosphate

EDTA - ethylenediaminetetracetate

Fmoc - fluorenylmethoxycarbonyl

FPLC - fast protein liquid chromatography

G - guanine

HOBt - hydroxybenzotriazole

LB - Luria-Bertani

MES - 2-(N-morpholino)-ethanesulphonic acid

mRNA - messenger RNA

Mtr - methoxytrimethylbenzene sulphonyl

NF - nuclear factor

NMR - nuclear magnetic resonance

OD - optical density

Opfp - O-pentafluorophenyl

OtBu - O-ter-butyl

p.f.u.	- plaque forming units
PBS	- phosphate buffer saline
PCR	- polymerase chain reaction
PDA	- piperazine di-acrylamide
PITC	- phenylisothiocyanate
PTH	- phenylthiohydantoin
RNA	- ribonucleic acid
SBTI	- soya bean trypsin inhibitor
SDS-PAGE	- sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	- thymine
tBu	- ter-butoxy
TFA	- trifluoroacetic acid
Tris	- tris(hydroxymethyl)aminomethane
Trt	- trityl
UV	- ultraviolet

Symbols for Amino Acids

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

Abstract

The adenovirus codes for a protease which is essential for virion infectivity. This protease requires the presence of a peptide cofactor in order to develop optimal activity. This peptide, GVQSLKRRRCF, originates from the C-terminal of a viral protein, pVI, and some evidence regarding its specificity came from observations showing that neither of the peptides GVQSLKRRRAF or KRRRCF was able to activate the protease, indicating that both the cysteine and the N-terminal were important in the activation process. However, the mechanism by which the peptide activates the protease has never been elucidated.

In this project, several factors contributing to the activation mechanism of the human adenovirus type 2 protease were studied, such as the peptide N-terminal length and composition, the environment close to the cysteine and the distance between the N-terminal and the cysteine, in view of assessing the relevance of each of these parameters in the activation process and proposing a mechanism of activation.

Based on the above studies, attempts of protease inhibition were also performed based on the activation process rather than on the blocking of the active site, and the relevance of these results was related with the proposed activation mechanism.

An attempt to clone an avian adenovirus protease was also performed, in order to try and compare the activation processes between the two proteases.

Table of Contents

1. Introduction

Lying in the frontier between live beings and inanimate substances, viruses represent one of the most fascinating examples of evolution towards a very compact, yet extremely effective form of parasitism.

Structurally, viruses consist of a protein shell which protects their genetic material, and in some cases an outer membrane made of lipids and glycoproteins envelops the protein capsid. Whilst outside a host, they present no biological activity whatsoever, rather behaving like 'inorganic' substances, waiting for the appropriate moment to unleash their infective power.

By themselves, viruses are not able to reproduce, requiring the invasion of a host cell in order to use their cellular machinery for the replication of DNA and protein synthesis, so that new virions can be produced. Throughout this operation, they subvert the cell cycle equilibrium, virtually shutting down many unessential cellular processes and using others for their own benefit, so that at the end of their infectious cycle the host cell has basically turned into a virion factory, usually at the expense of its own viability or, in some instances, resulting is such a disorganised cell that it loses control of its own replication and becomes malignant – a tumour is born. In some instances, however, viruses incorporate their genetic information into the genome of their host without destroying it, thus becoming permanent 'guests' waiting for any opportunity to produce outbursts of new infectious particles.

How can a virus be stopped? Ideally, in organisms with an immunological defence system, this deals with the invading virion before it penetrates any cell, or in the situation where cells have already been infected, the immune system chooses to destroy these cells before new virions are released. However, this system is unfortunately not always completely efficient, and other means have to be sought for. In this search for antiviral agents, an understanding of the viral infection mechanisms is necessary for potential weaknesses to be identified and exploited.

One of the 'weaknesses' of the adenovirus is that it needs to use a viral-coded enzyme to process some of its capsid proteins in order to become infectious. Therefore, a study of the operating mechanism of this molecule should enable the devising of a mode of inhibiting its action, thus allowing for a way of stopping the production of more virions. The broad aim of this project was then to unveil some of the characteristics of

the mechanism by which this enzyme becomes active, so that possible inhibitory substances can be conceived.

1.1. The Adenovirus

Adenovirus virions were observed (Horne *et al.*, 1959) to present an icosahedral shape (cf. Figure 1.1), with a diameter of 60 to 90 nm. They contain double-stranded DNA, a genome of 34 to 36 kbp with the potential to code for at least 50 proteins (Hierholzer *et al.*, 1991). They are responsible for several pathologies, including ocular, respiratory, and gastrointestinal infections (Wadell, 1994).

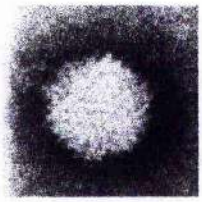


Figure 1.1 - Electron micrograph of human adenovirus type 2 virion (Norrby, 1969b).

Several aspects of adenovirus history, classification, pathogenicity, morphology and molecular biology are discussed below, in view of a better understanding on the importance of studying this virus.

1.1.1. History

In 1953, Rowe and co-workers (Rowe *et al.*, 1953) explanted surgically removed adenoids from children in order to evaluate different tissues for growth of polio viruses. After 4 weeks, the epithelial-like cells of some of the cultures exhibited a slowly progressive cytopathic effect. This change proved to be the result of the replication of previously unidentified viruses present in the adenoid tissues. One year later, Hilleman and Werner (Hilleman and Werner, 1954) isolated by tissue culture technique agents from respiratory secretions that induced cytopathic changes in human cells. An increase in neutralising and complement-fixing antibodies to the virus was shown to occur in the patients. The viruses discovered by the two groups were soon shown to be related (Huebner *et al.*, 1954) and were initially called *adenoid degeneration* (AD), *respiratory illness* (RI), *adenoidal-pharyngeal-conjunctival* (APC), or *acute respiratory disease* (ARD) agents.

Due to this variability in the nomenclature employed by the various laboratories involved in the study of these viruses, a group of researchers met and agreed on the

name *adenoviruses* (Enders *et al.*, 1956) as being generally acceptable and in accordance with the recommendations of the Subcommittee on Viruses of the International Nomenclature Committee, based on the original source of tissue (adenoid) in which the prototype viral strain was discovered.

It was in 1962 that a seminal discovery was made (Trentin *et al.*, 1962), that human adenovirus type 12 induced malignant tumours following inoculation into new-born hamsters. It was the first time that a human virus was shown to be oncogenic. Almost simultaneously, a report (Huebner *et al.*, 1962) described a similar property with human adenovirus type 18. Subsequently, further reports confirmed the oncogenicity of various other human and non-human adenoviruses in several species of rodents (Darbyshire, 1966; Girardi *et al.*, 1964; Huebner *et al.*, 1965; Huebner *et al.*, 1964; Huebner *et al.*, 1963; Hull *et al.*, 1965; Pereira *et al.*, 1965; Rabson *et al.*, 1964; Sarma *et al.*, 1965; Yabe *et al.*, 1964). The tumours formed have usually the characteristics of undifferentiated sarcomas, although malignant lymphomas were also observed sporadically (Larsson *et al.*, 1965). However, no evidence of linking of adenoviruses with malignant disease in humans has been reported so far; although there is one report of adenovirus-related RNA in neurogenic tumours (Ibelgaufts *et al.*, 1982), thorough research has failed to find adenovirus nucleic acids in human tumours (Green *et al.*, 1980; Mackey *et al.*, 1976). Yet, the ability of inducing tumours in animals, as well as to transform cells, has established adenoviruses as a precious model system for probing the mysteries of oncogenesis (Shenk, 1996).

The above *in vitro* transformation of cultured cells by adenoviruses was first demonstrated by McBride and Wiener (McBride and Wiener, 1964), who infected cultures of new-born hamster kidney cells with human adenovirus type 12 and detected transformed cells about 8 to 10 weeks later. Subsequently, a report (Freeman *et al.*, 1967) showed the transformation of rat embryo fibroblasts by the same adenovirus. The studies of adenovirus-infected cells have progressed much since then, and these have made innumerable contributions to the understanding of viral and cellular gene expression and regulation, DNA replication, cell cycle control, and cellular growth regulation (Shenk, 1996).

One of the major contributions to modern biology of the adenovirus system was however to host the discovery of mRNA splicing. Studies on the biogenesis of viral mRNA first showed that many mRNAs were produced from a large nuclear transcript

(Bachenheimer and Darnell, 1975; Weber *et al.*, 1977), and subsequent analysis of the adenovirus mRNA structure revealed the existence of introns (Berget *et al.*, 1977; Chow *et al.*, 1977). Nowadays, the usefulness of adenoviruses as vectors for gene therapy is the subject of intense exploration (Shenk, 1996).

1.1.2. Classification

Over 100 serotypes of adenovirus are classified within the family *Adenoviridae*, possibly the most diverse family of DNA viruses (Kitchingman, 1994). However, as not all of the family members contain the same common antigen, two genera have been established, based on a common antigen characterisation, which resides in the hexon of the outer capsid (Pereira *et al.*, 1963): *Aviadenovirus* (lack a common antigen) and *Mastadenovirus* (share genus-specific antigen – bovine adenovirus serotypes 4-8 are an exception (Wadell, 1994)).

1.1.2.1. Aviadenoviruses

This genus is limited to viruses of birds. The known hosts are fowl, turkey, goose, pheasant and duck. Aviadenoviruses can only replicate in avian cells, preferably from the homologous species (Russell, 1994).

The prototype avian adenovirus is also known as CELO (Chicken Embryo Lethal Orphan), and it has been the major subject of avian adenovirology for a number of years. The interest was partially based on the fact that CELO causes tumours in baby hamsters (Sarma *et al.*, 1965). Yet, interest in CELO virus has waned in recent years, as there are few serious health or economic consequences of its infection (Chiocca *et al.*, 1996).

Another avian adenovirus was discovered to cause an epidemic of reduced laying and soft-shell or shell-less eggs in some flocks of chickens in 1976, and was thus denominated EDS'76 (Egg Drop Syndrome 1976 virus) and shown to be a duck adenovirus (Russell, 1994), strain 127.

1.1.2.2. Mastadenoviruses

This genus comprises virus that infect mammals, including human, simian, bovine, equine, porcine, ovine, canine, murine and opossum viruses.

The serotype is defined through quantitation of neutralisation with hyperimmune sera, and the ratio of homologous to heterologous neutralisation titer must be greater

than 16. The designation of recombinant should be used only when the two parent genomes have been identified (Wadell, 1994).

There are 47 human adenovirus serotypes (cf. Table 1.1) known to date (Shenk, 1996), and these are distinguished on the basis of their resistance to neutralisation by antisera to other known serotypes. The type-specific neutralisation is predominantly the result of antibody binding to epitopes on the virion hexon protein and the terminal knob portion of the fiber protein (Norrby, 1969b; Toogood *et al.*, 1992). Various other schemes of classification have produced similar groupings, suggesting that the classification based on hemagglutination is a reasonable standard (Shenk, 1996).

Table 1.1 - Classification and properties of human adenoviruses (adapted from (Shenk, 1996; Wadell, 1994)).

Subgroup	Hemag-glutination Group	Serotypes	Oncogenic Potential (Tumours in Animals)	% of GC in DNA	Tropism Symptoms
A	IV (little or no agglutination)	12, 18, 31	High	48-49	Cryptic enteric infection
B	I (complete agglutination of monkey erythrocytes)	3, 7, 11, 14, 16, 21, 34, 35	Moderate	50-52	Respiratory disease; persistent infections of the kidney
C	III (partial agglutination of rat erythrocytes)	1, 2, 5, 6	Low or none	57-59	Respiratory disease persists in lymphoid tissue
D	II (complete agglutination of rat erythrocytes)	8, 9, 10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36-39, 42-47	Low or none (mammary tumours)	57-61	Keratocon-junctivitis
E	III	4	Low or none	57-59	Conjunctivitis; respiratory disease
F	III	40, 41	Unknown	52	Infantile diarrhoea

Although there is a noticeable variability within all the adenoviruses, they all present a similar structure and genome organisation, and many functional proteins exhibit a high conservation among the viruses of different subgroups. Therefore, the descriptions of morphology and molecular biology in the following sections will be based on the human adenovirus serotype 2 (Ad2), whose complete genome is known since 1984 (Roberts *et al.*, 1984), and also this is the serotype to which the protease studied in this work belongs.

1.1.3. Pathogenicity

There is a wide range of pathogenicity among the 47 human adenovirus serotypes (Wadell, 1994): Ad7 is the most severe cause of respiratory infections, whereas within all of the 28 serotypes of subgenus D which have a preference for the eye, only Ad8 is responsible for the most severe outbreaks of keratoconjunctivitis. Even within a serotype there are differences: Ad19 prototype has seldom been isolated for the past three decades whereas its variant Ad19a has been the cause of numerous outbreaks of keratoconjunctivitis since 1973.

Serotypes 1, 2, 5 and 6 are particularly involved in sporadic infections in adults, and these are the types commonly found latent in human adenoids and tonsils (Rowe, cited in (Andrewes *et al.*, 1978)). Serotypes 3, 4, 7, 14 and 21 in particular cause outbreaks of fever and pharyngitis, especially in service recruits and boarding schools, and the occurrence of conjunctivitis in some of those outbreaks has led to the use of the term 'pharyngoconjunctival fever' (Andrewes *et al.*, 1978).

Ad18 and Ad31 were obtained from infants with gastrointestinal disease, but it is still difficult to evaluate their role as a cause of diarrhoea. Ad14 causes outbreaks among military recruits, Ad11, Ad34 and Ad35 cause persistent infections of kidneys. Ad4 (subgenus E) is responsible for epidemic outbreaks of respiratory disease among military recruits, and a distinct genome Ad4a is second only to Ad8 as a cause of adenovirus-associated eye disease in Japan. Ad40 and Ad41 are second only to rotaviruses as a cause of infantile diarrhoea, but in contrast to rotaviruses they cause this illness throughout the year (Wadell, 1994).

1.1.4. Morphology

The adenoviruses are non-enveloped viruses, icosahedral in shape, and with a diameter ranging from 70 to 100 nm (cf. Figure 1.2). The virions are composed of DNA

(13% of total mass), protein (87% of total mass) and trace amounts of carbohydrates, as the fiber protein of the virion is modified by addition of glucosamine (Green and Piña, 1963; Ishibashi and Maizel, 1974a; Ishibashi and Maizel, 1974b). The virus consists of a protein shell enclosing a DNA-containing core, and the shell (capsid) is made of 252 subunits (capsomeres), of which 240 are hexons (surrounded by six neighbours) and 12 are pentons (surrounded by five neighbours) (Ginsberg *et al.*, 1966). Each penton is located at a vertex of the particle and it consists of a base anchored in the capsid and a projecting fiber, whose length varies among different serotypes (Norrby, 1966; Norrby, 1969a).

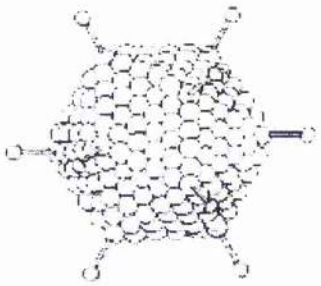


Figure 1.2 - Scheme of an adenovirus virion.

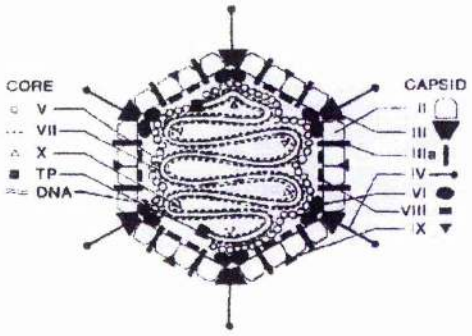


Figure 1.3 - Schematic section of an adenovirus particle. Virion constituents are designated by their polypeptide numbers with the exception of the terminal protein (TP). From (Stewart *et al.*, 1991).

The virion capsid is constituted of seven polypeptides (cf. Figure 1.3). Polypeptide II is the most abundant constituent, as hexon capsomeres are made of trimers of this protein. Polypeptides VI, VIII and IX were shown to be associated with the hexon (Everitt *et al.*, 1973), and these probably stabilise the hexon capsomere lattice, with polypeptides VI and VIII possibly bridging between the capsid and core components of the virion. Polypeptide III associates in pentameres to form the penton base protein (van Oostrum and Burnett, 1985), whereas polypeptide IIIa associates with hexon units that surround the penton, probably linking adjacent facets of the capsid and bridging between hexons and polypeptide VII of the core (Everitt *et al.*, 1975; Everitt *et al.*, 1973). Polypeptide IV trimerises to form the fiber protein (van Oostrum and Burnett, 1985) that projects from the penton base, and to the combination of penton base and fiber the designation penton capsomere is given.

The virion core contains four known proteins, as well as the viral genome. Polypeptides V, VII and mu are basic, arginine-rich components of the core (Hosakawa and Sung, 1976; Russell *et al.*, 1968), and all these proteins contact the viral DNA (Anderson *et al.*, 1989; Chatterjee *et al., 1986*). Although the function of mu protein is unknown (Hosakawa and Sung, 1976), polypeptide VII (the major core protein) is supposed to serve as a histonelike centre around which viral DNA is wrapped (Chatterjee *et al., 1986;* Mirza and Weber, 1982), and polypeptide V can bind to a penton base (Everitt *et al.*, 1975) and it might bridge between the core and the capsid, positioning one relative to the other. The remaining protein present in the core, the terminal protein, is covalently attached to the 5' ends of the viral DNA, and only two copies of it exist per virion. It serves as a primer for DNA replication (Challberg *et al.*, 1980; Challberg and Kelly, 1981; Enomoto *et al.*, 1981; Lichy *et al.*, 1981; Tamanoi and Stillman, 1982), and it mediates attachment of the viral genome to the nuclear matrix (Bodnar *et al.*, 1989; Fredman and Engler, 1993; Schaack *et al.*, 1990).

1.1.5. Molecular Biology

Adenoviruses have provided an invaluable tool for pioneering studies on elementary cellular mechanisms in molecular biology. The organisation of genetic functions on the viral genome, mapping and temporal regulation of viral gene expression, replication of the viral genome, splicing of RNA, the importance of *trans*-activating or enhancing genetic functions, promoter nucleotide sequence motifs interacting with specific host proteins, the role of sequence-specific promoter methylations in the long-term silencing of genes, translational control by low-molecular-weight 'virus-associated' (VA) RNA, integration of viral (foreign) DNA into the host genome, specific shut-off of cellular genes in adenovirus-transformed cells, and other principles, have been recognised by using the adenovirus system (Doerfler, 1994).

1.1.5.1. Genome Organisation

The adenovirus genome consists of a single linear, double-stranded DNA molecule, with a molecular weight in the range of 20-25$\times 10^6$ daltons (Green *et al.*, 1967; Van der Eb *et al.*, 1969), ranging from 36 to about 45 kb (Kitchingman, 1994).

Encoded in the viral chromosome are five early transcription units (E1A, E1B, E2, E3 and E4), two delayed early units (IX and IVa2), and one late unit (major late) that is processed to generate five families of late mRNAs (L1 to L5), all of which are

transcribed by RNA polymerase II, cf. Figure 1.4 (reviewed in (Pettersson and Roberts, 1986)). The chromosome contains also one or two (depending on the serotype) viral associated (VA) genes transcribed by RNA polymerase III. The genome map is represented by convention with the E1A gene at the left end, and both strands of the viral DNA are transcribed with the rightward reading strand on the conventional map coding for the E1A, E1B, IX, major late, VA RNA and E3 units, and the leftward reading strand coding the E4, E2 and IVa2 units.

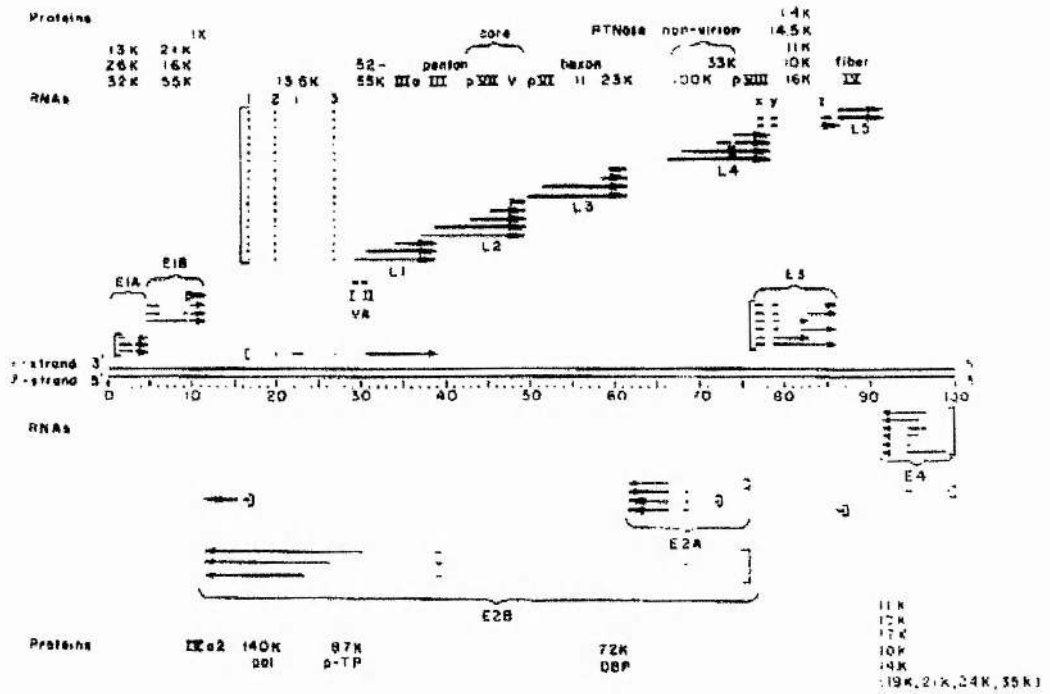


Figure 1.4 - Map of transcription and translation of adenovirus type 2. Early mRNAs are designated E and late mRNAs are designated L. Polypeptides are identified by the conventional numbering system (Roman numerals) for the virion structural components and by size (kDa) for non-structural peptides (Shenk, 1996).

Each of the adenovirus genes transcribed by RNA polymerase II originates multiple mRNAs that are differentiated by alternative splicing, and in some cases by the use of different poly(A) sites. As mentioned before, it was the analysis of adenovirus mRNA structure that led to the discovery of splicing (Berget *et al.*, 1977; Chow *et al.*, 1977).

Several individual adenovirus transcription units code for a series of polypeptides with related functions. As discussed further below, the E1A unit encodes two proteins that activate transcription and induce the host cell to enter the S phase of the cell cycle; E1B codes for two proteins which cooperate with E1A products to induce cell growth; E2 encodes three different proteins, all directly related to DNA replication; E3 encodes

products that modulate the response of the host to the adenovirus infection; and the late family of mRNAs is related with the production and assembly of capsid elements. The E4 unit is the only that seems to code for a diverse set of functions, mediating transcriptional regulation, mRNA transport and DNA replication (Shenk, 1996). The VA RNAs do not code for any polypeptide, but rather for polymerase III products that help in the regulation of the translation of viral mRNAs, perhaps also in the shut-off of cellular protein synthesis in adenovirus-infected cells (Doerfler, 1994).

1.1.5.2. Infectious Cycle

The replication cycle of the adenovirus is conventionally divided into two phases, separated by the onset of viral DNA replication (cf. Figure 1.5). The designation of *early* events applies to those which start as soon as the infecting virus interacts with the host cell, and includes adsorption, penetration, transcription, and translation of an early set of genes. The early viral gene products mediate viral gene expression and DNA replication, induce cell cycle progression, block apoptosis, and antagonise a variety of host antiviral measures. In HeLa cells infected at a multiplicity of 10 plaque-forming units (p.f.u.) per cell, the early phase lasts for about 5 to 6 hours, after which viral DNA replication is first detected. Concomitant with the onset of viral DNA replication, the late phase of the cycle begins with expression of a new set of *late* viral genes and assembly of progeny virions. The infectious cycle is completed after 20 to 24 hours in HeLa cells, and at the end of that, circa 10^4 progeny virus particles per cell have been produced, along with the synthesis of a substantial excess of virion proteins and DNA that are not assembled into virions (Green and Daesch, 1961). Cells infected at high multiplicity seldom divide (Horwitz, 1971), and therefore at the end of the infectious cycle the DNA and protein content of the infected cell has increased about twofold (Shenk, 1996).

Early and *late*, although convenient terms for the description of events occurring during the replication cycle, do not correspond to a sharp functional distinction between early and late events, as this separation is often unclear. For example, early genes are still expressed at late times after infection, and the promoter which controls expression of the major late transcription unit also directs a low transcription level early after infection. The viral genes that code for proteins IVa2 and IX begin to be expressed at an intermediate time (Pettersson and Roberts, 1986) and therefore form a "delayed early" category.

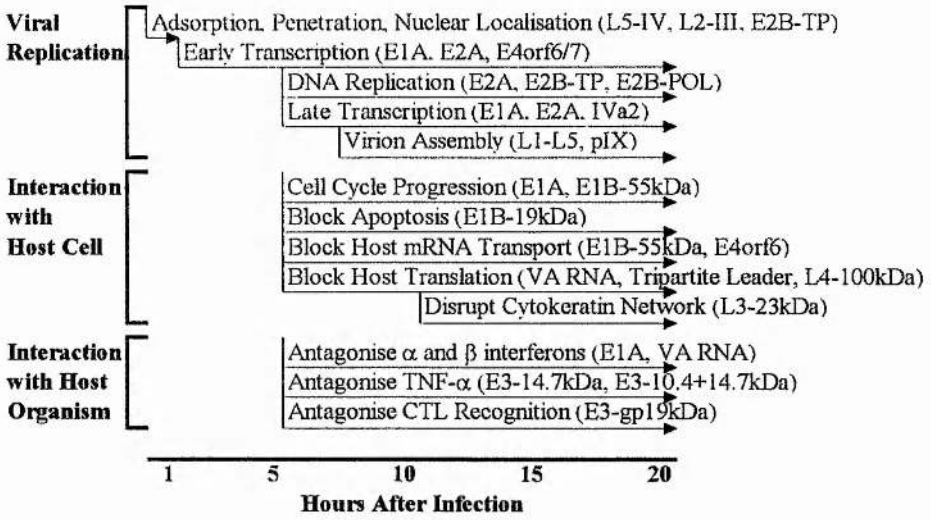


Figure 1.5 - Schematic diagram depicting the relative timing of the main events occurring during the adenovirus infection cycle. Viral gene products known to mediate each event are listed in parentheses. Times for stages of viral replication and interactions with host cell are as per growth of HeLa cells infected at a multiplicity of 10 p.f.u./cell. HeLa cells die at 20 to 24 hours after infection. Times for interaction with host organism are artificial and meant to place interactive events in a time frame relative to that of the viral replication cycle (from (Shenk, 1996)).

1.1.5.2.1.Adsorption and Entry

The attachment of adenovirus type 2 to cells is mediated by the fiber protein (Londberg-Holm and Philipson, 1969), whose distal knob domain at the C-terminal of the protein is thought to bind to a cellular receptor (Devaux *et al.*, 1987). The identity of this receptor is still a mystery, although three cellular membrane polypeptides have been captured on a penton-fiber affinity matrix and shown to inhibit Ad2 attachment to cells (Hennache and Boulanger, 1977). The finding that Ad2 binds to some cells but does not enter them effectively (Silver and Anderson, 1988) suggested the possibility that a second protein-protein recognition event should occur for internalisation, and this has now been identified. The penton base protein binds to specific members of a family of heterodimeric cell surface receptors designated *integrins* (Wickham *et al.*, 1993). Yet, the penton-integrin interaction on its own is not sufficient for viral binding to cells, as soluble fiber or fiber antibodies can block adsorption completely. Therefore, adsorption and internalisation is a two-component process which requires the interaction of both fiber and penton proteins with their related cellular targets.

Once adsorbed, the Ad2-receptor complexes diffuse into coated pits and are internalised by receptor-mediated endocytosis (Chardonnet and Dales, 1970a;

Chardonnet and Dales, 1970b; FitzGerald *et al.*, 1983; Varga *et al.*, 1991). The process is triggered by the penton-integrin interaction, since purified penton but not fiber protein is rapidly internalised by cultured cells (Wickham *et al.*, 1993). This process is quite efficient (Greber *et al.*, 1993), as 80-85% of the viruses that bind to the surface of a susceptible cell are internalised, and penetration occurs quickly, with half of the adsorbed viruses moving to the endosomes within 10 minutes, with a subsequent 90% of the latter moving successfully to the cytosol within a half time of circa 5 minutes. This movement to the cytosol is somehow triggered by the acidic pH of the endosome (Pastan *et al.*, 1986; Seth *et al.*, 1984; Svensson, 1985), and the penton base is thought to play an essential role in the process.

The virus particles are then transported through the cytosol to the nucleus by a process that probably involves microtubules (Dales and Chardonnet, 1973; Luftig and Weihing, 1975). About 40 minutes after penetration, virus particles can be seen at nuclear pore complexes, suggesting that release of DNA is occurring at the nuclear membrane. After 2 hours, about 40% of the internalised particles have released their DNA free from hexon proteins, although it is unknown which portion of the released DNA localises to the nucleus.

There is a sequential disassembly of the virion during the internalisation process (Greber *et al.*, 1993). This happens by selective dissociation and proteolytic degradation of the virion constituents. The process starts with the loss of the proteins at the vertices of the particle: the fiber, the polypeptide IIIa, located near the peripentonal hexons (cf. 1.1.4), and the penton are considerably lost by 15 minutes after penetration in the cell. Polypeptide VIII detaches form the particle as the penton capsomeres are lost, and degradation of polypeptide VI occurs shortly after the particle entry into the cytosol. The internal localisation of this polypeptide suggests that it might be degraded by the virus-coded protease (Mangel *et al.*, 1993; Tihanyi *et al.*, 1993; Webster *et al.*, 1993). As both polypeptides VI and VIII bridge from the DNA core to the capsid, their loss should prepare the particle for the release of its DNA. Somewhat later, polypeptide IX is released from the infecting particle, and finally the DNA-containing core is freed from hexons.

When the viral DNA reaches the nucleus, it associates with the nuclear matrix through its terminal protein (Bodnar *et al.*, 1989; Fredman and Engler, 1993; Schaack *et al.*, 1990). Apparently, this protein which arrives with the infecting genome is the first

viral gene product that functions within the nucleus to initiate the program of viral gene expression.

1.1.5.2.2.Activation of Early Viral Genes

Early adenovirus gene expression has three main objectives: (a) to induce the host cell in entering the S phase of the cell cycle, thus providing with an optimal environment for viral replication, in which E1A and E1B gene products are involved; (b) to set up viral systems protecting the infected cell from various antiviral defences of the host organism, where E3 and VA RNA genes contribute to these defences; and (c) to synthesise viral gene products needed for viral DNA replication. All these objectives rely on the transcriptional activation of the viral genome, and the main activating proteins are coded within the E1A gene (reviewed in (Nevins, 1992; Shenk and Flint, 1991)).

E1A is the first viral transcription unit to be expressed after the viral chromosome reaches the nucleus. This unit codes for two mRNAs during the early phase of infection which translate into two polypeptides usually referred to as 12S and 13S E1A proteins, and three additional E1A mRNA species accumulate later in the infectious cycle (Stephens and Harlow, 1987; Ulfendahl *et al.*, 1987), but no definite function has been described for their products. The primary E1A translation products undergo extensive phosphorylation (Harlow *et al.*, 1985), but the purpose of these modifications is not clear at present. The two above proteins encoded in the E1A region present three conserved regions, CR1, CR2 and CR3, separated by less highly conserved domains (Dyson *et al.*, 1992; Moran and Mathews, 1987; Nevins, 1992; Shenk and Flint, 1991; van Ormondt *et al.*, 1980), and they do not exhibit sequence-specific DNA binding (Chatterjee *et al.*, 1988; Ferguson *et al.*, 1985); instead, they bind to cellular proteins and modulate their function. The three conserved regions in the E1A proteins define domains that play major roles in protein-protein interactions. E1A proteins activate the expression of other adenovirus transcription units by increasing the rate of transcription (Nevins, 1981), and since they activate those genes in *trans*, they are usually referred to as *trans*-activators. The mechanism of *trans*-activation occurs through direct interaction of E1A proteins with auxiliary factors that mediate basal transcription, with activating proteins that bind to upstream promoter and enhancer elements, and with regulatory subunits that influence the activity of DNA-binding factors, and the three domains mentioned above mediate these protein-protein interactions which activate transcription (Shenk, 1996).

The expression of the adenovirus VA RNA genes is also induced by E1A proteins (Hoeffler *et al.*, 1988; Hoeffler and Roeder, 1985; Yoshinaga *et al.*, 1986), and they are transcribed by RNA polymerase III, and their role in infection is explained below (cf. 1.1.5.2.6).

Other than E1A proteins, two more gene products have been observed to activate adenovirus promoters: E4-17 kd polypeptide (Hardy *et al.*, 1989; Huang and Hearing, 1989; Neill *et al.*, 1990; Raychaudhuri *et al.*, 1990), and the E2 DNA-binding protein (Chang and Shenk, 1990; Morin *et al.*, 1989), but the mechanism of transcription stimulation of the latter is unclear at present.

Generally, the early genes remain active during all of the viral replication cycle, although their rate of transcription decreases gradually, in part due to cell death. However, there are three known down-regulatory events, and in each of them it seems that a viral protein, accumulating as a response to an activation event, acts subsequently to inhibit continued transcriptional stimulation effected by one or more early promoter elements. First, E1A proteins can repress the activity of several known enhancers, including the enhancer lodged upstream of the E1A gene itself (Borelli *et al.*, 1984; Velcich and Ziff, 1985). Second, whereas E2 DNA-binding protein activates some promoters, it apparently inhibits transcription from the E4 promoter by an unknown mechanism (Handa *et al.*, 1983; Nevins and Winkler, 1980). Third, the induction of AP1 (a cellular transcription factor) activity by E1A and cyclic AMP is transient (Muller *et al.*, 1989), and is antagonised by E4-14 kd which accumulates in response to the E1A-mediated activation of the E4 gene (Muller *et al.*, 1992).

As soon as the early mRNAs are synthesised, they are translated on polysomes together with cellular mRNAs. If at first they do not seem to present any competitive advantage, as the infection passes on to the late phase, cellular mRNAs are removed from polysomes (Shenk, 1996).

1.1.5.2.3.Activation of the Host Cell

It has been known for long that adenovirus infections induce quiescent cells into starting the S phase of the cell cycle, thus creating an environment suitable for viral replication (reviewed in (Tooze, 1981)). As mentioned above, the function of E1A proteins is essentially to modulate the cell cycle, and this is accomplished by interaction with a retinoblastoma tumour suppresser protein, pRB, as this protein has been shown to inhibit cell cycle progression, causing cells to stop at mid-to-late G_1 (Goodrich *et al.*,

1991; Hinds *et al.*, 1992). E1A proteins disrupt a series of complexes which contain different pRB family members, multiple transcription factor subunits (E2F), and cyclins with associated kinases. These complexes regulate the cell cycle progression in a normal healthy cell, and their dissociation by E1A deregulates normal cell cycle control, allowing quiescent cells to start DNA synthesis as a consequence of infection.

However, there are two distinct regions in the E1A proteins which can induce cells to progress from G_1 to S through two independent mechanisms: the already mentioned binding to pRB family members, and also the binding to another cellular protein, p300. Presumably, the viral proteins utilise simultaneously the two mechanisms because they work more effectively than either alone, or because there may exist cells in which one pathway functions better than the other (Shenk, 1996).

Another viral early protein, E1B-55 kd, also modulates cell cycle progression. This protein aims at the cellular p53 tumour suppresser protein, akin to members of the family of pRB in that it also regulates progression from G_1 to S. It seems that p53 normally works as a component of a G_1 checkpoint that is induced by DNA damage (Diller *et al.*, 1990; Hartwell, 1992; Lin *et al.*, 1992), as it can block cell cycle progression and contribute to the activation of genes known to be induced by DNA damage. Currently, p53 is thought to transcriptionally activate genes that prevent entry into S phase, although it is not clear whether transcriptional repression by p53 also contributes to this process. The E1B-55 kd protein of Ad5 binds to p53 within infected cells (Sarnow *et al.*, 1982) and it can block transcriptional activation by p53. Therefore, like E1A proteins, E1B-55 kd protein binds to a tumour suppresser protein, antagonises its normal activity and helps deregulating the cell cycle progression. However, in contrast to E1A proteins, E1B-55 kd alone is not enough to stimulate quiescent cells to start the S phase of the cell cycle. Supposedly, therefore, the E1B-55 kd protein works together with the E1A proteins in activating more efficiently quiescent cells. Also, as E1A proteins somehow stabilise p53 by increasing its steady state level (Lowe and Ruley, 1993), E1B proteins must contribute to the activation of quiescent cells by preventing the implementation of a cell cycle blockage by too high levels of p53.

1.1.5.2.4.Inhibition of Apoptosis

As mentioned above, E1A proteins in quiescent cells can induce effectively cellular DNA synthesis, as well as transient cell proliferation. Yet, E1A expression is not sufficient to induce long-term growth of primary cells, as besides inducing proliferation,

E1A proteins also cause apoptosis. Apoptosis, also designated *programmed cell death*, is associated with well-defined nucleus morphological changes as well as DNA fragmentation; this distinguishes it from *necrosis*, cell death characterised by extensive cytoplasmic destruction induced by an adverse environment or by damage (reviewed in (White, 1993; Williams and Smith, 1993).

Apoptosis induced by E1A involves the induction of p53 (Debbas and White, 1993; Lowe and Ruley, 1993). High-level expression of p53 can either cause the blocking of the cell cycle progression (reviewed in (Perry and Levine, 1993)) or the induction of apoptosis (Clarke *et al.*, 1993; Lowe *et al.*, 1993; Shaw *et al.*, 1992; Yonish-Rouach *et al.*, 1991). E1A proteins stabilise p53, causing it to accumulate in the nucleus (Lowe and Ruley, 1993), as mentioned previously, but both Ad5 E1B proteins can prevent E1A-induced apoptosis (Rao *et al.*, 1992). Whilst the Ad5 E1B-55 kd protein probably blocks apoptosis by binding to p53 and altering its function, either the E1B-19 kd protein or the cellular Bcl-2 proto-oncoprotein can block E1A-induced apoptosis in a more efficient way than E1B-55 kd (Rao *et al.*, 1992). However, the mechanism of p53-induced apoptosis, as well as the mechanism of E1B-19 kd apoptosis blocking, remain unexplained. But, as either E1B-19 kd or Bcl-2 can block p53-mediated transcriptional repression (Shen and Shenk, 1994), it is possible that p53 might mediate apoptosis, at least partially, by repressing transcription.

Therefore, apoptosis is an example of a cellular response to infection, with the potential of inhibiting viral growth and preventing its spread within the infected organism. Yet, adenoviruses evolved in such a way that allowed the coding of a gene product that effectively blocks the cellular defence (Shenk, 1996).

1.1.5.2.5. Viral DNA Replication

As a consequence of the expression and subsequent action of E1A proteins, the infected cell enters the S phase of the cell cycle, thus becoming the optimal environment for viral DNA replication (reviewed in (Challberg and Kelly, 1989; Hay and Russell, 1989; Stillman, 1989)). In Ad2 replication starts at about 5 hours after infection of HeLa cells, at a multiplicity of 10 p.f.u. per cell, continuing until the host cell dies (cf. Figure 1.5).

The viral chromosome presents inverted terminal repeats which function as replication origins. *In vivo* studies agree with a model in which adenovirus DNA replication takes place in two stages (Lechner and Kelly, 1977). In the initial stage,

synthesis is initiated at any of the terminus of the linear double stranded DNA, proceeding continuously until the other end of the genome is reached. Of the two DNA strands, only one is used as a synthesis template, so that the replication products are a double stranded DNA composed of a parental and a daughter strand, together with a single strand of DNA. The second stage of the replication process comprises the complementation of the displaced single parental strand of DNA template. This strand forms a structure known as the "panhandle" by annealing its two inverted terminal repeats and thus forming a circle with a double stranded "handle" which is identical in structure to either termini of the initial linear duplex DNA. This similarity permits the same initiation of replication complex that operates in the first stage to recognise and generate a second complementary strand, thus creating another double stranded DNA composed of a parental and a daughter strand (Shenk, 1996).

The inverted terminal repeats include the origin of replication, and three functional domains have been established. The first domain, encompassing the first 18 base pairs of the DNA, is known as domain A. A cellular protein, *ORP-A*, is known to bind to the first 12 bp (Rosenfeld *et al.*, 1987), although its binding does not seem to be essential for replication. A complex formed by two viral proteins binds to the following 9 to 18 bp (Chen *et al.*, 1990; Mul and van der Vliet, 1992; Temperley and Hay, 1992), the terminal protein (TP) and the DNA polymerase. Both these proteins are coded in the E2 region of the viral genome, and TP is synthesised as a precursor (pTP) which is active in the initiation of DNA replication (Challberg *et al.*, 1980; Stillman *et al.*, 1981) and found covalently bound to the 5' ends of the viral chromosome, as mentioned earlier. This precursor is cleaved later on during virus assembly by the viral protease to generate TP, which remains bound to the genome (Challberg and Kelly, 1981), although it seems that both the cleaved portions of the protein remain associated with the genomic termini (Schaack *et al.*, 1990). The preterminal protein functions as a primer for the replication of DNA (Rekosh *et al.*, 1977), by preserving its terminal sequence integrity during several rounds of DNA replication. With regard to the DNA polymerase, it possesses 5' to 3' polymerase activity as well as 3' to 5' exonuclease activity, which is possibly used as a proof-reading function during the polymerisation process (Field *et al.*, 1984).

Domain B goes from base pairs 19 to 39, and domain C from 40 to 51, and although these are not essential for replication, they do enhance the efficiency of initiation, by binding cellular factors. Thus, domain B binds nuclear factor I (NF-I) and domain C

binds nuclear factor III (NF-III), but whereas NF-I is known to interact with the pTP-polymerase complex and stabilise it at the origin (Bosher *et al.*, 1990; Chen *et al.*, 1990; Mul and van der Vliet, 1992; Mul *et al.*, 1990), stimulated by a third viral protein (E2 single-stranded DNA-binding protein, DBP (Cleat and Hay, 1989; Stuiver and van der Vliet, 1990)), the mechanism by which NF-III stimulates initiation remains unclear, and perhaps the finding that this nuclear factor induces DNA bending when NF-III binds to it (Verrijzer *et al.*, 1991) will be related to its function. It is possibly interesting to note that both these nuclear factors are also transcription factors.

The elongation of the DNA chain requires two virus E2-coded proteins, the polymerase and DBP, and a cellular protein, nuclear factor II (NF-II). The function of DBP, which coats the single-stranded replication intermediates, seems to be the enhancement of the efficiency of processing of the polymerase (Field *et al.*, 1984; Lindenbaum *et al.*, 1986), which will probably enable it to travel along the entire length of the genome after initiation at the terminus.

1.1.5.2.6.Activation of Late Gene Expression and Host Cell Shutoff

As already mentioned earlier, the late genes start to be efficiently expressed after the start of viral replication. In adenoviruses, these genes are under the control of the major late promoter, which shows low levels of activity at early stages of infection, but becomes extensively more active at a later time after infection (Shaw and Ziff, 1980). Apparently, at least two factors contribute to the activation of this promoter: a *cis*-acting modification in the viral genome, and the induction of at least one more virus-coded *trans*-acting factor.

Several explanations have been put forward to describe the *cis*-acting genome modification. It is possible that the viral chromatin has to undergo some composition modification during DNA replication in order to activate the major late promoter, or perhaps the replication enables the access of transcription factors to the promoter, as histones or histone-like proteins are removed and later rebound to DNA during replication, allowing a binding competition to DNA between transcription factors and histones. Another possibility is that the viral genome might establish a compartmentalised environment in the nucleus, thereby taking some time for some cellular factors to reach the appropriate environment for late gene expression. Discrete centres at which replication and transcription occur have been observed in infected nuclei by electron microscopy (Martinez-Palomo and Granboulan, 1967; Moyne *et al.*, 1978). Whatever

the mechanism, a transcription factor termed USF or MLTF has been observed to bind to the upstream site in the major late promoter (Carthew *et al.*, 1985; Miyamoto *et al.*, 1985; Sawadago and Roeder, 1985) only after the start of DNA replication (Toth *et al.*, 1992), and when bound this factor is known to activate transcription. In addition to USF/MLTF, a viral transcription factor also contributes to the activation of the major late promoter, coded by the IVa2 gene (Tribouley *et al.*, 1994).

After the onset of viral DNA replication and synthesis of the late mRNAs, the cytoplasmic accumulation of cellular mRNAs is blocked (Beltz and Flint, 1979), although their synthesis continues, which suggests that some sort of blocking of their transport occurs. This is mediated by a complex of two viral proteins (Sarnow *et al.*, 1984), E1B-55 kd (Babiss and Ginsberg, 1984; Pilder *et al.*, 1986) and E4-34 kd (Halbert *et al.*, 1985; Weinberg and Ketner, 1986), and these same proteins are necessary for efficient cytoplasmic accumulation of viral mRNAs late after infection (Babiss *et al.*, 1985; Bridge and Ketner, 1990; Halbert *et al.*, 1985; Pilder *et al.*, 1986; Weinberg and Ketner, 1986). The distinction between viral and cellular mRNAs is not based on the identity of individual mRNAs. A model has been proposed (Ornelles and Shenk, 1991) in which the E1B-E4 complex relocalises a cellular factor necessary for mRNA transport from its synthesis site to the nuclear pore, which would account for the simultaneous inhibition of host and activation of viral mRNA transport, as this transport factor would be moved from the sites of host transcription and processing to the viral centres. However, this putative transport factor has not yet been found.

Besides the facilitated transport of viral mRNAs from the nucleus, they are also preferentially translated in the cytoplasm late after infection (reviewed in (Zhang and Schneider, 1993)), and there are various regulatory components that work together in this selective translation. One of these components involves the cellular protein kinase R (PKR), which when activated by double-stranded RNA accumulated within adenovirus-infected cells (Maran and Mathews, 1988; O'Malley *et al.*, 1986) subsequently phosphorylates eIF-2α, thus inactivating this initiation factor and blocking translation of host cell mRNA. The adenovirus coded VA RNAs inhibit activation of PKR, and as they are found together with viral mRNAs (Mathews, 1980) they might protect viral protein synthesis.

The inactivation of eIF-4F (a cellular initiation factor with helicase properties) by dephosphorylation late in infection also helps in the selective translation of viral mRNA,

as the five families of mRNAs encoded by the adenovirus major late transcription unit all possess a non-coding region known as the *tripartite leader sequence*, a nucleotide sequence that virtually lacks secondary structure (Dolph *et al.*, 1990), which therefore is postulated (Huang and Schneider, 1991) not to need helicase activity in order to be translated, unlike cellular mRNAs.

Finally, a viral protein, L4-100 kd, seems to activate selectively late protein synthesis (Hayes *et al.*, 1990). This protein can bind to mRNA (Adam and Dreyfuss, 1987), suggesting a function at the polysome in facilitating viral translation.

1.1.5.2.7. Virus Assembly and Release from the Cell

The final stage of the virus infection is enabled with the production of large quantities of structural proteins. Trimeric hexon capsomeres are rapidly formed from monomers immediately after their synthesis in the cytoplasm (Horwitz *et al.*, 1969), helped by L4-100 kd (the same protein which stimulates late viral translation), which acts as a scaffold in the trimerisation by an unknown mechanism. Penton capsomeres assemble somewhat more slowly in the cytoplasm (Horwitz *et al.*, 1969), the fiber and the penton base being assembled independently and joining in the end to form the complete penton capsomere (Horwitz *et al.*, 1969; Velicer and Ginsberg, 1970). Following their production, hexon and penton capsomeres accumulate in the nucleus, where virion assembly takes place.

The process of assembly seems to start with the formation of an empty capsid (Philipson, 1984; Sundquist *et al.*, 1973), with the subsequent entry of a viral DNA molecule. This DNA-capsid recognition event is mediated by the packaging sequence, a *cis*-acting DNA element located about 260 bp from the left end of the viral genome (Grable and Hearing, 1992; Hammarskjold and Winberg, 1980; Hearing *et al.*, 1987; Tibbetts, 1977). Presumably, one or more unknown proteins bind at that sequence and mediate the interaction between DNA and the capsid. The sequence is only effective when near to the end of the genome, suggesting that proteins which interact at the terminal origin of replication might have a role in this process. Capsid assembly does not seem to occur in the absence of interacting DNA, rather being probably initiated in association with viral DNA, mediated by the packaging sequence. The mechanism of DNA entry is controversial (D'Halluin *et al.*, 1980; Hasson *et al.*, 1992; Weber *et al.*, 1985): some suggest that the DNA enters the capsid as it is replicated, others suggest these events occur in separate nuclear compartments. L1-52/55 kd seems to facilitate the

process (Hasson *et al.*, 1992; Hasson *et al.*, 1989), probably also acting as a scaffold. The L3-coded protease (Mangel *et al.*, 1993; Tihanyi *et al.*, 1993; Webster *et al.*, 1993) functions late in the assembly process by cleaving virion capsid components, thus stabilising the virion particle and rendering it infectious.

Once the particle is assembled, it needs to exit the cell. Apparently, there are at least two viral mechanisms that facilitate the release and proliferation of progeny virus, both involving the disruption of intermediate filaments which are components of the cytoskeleton. After infection vimentin is cleaved very rapidly, apparently both by a viral gene expression independent process related with the adsorption process (Belin and Boulanger, 1987), and in response to protein E1B-19 kd (White and Cipriani, 1989; White and Cipriani, 1990), thus collapsing the extended vimentin system into the perinuclear region (Defer *et al.*, 1990; Zhai *et al.*, 1988). Also late in the infectious cycle, the viral protease cleaves the cellular cytokeratin K18 (Chen *et al.*, 1993), which prevents its polymerisation into filaments and causes it to accumulate in clumps, thus disturbing the mechanical integrity of the cell and rendering it more prone to lysis and subsequent release of the progeny virus.

1.2. The Adenovirus Protease

The role of proteases as biological regulators was neglected for many years. Although for many early researchers proteases seemed to be boring catalysts present only to annoy biochemists, in recent years it became obvious that proteolysis plays a vital role in many biological processes (reviewed in (Wolf, 1992)). In particular, it became evident that viral proteases play a vital role in the morphogenesis of virus particles (reviewed in (Hellen and Wimmer, 1992)), namely in two processes: virus assembly, and subsequent maturation, related to infectivity.

Several characteristics of the adenovirus protease make it an attractive target for antiviral therapy: its role in virion infectivity, its unusual specificity, and its mechanism of activation.

1.2.1. History

The first evidence of proteolytic processing of adenovirus proteins late in infection was reported in 1973 (Anderson *et al.*, 1973; Eron *et al.*, 1974; Ishibashi and Maizel, 1974a), but the suggestion that it might be virus-encoded came from temperature-sensitive (ts1) mutant studies (Bhatti and Weber, 1978; Weber, 1976), where at the non-

permissive temperature (39°C) some viral protein precursors were observed not to be cleaved.

Whilst initial attempts to characterise the protease using inhibition studies (Bhatti and Weber, 1979) suggested that it was a chymotrypsin-like, non-metallo, neutral serine protease, sequencing studies lead to the finding of the gene coding for the protease of adenovirus type 2 within the viral chromosome (Akusjärvi *et al.*, 1981; Hassel and Weber, 1978; Kruijer *et al.*, 1980; Yeh-Kai *et al.*, 1983) and its subsequent translation into amino acid sequence, revealing a molecular weight of 23 kDa, and this was confirmed by ts1 mutant studies (Weber and Houde, 1987).

An initial study claimed that the protease was specific for Gly-Ala bonds (Sung *et al.*, 1983; Tremblay *et al.*, 1983), although none of the peptides Gly-Ala, Ac-Gly-Ala or Ac-Gly-Gly-Ala-ONH_2 was able to be cleaved by the protease (Tremblay *et al.*, 1983), this being explained in terms of the need of a secondary structure for the enzyme to bind to the substrate. Six viral substrates are acknowledged to be processed by the protease (Tremblay *et al.*, 1983): pVI, pVII, pVIII, pTP, pIIIa and L2-11 kDa.

Based on the cleavage specificity suggested above, it was proposed (Chatterjee and Flint, 1987) that the protease is synthesised as a 23 kDa zymogen which by means of autocatalysis cleaves an Ala-Gly bond at amino acids 46 and 47 of the protease, thus maturing into a 19 kDa phosphorylated active form.

More detailed substrate specificity investigations (Anderson, 1990; Webster *et al.*, 1989a) finally agree to a consensus cleavage sequence of (I,L,M)XGG-X or (I,L,M)XGX-G, with the cleavage occurring at the G-X or X-G bonds, respectively. Furthermore, more inhibition studies (Webster *et al.*, 1989b) propose for the first time that the protease is a cysteine rather than a serine protease. A protease deletion mutant lacking the first 9 N-terminal amino acids was shown not to be active (Anderson, 1990), dismissing the proposal of autocatalytic activation. Also, immunoblot titrations showed (Anderson, 1990) that each virion should contain about 10 protease molecules.

The first protease recombinant (Houde and Weber, 1990) still seemed to support the possibility of autocatalytic activation, and the same report proposes a serine protease catalytic triad of His-54, Asp-102 and Ser-160, based on the alignment of the protease sequences known by then.

In 1993, consensus was reached on a cysteine protease classification (Tihanyi *et al.*, 1993; Webster *et al.*, 1993) with Cys-104 or Cys-122 as likely candidates for the active

site thiol. Also, the autocatalytic proposal for activation was definitely abandoned (Tihanyi *et al.*, 1993; Webster and Kemp, 1993), and instead a peptide was found (Mangel *et al.*, 1993; Webster *et al.*, 1993) to activate the protease, which was shown to be derived from the C-terminal of the viral protein pVI, with the sequence GVQSLKRRRCF. A thiol-disulphide exchange mechanism of activation was proposed (Webster *et al.*, 1993) in which an internal disulphide bond rearrangement might involve the conserved Cys-104, Cys-122 and Cys-126, although the protease does not seem to contain any internal disulphide bonds (Tihanyi *et al.*, 1993). Viral DNA (or any negatively charged polymer) also appears to contribute to the activation of the protease (Mangel *et al.*, 1993; Mangel *et al.*, 1996), although there is not complete agreement on this (Matthews and Russell, 1995; Webster *et al.*, 1994).

In adenovirus infected HeLa cells, non-viral substrates, cytokeratin-7 and cytokeratin-18, were found (Chen *et al.*, 1993; Zhang and Schneider, 1994) to be cleaved by the protease, which should contribute to the weakening of the cell mechanical integrity and thus promote host cell lysis as well as release of progeny virions.

As more protease sequences were published, it became apparent that the mechanism of activation involves one of the only two conserved cysteines, Cys-104 or Cys-122 (Grierson *et al.*, 1994; Rancourt *et al.*, 1994), whereas the remaining one should belong to the active site.

Further studies on the ts1 mutant revealed that the mutation at proline-137 to leucine produced a slight structural change which affected its solubility (Keyvani-Amineh *et al.*, 1995a), and that the mutation did not affect enzyme activity but rather its packaging into the virion particles (Rancourt *et al.*, 1995).

Investigations on the role of the protease in the virus entry into cells (Cotten and Weber, 1995; Greber *et al.*, 1996) indicated that it might be in an inactive oxidised form whilst in the oxidising environment outside the host cell, whereas it should become active when it enters the reducing environment either in the endosome or the cytosol. The degradation of protein VI during virus entry is thought to be due to the protease, which would seem to violate its substrate specificity under some circumstances (Greber *et al.*, 1996; Matthews and Russell, 1995) by cleaving sites such as PEGR-G (Greber *et al.*, 1996). Interestingly, the ovine strain 287 protease was also discovered not to agree with the consensus sequence (Vrati *et al.*, 1996), cleaving sequences such as MRAT-G and NTGW-G.

Human adenovirus type 2 protease expressed in insect cells was also shown to be active in the absence of activating peptide (Keyvani-Amineh *et al.*, 1995b), cleaving ovalbumin and baculovirus protease, whereas expression of protease mutants also in insect cells showed (Jones *et al.*, 1996) that cysteine-122 of the protease should be the active site thiol and cysteine-104 involved in the activation mechanism with the activating peptide, and these results were soon confirmed (Rancourt *et al.*, 1996).

More inhibition studies (Brown *et al.*, 1996; Sircar *et al.*, 1996) showed that cysteine inhibitors successfully depress the formation of infectious particles *in vivo* (Sircar *et al.*, 1996), and bovine pancreatic trypsin inhibitor was thought (Brown *et al.*, 1996) to inhibit by precipitating viral DNA, one of the putative cofactors of enzyme activity (Mangel *et al.*, 1993). Also these studies led to a new protease count of about 50 molecules per virion (Brown *et al.*, 1996). Some isoforms of the protease were described (Keyvani-Amineh *et al.*, 1996), which seem to be related to its oxidation state and influence its activity.

Finally, in the summer of 1996 the tridimensional structure of the protease·peptide complex was published (Ding *et al.*, 1996), confirming cysteine-122 to be the active site thiol and the activating peptide to be covalently bound to cysteine-104 in the protease.

1.2.2. Classification and Mechanism of Proteolysis

Proteases are classified according to the essential catalytic residues at their active sites (Bond and Butler, 1987). There are four distinct classes of proteases identified to this moment: serine (EC 3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23) and metallo-proteases (EC 3.4.24). The serine proteases present an unusually reactive serine side chain at the active site which attacks nucleophilically the carbonyl carbon of the peptide bond and involve an intermediate acylation of the serine. Cysteine proteases (formerly known as thiol proteases) work in a very similar fashion to that of serine proteases, but instead of a nucleophilic serine there is a cysteine at the active site, and the mechanism of cleavage is depicted in Figure 1.6 and Figure 1.7. Aspartic proteases (formerly known as acid proteases) contain two aspartic residues at the active site, and they are thought to work via acid-base catalysis (Bond and Butler, 1987). Metallo-proteases contain a metal ion (usually zinc) at their active site, which probably enhances the nucleophilicity of H_2O as well as polarising the peptide bond, prior to a nucleophilic attack (Bond and Butler, 1987).

The usual approach for the identification of the protease class to which a certain enzyme belongs is by assaying the protease in the presence of several inhibitors known to block with a reasonable specificity a certain type of active site (Barrett, 1977). However, most inhibitors for serine proteases also inhibit cysteine proteases, and the first inhibition studies that led to the classification of the adenovirus protease as a serine protease (Bhatti and Weber, 1979) failed to verify whether this was the situation. Subsequent studies corrected this assertion (Webster *et al.*, 1989b), and its classification as a cysteine protease is now generally accepted and verified in several independent studies (Brown *et al.*, 1996; Cotten and Weber, 1995; Greber *et al.*, 1996; McGrath *et al.*, 1996; Rancourt *et al.*, 1994; Sircar *et al.*, 1996; Tihanyi *et al.*, 1993; Weber and Tihanyi, 1994).

As mentioned above, the cleavage mechanism of a cysteine protease is based on a thiol nucleophilic attack on the carbonyl carbon of the peptide bond, followed by the acylation of the thiol group and release of the N-terminal free peptide chain:

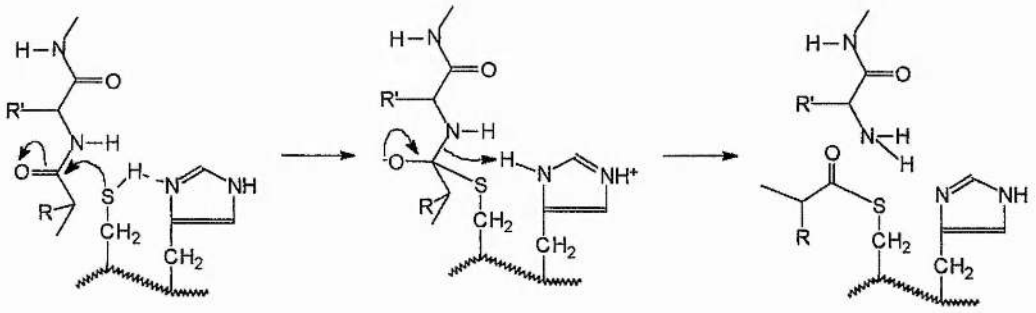


Figure 1.6 - Mechanism of cleavage by a cysteine protease. The first step consists of the active site thiol acylation, with the release of the free N-terminal chain.

The next step involves the hydrolysis of the thiol acyl:

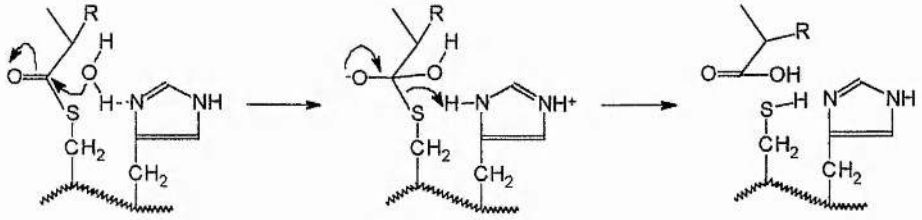


Figure 1.7 - Mechanism of cleavage by a cysteine protease. The second step (deacylation) consists of the hydrolysis of the thiol acyl, with the release of the free C-terminal chain.

From the mechanism described above, it can be seen that the active site is composed of a reactive cysteine side chain coupled with a histidine side chain which functions as a proton acceptor and donor. Therefore, another means of identifying whether a certain

protease is of the cysteine class is by comparing the sequences of the several known serotypes for that protease and searching for conserved cysteine residues which can be assigned to the active site. This is shown in Figure 1.8.

```
Virus                    1       10        20        30        40        50
Bovine   type    7 .....MSGLSEKEVFL LLSSLQCTHG FLGTFDCRFP GFINKVKVQT AIINTGPREQ
Ovine    strn  287 .....MSGTSESELKN LISSLHLNNG FLGIFDCRFP GFLQKSKIQT AIINTGPREQ
Avian    strn  127 .....MSGTSESELKA LMKSLGIAGN FLGTFDCTFP GFINKHKRQT AIINTGSRAS
Human    type    2 ......MGSSEQELKA IVKDLGCGPY FLGTYDKRFP GFVSPHKLAC AIVNTAGRET
Human    type    5 ......MGSSEQELKA IVKDLGCGPY FLGTYDKRFP GFVSPHKLAC AIVNTAGRET
Human    type   40 ......MGSSEQELVA IVRELGCGPY FLGTFDKRFP GFMAPHKLAC AIVNTAGRET
Human    type   41 ......MGSSEQELVA IARDLGCGSY FLGTFDKRFP GFMAPNKLAC AIVNTAGRET
Human    type   12 ......MGSSEQELTA IVRDLGCGPY FLGTFDKRFP GFVSRDRLSC AIVNTAGRET
Human    type    3 MTCGSGNGSSEQELKA IVRDLGCGPY FLGTFDKRFP GFMAPDKLAC AIVNTAGRET
Canine   type    1 ...MAEGGSSEEELRA IVRDLAVTPF FLGTFDKRFP GFISSQRITC AVVNTAGRET
Bovine   type    3 ......MGSREEELRF ILHDLGVGPY FLGTFDKHFP GFISKDRMSC AIVNTAGRET
Human    type    4 ....MAAGSGEQELRA IIRDLGCGPY FLGTFDKRFP GFMAPHKVAC AIVNTAGRET
Murine   type    1 ......MGSSETELRQ LVADLGIGS. FLGIFDKHFP GFISVNKPAC AIVNTASRET
Porcine  type    3 ......MGSTEDELRA MARDLQLPR. FLGTFDKSFP GFLQESQRCC AIVNTAARHT
Avian    type    1 .....MSGTTETQLRD LLSSMHLRHR FLGVFDKSFP GFLDPHVPAS AIVNTGSRAS

Virus               51      60        70        80        90       100       110
Bovine   type    7 GGIHWIALAW DPKSYQMFIF DPLGWKNDQL MKYYKFSYSN LIKRSALSS. PDKCVKVIKN
Ovine    strn  287 GGIHWITLAL EPISYKLFIF DPLGWKDTQL IKFYNFSLNS LIKRSALNN. SDRCITVERN
Avian    strn  127 GGLHWLAFAW DPLRYTIYMF DPLGWKEKDL FKLYGFSYKT MIKRSALQS. DNRCVKLVKN
Human    type    2 GGVHWMAFAW NPRSKTCYLF EPFGFSDQRL KQVYQFEYES LLRRSAIASS PDRCITLEKS
Human    type    5 GGVHWMAFAW NPHSKTCYLF EPFGFSDQRL KQVYQFEYES LLRRSAIASS PDRCITLEKS
Human    type   40 GGVHWLALAW NPKNRTCYLF DPFGFSDERL KQIYQFEYEG LLKRSALAST PDHCITLIKS
Human    type   41 GGVHWLALAW NPKSHTCYLF DPFGFSDERL KQIYQFEYEG LLKRSALAST PDHCITLVKS
Human    type   12 GGVHWLAFGW NPKSHTCYLF DPFGFSDQRL KQIYQFEYES LLRRSALAAT KDRCVTLEKS
Human    type    3 GGEHWLAFGW NPRYNTCYLF DPFGFSDERL KQIYQFEYEG LLRRSALA.T KDRCITLEKS
Canine   type    1 GGVHWLAMAW NPRSKTFYMF DPFGFSDSKL KQVYSFEYEG LLRRSAIAST PDRCVTLAKS
Bovine   type    3 GGVHWLAMAW HPASQTFYMF DPFGFSDQKL KQIYNFEYQG LLKRSALTST ADRCLTLIQS
Human    type    4 GGEHWLAFAW NPRSNTCYLF DPFGFSDQRL KQIYQFEYEG LLRRSALA.T KDRCVTW.KS
Murine   type    1 GGVHWLAMAW YPTSSTFYLF DPFGFSDRKL QQVYKFEYER LLKRSAVSSS SSKCVTLVKS
Porcine  type    3 GGRHWLAVAW EPASRTFYFF DPFGFSDREL AQVYDFEYQR LLRKSAIQST PDRCLTLVKS
Avian    type    1 GGMHWIGFAF DPAAGRCYMF DPFGWSDQKL WELYRVKYNA FMRRTGL.RQ PDRCFTLVRS

Virus               111     120       130       140       150       160       170
Bovine   type    7 SQSVQCTCAG SCGLFCVFFL YCFYKYKSNA FKNCLFQSLY GSIPS...LT PPNPTNLHKN
Ovine    strn  287 TQSVQCTCAG SCGLFCIFFL YCFHFYKQNV FKSWLFQKLN GSTPS...LI PCEPHLLHEN
Avian    strn  127 TEAVQCTCAG SCGLFCVFFL YCFNLCHINP FEASIFQAMH GTSPA...LY PSKPHLLAAN
Human    type    2 TQSVQGPNSA ACGLFCCMFL HAFANWPQTP MDHNPTMNLI TGVPNSMLNS PQVQPTLRRN
Human    type    5 TQSVQGPNSA ACGLFCCMFL HAFANWPQTP MDHNPTMNLI TGVPNSMLNS PQVQPTLRRN
Human    type   40 TQTVQGPFSA ACGLFCCMFL HAFVNWPTSP MERNPTMDLL TGVPNSMLQS PQVVPTLRHN
Human    type   41 TQTVQGPFSA ACGLFCCMFL HAFIHWPSNP MEQNPTMDLL TGVPNSMLQS PQVEPTLRRN
Human    type   12 TQTVQGPFSA ACGLFCCMFL HAFTHWPDHP MDKNPTMDLL TGVPNCMLQS PQVVGTLQRN
Human    type    3 TQSVQGPRSA ACGLFCCMFL HAFVHWPDRP MDGNPTMKLV TGVSNSMLQS PQVQPTLRRN
Canine   type    1 NETIQGPNSA ACGLFCCMFL HAFVNWPDNP FNHNPTMGPL KSVPNYKLYD PTVQHVLWEN
Bovine   type    3 TQSVQGPNSA ACGLFCCMFL HAFVRWPLRA MDNNPTMNLI HGVPNNMLES PSSQNVFLRN
Human    type    4 HQTCRVRV.G RCG.FSAACS TACA.WP.TP MDKNPTMNLL TGVPNGMLQS PQVEPTLRRN
Murine   type    1 HQTVQGPHSA ACGLFCVLFL AAFGKYPQNP MNNNPIMGPI EGVPNDQMFN PCYTKTLYRN
Porcine  type    3 TQSVQGPHSA ACGLFCLLFL AAFARYPDSP MAYNPVMDLV EGVDNERLFD ADVQPIFRAN
Avian    type    1 TEAVQCPCSA ACGLFSALFI VSFDRYRSKP MDGNPVIDTV VGVKHENMNS PPYRDILHRN

          Virus               171     180       190       200   204       214
          Bovine   type    7 QDFLYKFFKE KSLYFRQNEE YIVSNTKIGL IKSHI..... ....
          Ovine    strn  287 QTFLYDFLNA KSVYFRKNYR TFIENTKTGL IKTH..... ....
          Avian    strn  127 QQMLYDFLRS HSSYFVNNER TLVCNTKLNL INIHQ..... ....
          Human    type    2 QEQLYSFLER HSPYFRSHSA QIRSATSFCH LKNM..... ....
          Human    type    5 QEQLYSFLER HSPYFRSHSA QIRSATSFCH LKNM..... ....
          Human    type   40 QERLYRFLAQ RSPYFQRHCE RIKKATAFDQ MKNNM..... ....
          Human    type   41 QERLYRFLTQ HSPYFRRHRE RIEKATAFDQ MKNAQVLFHN KIFY
          Human    type   12 QNELYKFLNS LSPYFRHNRE RIEKATSFTK MQNGLK.... ....
          Human    type    3 QEVLYRFLNT HSSYFRSHRA RIERATAFDR MDMQ..... ....
          Canine   type    1 QEKLYKFLEK NSAYFRAHAA AIKTRTAFNK LKQ..... ....
          Bovine   type    3 QQNLYRFLRR HSPHFVKHAA QIEADTAFDK MLTN..... ....
          Human    type    4 QEALYRFLNS HSAYFRSHRA RIEKATAFDR MNQDM..... ....
          Murine   type    1 QQWVYSYLNK NSLYFRLHVE LIKKNTAFDK LLVRK..... ....
          Porcine  type    3 QEACYAFLAR HSAYFRAHRH AIMEQTHLHK ALDMQ..... ....
          Avian    type    1 QERTYYWWTK NSAYFRAHQE ELRRETALNA LPENHV..... ....
```

Figure 1.8 - Alignment of protease sequences from 15 adenovirus serotypes with increased gap length penalty and manual adjustments based on DNA sequence alignment. Alignments created using the Pileup program of the University of Wisconsin GCG package. The single conserved histidine and the two conserved cysteines are highlighted, as well as the glutamic/aspartic acid thought to be involved in the catalytic site. Sequences taken from N.C.B.I. ENTREZ Browser (http://www3.ncbi.nlm.nih.gov/Entrez).

From the sequence alignment, no serines were found to be fully conserved, whereas two cysteines, Cys-104 and Cys-122, are conserved in all of the known serotypes, as

well as the necessary His-54, again confirming the adenovirus to possess a cysteine protease.

1.2.3. Regulation of Activity: The Activating Peptide

First described in 1993 (Mangel *et al.*, 1993; Webster *et al.*, 1993), deriving from the C-terminal of the viral protein pVI, the peptide GVQSLKRRRCF was found to enhance the activity of the protease by what was assumed to be a thiol-disulphide interchange in which it would presumably rearrange some intramolecular disulphide bond within the protease, thus exposing the active site thiol (Webster *et al.*, 1993). This assumption was based on the finding (Webster *et al.*, 1993) that the oxidised (dimeric) form of the peptide was more able to activate the protease, together with the fact that the peptide GVQSLKRRRAF was unable to cause any observable activation. Another interesting result (Webster *et al.*, 1993) was that neither the peptide KRRRCF nor the unrelated peptides CKQDPFLRFGK or CQIRSATSFCHLKNM were able to activate the protease, suggesting a high specificity in the activation mechanism by the activating peptide. However, no further studies were made in order to elucidate the nature of this specificity, and this was the main objective of this project.

Other studies (Greber *et al.*, 1996; Jones *et al.*, 1996; Mangel *et al.*, 1996) only led to the conclusion that the peptide should bind to the protease, although no requirements for the specificity of the binding were ever determined.

1.2.3.1. Mechanism of Activation

In a similar approach to that of identifying what are the important residues in the protease, an alignment of the known activating peptide serotypes was produced in order to determine which are the conserved amino acids in most of the sequences. Although currently 14 serotypes are known (cf. Figure 1.9), at the start of this project only the human types 2, 5, 12, 40 and 41, together with the murine type 1, were published (Davidson *et al.*, 1993; Weber and Tihanyi, 1994).

From the sequences known by then, several characteristics were common to all serotypes: (a) all peptides had the same length; (b) cysteine-10 was always present at the same position; (c) the consensus sequence was of the form **GXXXX**KRRRCX, which seemed to suggest a role for the four positively charged residues to the left of the cysteine. The relevance of all of the above characteristics was assessed in this project.

```
Human   type      2 IVGL-GVQSLKRRRCF
Human   type      5 IVGL-GVQSLKRRRCF
Human   type     12 IVGL-GVKSLKRRRCY
Human   type     31 IVGL-GVKSLKRRRCY
Human   type     40 IVGL-GVKSLKRRRCY
Human   type     41 IVGL-GVKSLKRRRCY
Canine  type      1 IVGV-GLSNVKRRRCF
Canine  type      2 IVGV-GLSNVKRRRCF
Equine  type      1 IVGV-GLHGVKRRRCFY
Murine  type      1 IMGL-GLQPIKRRRCF
Ovine   strain  287 MTGD-GVNFNTRRYCY
Avian   strain  127 MVGD-GVRYGSQRYCY
Avian   type      1 LSGT-GVATATRRMCY
Avian   type     10 LSGT-GVNVSSRRLCY
```

Figure 1.9 - Alignment of pVI C-terminal sequences from 14 adenovirus serotypes. Amino acids in `blue` highlight the conserved residues, whereas the dash indicates the putative site of cleavage by the adenovirus protease. Sequences taken from N.C.B.I. ENTREZ Browser (http://www3.ncbi.nlm.nih.gov/Entrez). Equine type 1 and avian type 10 sequences were obtained from translation of DNA sequences upstream from the hexon gene (accession numbers L79955 and U26221, respectively), where pVI is usually located, and as such are subject to confirmation.

1.2.3.2. Activation in Other Proteases

Many enzymes are synthesised as inactive precursors (*zymogens*, or *proenzymes*), subsequently activated by some mechanism which usually allows control over the activity of the enzyme. Proteases are no exception, and perhaps the most well known examples are pepsin, trypsin and chymotrypsin, all synthesised as precursors which are then activated by proteolysis (Stryer, 1988). However, this activation process is irreversible, and the control over activity has then to be exerted by means of specific inhibitors.

Yet, there are other known mechanisms of protease activation or enhancement of activation, encompassing activation by ions, glycosaminoglycans, lipoproteins, pH change, etc., and these can usually be associated with several pathologies, which stresses the importance of proteases in the control of biological processes. For instance, the cytotoxicity of arsenic is thought (Yih and Lee, 1994) to be due to arsenite ions (AsO_3^{3-}) being able to induce the activity of a cathepsin-B like protease. Rat mast cell protease 1 seems to be active only when bound to heparin (Pejler and Maccarana, 1994), whereas apoliprotein(a) was reported (Pursiainen *et al.*, 1994) to be activated when bound to lipoprotein(a), and this could be a factor determining the atherogenic potential of lipoprotein(a). Cathepsin-D in brain cells was found (Kohnken *et al.*, 1995) to present pH-dependent reversible activation, suggesting that its intracellular localisation could play a role in Alzheimer's disease. Tetanus and botulinum-B neurotoxins appear to be

zinc proteases (Schiavo *et al.*, 1992) which when activated by reduction of the interchain disulphide bond can cleave synaptobrevin, an integral membrane protein of small synaptic vesicles.

Several protozoan cysteine proteases also seem to be activated by proteolytic cleavage (Eakin *et al.*, 1992; Nene *et al.*, 1990), but curiously the activity of at least one of them (Lonsdale-Eccles *et al.*, 1995) also appears to be enhanced by a kininogen-like moiety, to which usually is attributed cysteine protease inhibitor behaviour.

Also the solvent can be responsible for enhancement of activity: the addition of formamide to a chymotrypsin aqueous solution markedly enhances its proteolytic activity (Yamamoto and Kise, 1994), whereas tertiary amines have a similar effect in both subtilisin and chymotrypsin (Yamamoto and Kise, 1993).

Several virus proteases have also been reported to be activated either by a solvent change or by binding to another molecule, as well as by proteolytic maturation. Whilst both equine arthritis virus (Snijder *et al.*, 1992) and murine coronavirus (Baker *et al.*, 1993) code for cysteine like proteases which are matured by autocatalysis, the protease of herpes simplex virus type 1 interestingly presents an activity quite dependent on solvent changes (Hall and Darke, 1995), effected by kosmotropes (water structure-forming cosolvents) such as glycerol, sodium citrate or phosphate, which seem to induce a conformational change in the molecule.

The binding of other molecules to viral proteases has also been reported to activate them. These molecules can be of cellular origin, as in the case of poliovirus 2A protease being activated by the eukaryotic translation initiation factor 3 (Wyckoff *et al.*, 1990), or of viral origin, such as reported (Wiskerchen and Collett, 1991) for the p80 protease of bovine viral diarrhoea virus which needs the presence of 133-kDa polyprotein for development of activity, or Sindbis virus nsP4 protease which appears to require the protein nsP3 for efficient proteolysis (de Groot *et al.*, 1990).

Other examples in particular are known in the *Flaviviridae* virus family: yellow fever virus NS3 protease domain requires NS2B domain (Chambers *et al.*, 1991), dengue virus type 4 NS3 protease domain needs NS2B domain (Falgout *et al.*, 1993), and hepatitis C virus NS3 protease was shown to request binding to NS4A protein for efficient proteolytic activity (Butkiewicz *et al.*, 1996; Kim *et al.*, 1996; Shimizu *et al.*, 1996).

A final interesting example of activation is the human cytomegalovirus protease, a serine protease which was shown (Margosiak *et al.*, 1996; Qiu *et al.*, 1996; Shieh *et al.*,

1996; Tong *et al.*, 1996) to dimerise in order to achieve full proteolytic activity. Inhibition of activation has also been reported (Baum *et al.*, 1996a; Baum *et al.*, 1996b) to be achieved by preventing disulphide bond formation, using 1,1'-(dithio-di-*o*-phenylene)-*bis*-(5-phenylbiguanide) or flavins, which corroborates the dimerisation mechanism as a means of activation.

2. Materials and Methods

This chapter describes all the techniques and methods used in this work, with a theoretical introduction explaining the principles underlying each technique.

2.1. Production and Purification of Recombinant Proteases

The cloning of adenovirus proteases in *E. coli* provided with a more abundant source of material, as well as the ability of easily obtaining mutants for the study of the importance of selected residues on the activation/activity of the protease. It also allowed the preparation of reasonable amounts of proteases from other adenovirus serotypes using their DNA as a template without having to resort to viral cultures.

Most of the mutants used in this work had been already cloned, and in such instances the transformed cells were used for culture and expression of the protease in the same way as for the cloned proteases.

Several purifications were attempted, which consisted on various combinations of the methods cited herein.

2.1.1. Production of Recombinant Protease

The original viral DNA was amplified from the protease gene using polymerase chain reaction (PCR). For this purpose, primers were synthesised flanking the protease gene, with some point mutations in order to create restriction sites which would later enable the insertion of the amplified gene into a plasmid vector.

The proteases were cloned using a pET-11c (plasmid for expression by T7 RNA polymerase) system (Novagen), which utilises the bacteriophage T7 RNA polymerase promoter to direct high level expression of cloned genes in *E. coli* (Rosenberg *et al.*, 1987). This vector contains a β-lactamase gene (which confers resistance to ampicillin; Amp^r) as a selectable marker (cf. Figure 2.1).

As standard laboratory stocks of *E. coli* do not produce T7 RNA polymerase, all DNA manipulations can be performed in any $RecA^-$, K12 *E. coli* strain normally used for cloning (such as XL1-Blue, selected with tetracycline, Stratagene) without concern about background expression (Glover and Hames, 1995).

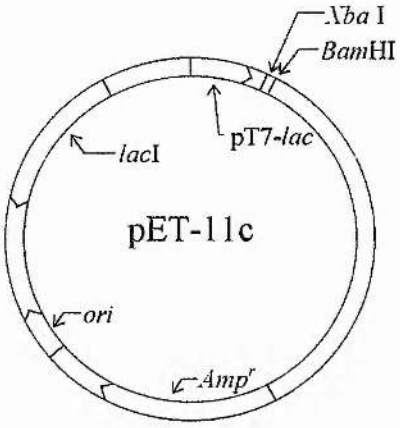


Figure 2.1 - Schematic representation of the structure of the pET-11c plasmid: *ori*, origin of replication; *Amp*^{r}, β-lactamase gene conferring ampicillin resistance; *lac*I, *lac*I gene; pT7-*lac*, T7-*lac* promoter; *Xba* I and *Bam*H I, restriction enzyme cleavage sites (Glover and Hames, 1995).

When protein expression is required, the plasmid is transformed into a host strain (BL21, F$^-$*ompT*r$_B^-$m$_B^-$) that contains a chromosomally integrated gene for the T7 RNA polymerase, which is present as a bacteriophage DE3 lysogen, with the above gene under control of the *lacUV5* promoter. The BL21 host has the advantage of lacking the outer membrane protease (*ompT*) as well as the *lon* protease (Glover and Hames, 1995), which may help in stabilising foreign proteins that are sensitive to endogenous protease activity. However, if the cloned protein is toxic the low level expression of the T7 RNA polymerase in the BL21(DE3) strain can be sufficient to prevent a stable transformation or propagation of the plasmid, and two solutions can used to circumvent this problem: the use of the T7-*lac* promoter in the pET vector, which will place the cloned protein under the control of *lac*, thus reducing its background expression in the absence of isopropylthiogalactopyranoside (IPTG); or alternatively, inhibiting T7 polymerase by expressing its natural inhibitor, T7 lysozyme, using a second plasmid that codes for this protein: pLysS, which produces small amounts of the above protein so that it does not inhibit induced levels of T7 polymerase. Furthermore, T7 lysozyme also cleaves a bond in the peptidoglycan layer of *E. coli*, thus allowing cells to be lysed under mild conditions such as freeze-thawing.

The above system can be switched on by adding IPTG to the culture, thus inducing the expression of T7 polymerase, which will in turn trigger expression of the cloned gene.

2.1.1.1. Agarose Gels

Agarose gels are used to separate DNA or RNA according to its molecular weight, much in the same way as SDS-PAGE (Method 2.2.1) separates proteins. However, unlike proteins DNA already has an overall negative charge, which allows straightforward electrophoresis of samples.

2 M Tris-acetate 0.05 M EDTA, pH 8, (TAE, 50×) electrophoresis buffer was prepared as a stock solution, which was diluted accordingly for use as chamber buffer.

Samples were prepared by adding 80 µl of sample to 20 µl of loading buffer (40% glycerol, 5×TAE, 50 mM EDTA, 5 units/ml RNAse (QIAGEN), 0.05% bromophenol blue). Agarose gels were prepared by making solutions of 1-2% agarose, 0.0005% ethidium bromide (from a 0.05% stock solution), dissolved by heating in microwave at full power for 2 min, and left to cool down to set. Samples were then loaded on the gel and run at a constant voltage of 100 V.

Separated bands were visualised using UV illumination of gel, and appropriate DNA was excised with the aid of a scalpel for further processing.

2.1.1.2. Polymerase Chain Reaction

This method is used to amplify from a given template a certain portion of DNA whose limits are defined by the oligonucleotide primers provided. There are three temperatures in each amplification cycle: the temperature required for DNA melting (denaturation from double strand to single strand), followed by cooling for oligonucleotide primers to anneal with their target sequences, and a final optimum temperature for the primers to be extended with DNA polymerase. This cycle is then allowed to repeat for several times, so that the products of an amplification cycle can be used as templates for the next, thus allowing an exponential increase of the desired product (Sambrook *et al.*, 1989).

Oligonucleotide primers were designed with the aid of the Macintosh application Amplify (Engels, 1992), and synthesised using a Beckman Oligo 1000 DNA Synthesizer, at a concentration of 30 µM.

Viral DNA (0.5-2 µg) was washed with 1 ml ethanol together with 0.5 ml of chloroform and 40 µl of 3M sodium acetate, left for 5 min and centrifuged at 13000 rpm for 5 min (MSE Micro Centaur). The supernatant was discarded, 1 ml of 70% ethanol was added, and the mixture was centrifuged again at 13000 rpm for 1 min. The

supernatant was removed and the DNA pellet was left to dry, and resuspended in 250 µl HPLC filtered (0.2 µm) water.

A PCR mixture was prepared with 2 µl of the above DNA solution, 2 µl of each primer, 10 µl of dNTP solution (0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.5 mM dTTP), 2 µl of 100 mM $MgSO_4$ (BioLabs), 10 µl of 10× ThermoPol buffer (0.1 M KCl, 0.1 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.8, 20 mM $MgSO_4$, 1% Triton X-100, BioLabs), 1 µl of Vent DNA polymerase (in 0.1 M KCl, 0.1 mM EDTA, 10 mM Tris-HCl pH 7.4, 1 mM DTT, 0.1% Triton X-100, 50% glycerol, BioLabs) and 71 µl of HPLC filtered (0.2 µm) water. 100 µl of liquid paraffin (Boots) were overlaid on the solution to prevent evaporation.

This mixture was submitted to 35 cycles of 94°C for 1.5 min, 55°C for 1.5 min and 72°C for 2 min, and a final 7 min at 72°C (Techno Programmable Dri-Block PHC-1). The mixture was then run on an agarose gel (Method 2.1.1.1) to extract the amplified segment.

2.1.1.3. DNA Gel Extraction Protocol

This procedure is used to extract DNA from agarose gels (Method 2.1.1.1) and purify it for further use. It is essentially the same as described in (QIAGEN, 1995b).

To the excised gel band, 3 volumes (µl) of buffer QX1 (QIAGEN) were added per volume of gel band (assuming a density of 1 mg/µl), and this was incubated at 50°C for 10 min. One gel volume of isopropanol was then added and mixed, and this was centrifuged through a QIAquick spin column (QIAGEN) for 1 min at 13000 rpm (MSE Micro Centaur). The flow-through was discarded and the column was centrifuged with a further 0.5 ml of buffer QX1 for 1 min at 13000 rpm. The column was then washed by spinning with 0.75 ml of buffer PE (QIAGEN) for 1 min at 13000 rpm, the flow-through was discarded and the column centrifuged for an additional 1 min at 13000 rpm to remove any residual wash buffer.

The DNA was then eluted by spinning the column with 30 µl of HPLC filtered (0.2 µm) water at 13000 rpm for 1 min.

2.1.1.4. Restriction of Plasmid and Insert

This procedure aims at cutting the ends of the insert and the plasmid so that the ends of both match in one direction only, to ensure proper translation of the insert. For this

purpose, different restriction enzymes are used for each end of the insert, as well as in the plasmid.

The insert was cut by mixing 25 µl of the DNA solution obtained by Method 2.1.1.3 to 1 µl of each of the two restriction enzymes Xba I and BamH I (Promega) and 3 µl of Multicore buffer (Promega), and incubating at 37°C for 2 h.

The plasmid was cut by mixing 10 µl of the plasmid pET-11c solution (50 µg/ml, Novagen) to 1.5 µl of each of the two restriction enzymes Xba I and BamH I (Promega), 1.5 µl of Multicore buffer (Promega) and 0.5 µl of HPLC filtered (0.2 µm) water, and incubating at 37°C for 2 h.

The restricted insert and plasmid were run on an agarose gel (Method 2.1.1.1) and purified using Method 2.1.1.3.

2.1.1.5. Ligation of Insert

The ligation of the insert uses an enzyme, ligase, to join the ends of the insert obtained by PCR and plasmid (Method 2.1.1.2), whose ends were cut by restriction enzymes (Method 2.1.1.4).

To 9 µl of restricted insert, 1 µl of restricted plasmid (pET-11c, Novagen) and 7 µl of filtered (0.2 µm) HPLC water (Rathburn) were added and incubated at 45°C for 5 min, followed by incubation on ice for another 5 min. Finally 2 µl of T4 DNA ligase buffer (Promega) and 1 µl of T4 DNA ligase (Promega) were added and left incubating at room temperature for 4 h.

As a control for religation of plasmid, a mixture of 1 µl of plasmid and 16 µl of filtered (0.2 µm) HPLC water (Rathburn) was used and subjected to the same treatment described above.

2.1.1.6. Transformation of Cells

Transformation of cells consists of introducing a foreign plasmid inside the cell by permeabilising temporarily the cell wall to the passage of DNA.

A single colony of bacteria from a stock agar plate or glycerol was inoculated in 1 ml of Luria-Bertani medium (LB: 1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, pH 7.0 (Sambrook *et al.*, 1989)) with 50 µg/ml carbenicillin (in the case of BL21(DE3) cells no antibiotic was used) and left growing overnight (~15 h) at 37°C. This was used to inoculate 50 ml of LB with antibiotic (see above) which was grown at 37°C until the culture reached an optical density (OD) of 0.4-0.6 (600 nm). The cells

were then centrifuged at 4°C, 3000 rpm for 10 min (IEC Centra-3R) and the supernatant was discarded.

The cell pellet was resuspended in 10 ml of filter sterilised (0.2 μm), ice cold 0.1 M $MgCl_2$, and centrifuged at 4°C, 3000 rpm for 10 min. Again the supernatant was discarded and the cells resuspended in 4 ml of filter sterilised (0.2 μm), ice cold 0.1 M $CaCl_2$, then left on ice for 30 min. The cells were then competent for 4-5 days, at their optimum after 24 h.

To 200 μl of competent cells 10 μl (containing at most 300 ng of DNA) of the ligation mixture (Method 2.1.1.5) was added and incubated on ice for 30 min, followed by incubation at 42°C for 90 s, and again on ice for 5 min. Aliquots of 30, 70 and 100 μl of cells were then plated on agar plates (Method 2.1.1.7) containing antibiotic (see above).

2.1.1.7. Agar Plates

LB-Agar medium was prepared by adding 15 g of bacto-agar per litre of Luria-Bertani medium (cf. Method 2.1.1.6) and autoclaving, followed by the addition of 50 μg/ml carbenicillin when necessary (cf. Method 2.1.1.6). The agar was then poured in sterile plastic Petri dishes and left to set.

Cells were then spread on the plate at the required volume, left growing overnight (~15 h) at 37°C and then stored at 4°C, where they can be kept for about two weeks as a stock.

2.1.1.8. Mini-Prep

This protocol is used to check whether a certain bacterial strain has a plasmid successfully transformed into it. It is essentially based on (QIAGEN, 1993), with modifications.

Transformed cells were plated in agar plates with 11.8 μg/ml tetracycline (in the case of XL1-Blue) as well as 50 μg/ml carbenicillin and left growing at 37°C overnight (~15 h). Single colonies were picked and inoculated in 5 ml LB (cf. Method 2.1.1.6), 50 μg/ml carbenicillin, 11.8 μg/ml tetracycline (in the case of XL1-Blue) and left growing at 37°C overnight (~15 h). They were then centrifuged at 3000 rpm for 5 min (IEC Centra-3R), the supernatant discarded and the cells resuspended in 0.3 ml of buffer P1 (QIAGEN). Then, 0.3 ml of buffer P2 (QIAGEN) was added and the mixture incubated at room temperature for 5 min, followed by addition of 0.3 ml buffer P3

(QIAGEN) and incubation on ice for 15 min, shaking occasionally. The precipitate was then centrifuged for 15 min at 13000 rpm (MSE Micro Centaur) and the precipitate discarded.

To the volume of supernatant 0.8 volumes of isopropanol were added and this was centrifuged again for 15 min at 13000 rpm, and the supernatant carefully removed. The pellet was washed with 0.5 ml 70% ethanol, centrifuged again for 15 min at 13000 rpm, the supernatant discarded and the DNA resuspended in 10 µl HPLC filtered (0.2 µm) water.

2.1.1.9. Maxi-Prep

This protocol is used for obtaining up to 500 µg of plasmid DNA, which can be further used for cell transformation. It is essentially as described (QIAGEN, 1995a).

Transformed cells were plated on agar plates with 11.8 µg/ml tetracycline (in the case of XL1-Blue) and 50 µg/ml carbenicillin and left growing overnight (~15 h) at 37°C. Single colonies were picked and inoculated in 5 ml LB (cf. Method 2.1.1.6), 50 µg/ml carbenicillin, 11.8 µg/ml tetracycline (in the case of XL1-Blue) and left growing at 37°C overnight (~15 h). They were then used (2 ml) to inoculate 500 ml LB, 50 µg/ml carbenicillin, 11.8 µg/ml tetracycline (for XL1-Blue), which was left growing at 37°C overnight (~15 h).

Cells were then centrifuged at 4500 rpm for 20 min (Beckman J6-HC), the supernatant discarded and the pellet resuspended in 10 ml of buffer P1 (QIAGEN). To this mixture, 10 ml of buffer P2 (QIAGEN) were added, mixed gently, and left incubating at room temperature for 5 min. Then, 10 ml of ice cold buffer P3 (QIAGEN) was mixed gently and this was left to incubate for 20 min on ice. The precipitate was centrifuged at 4°C, 12000 rpm for 30 min (Beckman J2-21), the precipitate discarded and the supernatant again centrifuged at 4°C, 12000 rpm for 15 min, and any further precipitate again discarded.

A QIAGEN-tip 100 (QIAGEN) column was the equilibrated with 10 ml of buffer QBT (QIAGEN), allowing the column to empty by gravity flow. The supernatant previously obtained was then applied to the column and it was allowed to enter the column by gravity flow. The column was then washed twice with 30 ml of buffer QC (QIAGEN) and the DNA eluted with 15 ml of buffer QF (QIAGEN). To the eluted DNA 10.5 ml of isopropanol were added and the precipitated DNA was centrifuged at 4°C, 10000 rpm for 30 min, with the supernatant being carefully discarded, and the DNA

pellet was washed with 5 ml of 70% ethanol and centrifuged at 4°C, 10000 rpm for 15 min. The 70% ethanol was discarded and the precipitated DNA was redissolved in 250 µl of HPLC filtered (0.2 µm) water.

2.1.1.10. Determination of DNA Concentration and Purity

This procedure enables an approximate determination of the concentration of several kinds of DNA, based on their absorbance at 260 nm. The following table is used for the above purpose:

Table 2.1 - Conversion factors between absorbance at 260 nm and DNA concentration (Sambrook *et al.*, 1989).

Type of DNA	Conc. (µg/ml)
Double stranded (ds) DNA	$A_{260\ nm} \times 50$
Single Stranded (ss) DNA	$A_{260\ nm} \times 40$
Oligonucleotides	$A_{260\ nm} \times 20$

2.1.1.11. DNA Sequencing

There are two main techniques for the sequencing of DNA: the method of chemical cleavage (Maxam and Gilbert, 1977) and the controlled interruption of replication (Sanger *et al.*, 1977).

The first method is based on the radioactive labelling of DNA on one end of its chain followed by limited scission of the chain, using conditions specific for each nucleotide or pair of nucleotides. Thus, dimethylsulphate can damage adenine (A) and guanine (G) by methylation, and these bases are then readily removed by heating at neutral pH, leaving the free sugar. The backbone can then be cleaved by heating in the presence of alkali, with concomitant elimination of the sugar. However, as methylation of G occurs at a much faster speed than that of A, the above reaction yields almost only breakage at G nucleotides. But if the methylated purine is treated with dilute acid both kinds of purine are cleaved. In the case of cytosine (C) and thymine (T), these pyrimidines are removed by hydrazine, followed by cleavage of the backbone by piperidine. Yet, in the presence of 2 M NaCl, the removal of T is suppressed. Therefore, if four lanes of a polyacrylamide gel are loaded with each of these reaction products (G, A+G, C, C+T), fragments can be separated depending on the place of cleavage, ordered according to the number of bases they contain, which will enable the deduction of the nucleotide sequence on the original DNA.

The principle underlying the second technique is the generation of DNA fragments by means of controlled interruption of enzymatic replication. This is achieved by means of extending oligonucleotide primers that complement the DNA to be sequenced by using DNA polymerase I, in much the same way as for PCR. However, for each different nucleotide lane there is in the incubation mixture a 2',3'-dideoxy analogue of one of the nucleotides, whose incorporation blocks further extension of the new chain, due to the lack of the 3'-hydroxyl terminus needed for the formation of the phosphodiester bond. Hence, fragments of various lengths will be generated which can be separated electrophoretically as in the previous method. The labelling in this method, however, is done by using radioactively labelled deoxyribonucleoside triphosphates, or alternatively using a primer tagged with a fluorescent compound (Smith *et al.*, 1986).

The sequencing was done automatically, based on the latter method, at the University of St. Andrews DNA Sequencing Facility using a Perkin Elmer ABI PRISM DNA Sequencer, using a solution of 12 µl containing 500 ng of the DNA to be sequenced and 3-5 pmol of the appropriate primer, which was amplified using a Perkin Elmer GeneAmp PCR System 2400, with a temperature program of 20 s at 96°C, 25 cycles of 10 s at 96°C, 5 s at 50°C and 4 min at 60°C, and a final temperature of 4°C.

2.1.1.12. Glycerol Stocks

Glycerol stocks are used to keep bacterial strains stored indefinitely (Sambrook *et al.*, 1989).

These are made by adding 1 volume of autoclaved 30% glycerol to 1 volume of bacterial culture in polypropylene containers, stirred to mix, and frozen at -70°C until required.

2.1.1.13. Cell Culture, Induction and Harvesting

Two methods were used for growing cells: the first used conical flasks and the second used a fermenter (Microferm Fermentor, New Brunswick Scientific). The advantage of using conical flasks is that the volume of culture can be chosen, but it has the disadvantage of a poorer reproducibility of experimental conditions and poorer control of variables such as agitation and aeration.

Both methods started with an inoculation of 5 ml LB medium (cf. Method 2.1.1.6), 50 µg/ml carbenicillin, with either a single colony of transformed *E. coli* from a stock plate or a glycerol stock. This inoculum was left to grow for 6-8 h at 37°C, and 1 ml of

this was used to inoculate 2×250 ml M9 (18.7 mM NH_4Cl, 22.0 mM KH_2PO_4, 42.3 mM Na_2HPO_4, 0.4% glucose, 1 mM $MgSO_4$, 50 µg/ml ampicillin/carbenicillin) in 2×2 l (for the conical flasks method) or 2×500 ml M9 in 2×1 l conical flasks (for the fermenter inoculum), which was left growing with constant shaking overnight (~15 h) at 37°C.

For the conical flasks method, the absorbance (600 nm) was checked, and if the value exceeded 0.6 the culture was diluted 1:3-5 in M9 medium and left to grow until the absorbance reached the desired value. When the absorbance (600 nm) was between 0.4-0.6 the culture was induced with 1 ml 0.25 M IPTG per 250 ml culture and left growing at 37°C for a further 5 h with constant shaking. The cells were then centrifuged at 10000 rpm (Beckman J2-21), the supernatant discarded, the cell pellet resuspended in 100 ml suspension buffer (50 mM Tris-HCl, 5 mM EDTA, 4% glycerol, pH 8), centrifuged again as before, the supernatant discarded and the cell pellet resuspended in 6 ml of suspension buffer, stored in polypropylene tubes at -20°C until required.

For the fermenter method, the 1 l inoculum was used to inoculate 10 l of M9, the temperature set to 37°C, the drive speed set to 300 rpm, the aeration set to 12 p.s.i., and the cells were left to grow for ~3 h, until the absorbance (600 nm) reached ~0.6. Then, 1.1 g/ml IPTG were added to the culture, the temperature was set to 25°C and the cells were left inducing for ~5 h. The culture was then centrifuged at 4500 rpm for 20 min (Beckman J6-HC), the supernatant discarded and the cell pellet resuspended in 30 ml of suspension buffer, stored in polypropylene tubes at -20°C until required.

2.1.1.14.Analysis of Protein Expression by SDS-PAGE

This procedure aims at assessing the increase of protein expression after induction, by taking samples of the induced culture at given time points and separating them by SDS-PAGE (Method 2.2.1.1). Typically, the induced protein can be identified as the band whose intensity increases noticeably with time.

Samples (1 ml) were centrifuged at 13000 rpm for 1 min and the supernatant discarded. The pellet was resuspended in 100 µl PBS (0.14 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4 (Sambrook *et al.*, 1989)) and 20 µl of β-MeOH were added. This was sonicated at an amplitude of 6.5 microns (MSE Ultrasonicator), boiled for 2 min and loaded onto the gel, which was run as per Method 2.2.1.1.

2.1.2. Purification of Protease

After expression, the protease is lodged inside the bacterial cells and it must be extracted and purified in order to be used. This process involves disruption of the cells, digestion of all the bacterial DNA and separation of the protease from the host proteins. The disruption of the cells is achieved by means of freeze-thawing together with the action of lysozyme, which degrades the peptidoglycan layer of *E. coli*. The step of DNA degradation is accomplished by the use of DNAse, an enzyme that degrades all DNA to oligonucleotides, thus facilitating its separation. A preliminary separation of the protease from the host proteins is achieved by centrifugation, as the protease remains in the supernatant. After the protease extraction follows its purification, by Fast Protein Liquid Chromatography (FPLC), using anionic and cationic exchange columns.

FPLC is a fast way of purifying proteins, as the mobile phase is driven through the stationary phase with the aid of pressure rather than by means of gravity flow. In ion exchange chromatography, analytes are separated according to their charge, and in the case of proteins this is pH dependent (one protein will have net positive charge below its isoelectric point and negative above it) (Harris and Angal, 1989). Therefore, a positively charged protein will bind to a cationic (negative) column but not to an anionic (positive), and different proteins will bind with different affinities to anionic or cationic columns. Once bound, a protein can be eluted by applying a high salt concentration to the column, which will exchange the charged protein by an ion, thus regenerating the column (Harris and Angal, 1989).

2.1.2.1. Extraction of Protease

From the two methods used to produce the protease, via conical flask culture or fermenter (cf. Method 2.1.1.13), different size pellets are obtained. However, the extraction method is essentially the same, the only variation being due to the amounts of reagents used. Therefore, the following description applies to ~2 ml pellets originated from the conical flasks method. The differences regarding the fermenter pellets are indicated in (parenthesis).

To each 2 ml thawed cell pellet, 0.67 ml suspension buffer (cf. Method 2.1.1.13), 6 mM β-MeOH, were added, making it up to 4 ml with suspension buffer (to each ~35 ml pellet, 5 µl β-MeOH were added). To this mixture, 50 µl of 10 mg/ml lysozyme were added and incubated at 37°C for 15 min, followed by 3 steps of freeze-thawing

(liquid nitrogen and 37°C). Then, 50 µl (200 µl) of 1 M $MgSO_4$ and 50 µl (100 µl) of 2 mg/ml DNAse I (Boehringer Manheim) were added and the mixture incubated at 37°C for 15 min.

The suspension was then centrifuged for 10 min at 13000 rpm (15000 rpm) on a MSE Micro Centaur minicentrifuge (Beckman J2-21 centrifuge) and the precipitate discarded. The resulting solution was then further purified by FPLC.

In brief, the FPLC system used consisted of two pumps (P-500, Pharmacia) allowing a maximum pressure of 5 MPa connected to a gradient programmer (GP-250, Pharmacia) which enables the use of isocratic, stepwise and continuous gradients for the elution of proteins. Detection was performed at 280 nm using a UV detector (2238 Uvicord SII, LKB Bromma).

2.1.2.2. FPLC with DEAE column

Diethylaminoethyl (DEAE) is a weak anion exchange group for ion-exchange chromatography. The strength of an ion-exchange group relates to the pH range in which the group remains ionised (Harris and Angal, 1989).

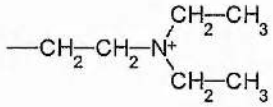


Figure 2.2 - The structure of DEAE, a weak anionic exchange group.

This group will bind any protein whose net charge is negative at the operating pH. It was observed that the protease does not bind to this group at a pH of 8, whereas quite a few bacterial proteins do bind. Furthermore, the particular elution time of the protease enabled it to be separated from most eluted proteins, as the protease migrates quite fast through DEAE.

The buffer used was either 25 or 50 mM Tris-HCl, pH 8.0, and depending on the sample volume one of two columns was used: a 30 × 1 cm column for the smaller samples and a 40 × 1.6 cm column for larger samples. In both the packing was DEAE-Sepharose Fast Flow (Sigma). Typical flow rates were 1-2 ml/min, depending on the purification procedure.

2.1.2.3. *FPLC with Heparin/CM column*

Heparin and carboxymethyl (CM) are both cationic exchange groups, the first being a strong ion exchange group and the latter a weak (cf. Method 2.1.2.2).

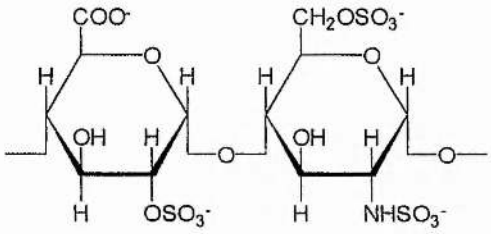


Figure 2.3 - The structure of heparin, a strong cationic exchange group.

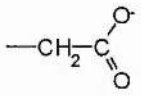


Figure 2.4 - The structure of CM, a weak cationic exchange group.

Once the protease has been separated from most proteins, and as it does not bind to positively charged groups, it should bind to negative groups. This was found to be the case. Therefore, running the DEAE purified protease through either or both these cationic exchange groups should enable a final separation of the protease from any remaining contaminants.

The buffer used was 25 mM Tris-HCl, pH 8.0, and two combinations of columns were tried: a mixed 10 mm internal diameter column with 2 cm heparin (Sigma) packed above 6.5 cm of CM-Sepharose (Sigma), and a 5×1 cm heparin column followed by a 11×1 cm CM-Sepharose column. Typical flow rates were 1-2 ml/min, depending on the purification procedure. The protease was typically eluted at an isocratic gradient of 10% 1 M NaCl, 25 mM Tris-HCl, pH 8.0.

2.2. *Determination of Protein Concentration*

Protein concentration was determined by comparison to solutions of known concentration of Soya bean trypsin inhibitor (SBTI), either by SDS-PAGE or using the Bradford reagent.

2.2.1. SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is the name given to a technique by which proteins are separated based on their molecular weight (Shapiro *et al.*, 1967), improved by Laemmli (Laemmli, 1970).

This consists of denaturating the proteins by the action of an anionic detergent such as sodium dodecyl sulphate (SDS), which also binds to the main chain of proteins,

conferring a large net negative charge to the protein, mercaptoethanol (β-MeOH) being also added to reduce any disulphide bonds. The SDS complexed protein has then a charge roughly proportional to its molecular weight, and it can be separated on an electric field.

In order to separate the proteins, a support must be provided, and this is accomplished by polyacrylamide gels, which can be synthesised by polymerisation of acrylamide using methylenebisacrylamide as a cross-linking reagent, in the presence of persulphate (Stryer, 1988) and N,N,N',N'-tetramethylethylenediamine (TEMED), a photochemical catalyst (Pombeiro, 1983) which stabilises the formation of sulphate radicals (Sambrook *et al.*, 1989):

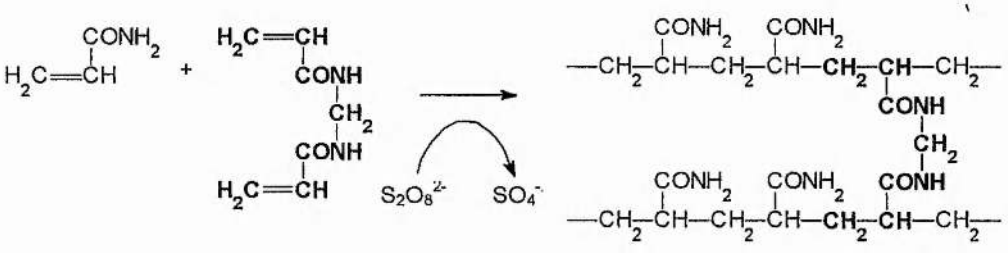


Figure 2.5 - Reaction of acrylamide with methylenebisacrylamide forming polyacrylamide, caused by the presence of persulphate.

The polymerisation reaction is inhibited by atmospheric oxygen. By varying the concentrations of acrylamide and methylenebisacrylamide, various pore sizes can be achieved that can be used to separate different ranges of molecular weights. Usually, a less concentrated mixture is used to prepare the stacking gel, whose role is to "order" the proteins, so that they can be separated more easily in the more concentrated separating gel that lies contiguously below the stacking gel (Pombeiro, 1983).

The mobility of most proteins under these conditions is linearly proportional to the logarithm of their mass. After being separated on the gel, they can be stained by either Coomassie blue (Chrambach *et al.*, 1967; de St. Groth *et al.*, 1963) or silver staining (Oakley *et al.*, 1980; Switzer *et al.*, 1979).

Whereas Coomassie blue, an anionic dye, stains by attaching itself to the NH_3^+ groups of proteins in slightly acidic media, van der Waals forces keeping the reactants together (de St. Groth *et al.*, 1963), the mechanism of staining with silver appears to be more complex, and possibly not fully understood. Also, there are several protocols for silver staining (see, for example, (Nielsen and Brown, 1984; Shevchenko *et al.*, 1996; Wray *et al.*, 1981) and references therein) which can be classified into three basic

categories (Merril, 1990): diamine or ammoniacal silver stains, non diamine chemical development silver stains, and photoreduction silver stains. The essence of these methods seems to be the ability of silver to complex with charged groups in the proteins with its subsequent reduction: carboxyl groups complexing with silver ions (Wray *et al.*, 1981), metallic silver precipitated by reduction with formaldehyde (Dion and Pomenti, 1983; Kurosaki *et al.*, 1984) and/or silver ions complexing with any charged group (Dion and Pomenti, 1983; Nielsen and Brown, 1984) are but a few of the published possibilities.

In the method used, after the fixation of proteins to the gel, a sensitising agent (such as $Na_2S_2O_4$) is used which will precipitate microgranules of silver sulphide (thus creating latent images of bands (Shevchenko *et al.*, 1996)), followed by addition of silver nitrate and formaldehyde. Development is done in an alkaline medium (Na_2CO_3), so that formaldehyde can reduce ionic silver to metallic silver (Merril, 1990), together with $Na_2S_2O_4$. The development is stopped by acidifying with acetic acid.

2.2.1.1. SDS-PAGE Mini Gels

Mini gels were prepared using a Bio-Rad Mini Gel kit. Separating gel solutions had a final concentration of 2.1 M acrylamide (15% (w/v), from a 30% stock solution), 5.6 mM methylenebisacrylamide (0.87% (w/v), from a 2% stock solution), 0.373 M Tris-HCl (from a 1 M, pH 8.7 stock solution) and 3.5 mM SDS (0.1% (w/v), from a 10% stock solution), 4.4 mM TEMED and 1.5 mM $(NH_4)_2S_2O_8$ (from a 0.44 M (10% (w/v)) stock solution freshly prepared), which was topped with water saturated butanol until the gel set.

Stacking gel solutions had a final concentration of 0.72 M acrylamide (5.1% (w/v), from a 30% stock solution), 9.1 mM methylenebisacrylamide (0.14% (w/v), from a 2% stock solution), 0.125 M Tris-HCl (from a 1 M, pH 6.9 stock solution) and 3.5 mM SDS (0.1% (w/v), from a 10% stock solution), 17 mM TEMED and 2.2 mM $(NH_4)_2S_2O_8$ (from a 0.44 M (10% (w/v)) stock solution freshly prepared).

Samples (15-20 µl) were added to 5 µl of Magic Mix (0.35 M SDS, 0.25 M Tris, 25% (v/v) glycerol, 0.05% bromophenol blue) and boiled for 2-5 min. They were then loaded onto the gel, and a constant current of 40 mA was applied per gel, which ran for ~40 min.

2.2.1.2. Coomassie Blue Staining

Gels were stained with ~20 ml of a solution of 2.5 g of Coomassie Brilliant Blue R250 in 45% (v/v) methanol, 10% (v/v) acetic acid, 45% (v/v) water, for 10-15 min, and destained several times with 10-30 min ~20 ml washes of an aqueous solution of 25% (v/v) methanol and 7.5% (v/v) acetic acid.

2.2.1.3. Silver Staining

Gels stained with Coomassie Blue (Method 2.2.1.2) were washed 3 times for 20 min in 30% (v/v) ethanol, then washed twice for 15 min with water. They were then soaked for 1 min in 1.4 mM $Na_2S_2O_4$, washed twice for 1 min in water, shaken for 30 min in a solution 12 mM $AgNO_3$, 1 mM HCHO, followed by a 1 min wash with water. It was then washed with a solution 0.57 M Na_2CO_3, 6 mM HCHO and 20 μM $Na_2S_2O_4$ until bands developed.

It was then washed with 3.5% acetic acid for 10 min for the developing to stop and washed 4 times for 30 min with water.

2.2.2. Bradford Reagent

This method is based in the binding of Coomassie blue to proteins (Bradford, 1976), as described in the Coomassie blue staining (Method 2.2.1). The binding of the dye to protein causes a shift in the absorption maximum of the dye from 465 to 595 nm, and it is the increase of absorbance at the latter wavelength that is monitored (Bradford, 1976).

2.2.2.1. Procedure for Bradford Reagent

Several dilutions of protein standards (SBTI) were prepared for a standard curve. To 0.8 ml of standards or sample 0.2 ml of Dye Reagent Concentrate (Bio-Rad) were added and vortexed. After 5 min the absorbance at 595 nm was read against a reagent blank. The absorbance was plotted against concentration of standards and sample concentration was read from the standard curve.

2.3. Western Blotting

Western Blotting consists of transferring electrophoretically proteins from a SDS-PAGE gel (Method 2.2.1) onto to a polyvinylidene difluoride (PVDF) membrane in such a way that a faithful replica of the original gel pattern is obtained (Burnette, 1981; Towbin *et al.*, 1979). These bound proteins can then be marked with a specific antibody

against which a second antibody is directed. The second antibody can then be either radioactively labelled or conjugated to fluorescein or peroxidase. The specific protein can finally be detected by either autoradiography or the peroxidase reaction product (Leong and Fox, 1990), respectively.

Peroxidase reaction was the method used, in which luminol, reacting with a peroxide in the presence of an oxidising agent such as peroxidase in a basic media, is oxidised with the concomitant production of light (Roswell and White, 1978):

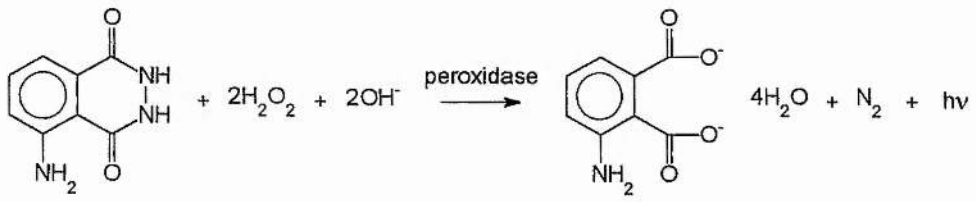


Figure 2.6 - Oxidation of luminol with the production of light.

If the blot is to be reacted again with the antibodies, it has to be stripped of the previous antibodies first.

2.3.1. Western Blotting Procedure

A blotting sandwich was assembled as follows: lower (positive) electrode, four layers of filter paper soaked in transfer buffer (48 mM Tris, 39 mM glycine, 20% (v/v) methanol, 0.13 mM SDS), polyvinylidene difluoride (PVDF) or nitro-cellulose membrane, SDS-PAGE gel (Method 2.2.1), four layers of filter paper soaked in transfer buffer and upper electrode (negative). This was run at a constant current of 50 mA for 1 hour.

The resulting blotted membrane was then washed for 10 min with 20 ml blocking buffer (500 ml of PBS (0.14 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4 (Sambrook *et al.*, 1989)), 25 g of Boots dried milk powder and 500 µl of polyoxyethylenesorbitan monolaurate (Tween 20), microwaved at full power for 2 min), followed by incubation with the first antibody (5 µl in 20 ml blocking buffer) for 30 min at room temperature.

The membrane was then rinsed with 10 ml blocking buffer, washed for 10 min with 20 ml of blocking buffer, rinsed again with 10 ml of blocking buffer, again washed for 10 min with 20 ml of blocking buffer and finally rinsed once again with 10 ml of blocking buffer.

The membrane was then incubated with the second antibody (Amersham peroxidase conjugated anti-rabbit IgG (4 µl in 20 ml blocking buffer) for 30 min at room temperature, washed again as above and rinsed with PBS.

The blot was then washed for 1 min in a mixture of Amersham reagents 1 (2.5 mM luminol (Fluka) from a 250 mM stock solution in dimethyl sulphoxide (DMSO, Sigma), 0.396 mM p-coumaric acid from a 90 mM stock solution in DMSO, and 0.1 M Tris-HCl from a 1 M stock solution of Tris-HCl pH 8.5) and 2 (0.0192% (v/v) H_2O_2 and 0.1 M Tris-HCl from a 1 M stock solution of Tris-HCl pH 8.5), drained and placed between two acetate sheets, and X-ray film (Fuji Medical X-Ray Film RX) was exposed to the blot for 1-5 min, and developed with a Kodak M35 X-OMAT Processor.

2.3.2. Stripping of Western Blot from Antibodies

PVDF membranes with antibodies previously bound were stripped by incubating the membrane at 70°C for 30 min in 0.10 M β-MeOH, 69 mM SDS, 62.5 mM Tris-HCl (from a 1 M, pH 6.9 stock solution), followed by washing with PBS.

2.4.- Peptide Synthesis and Purification

The concept of solid phase peptide synthesis was first proposed by Merrifield (Merrifield, 1963), marking a radical change on the field of peptide synthesis. The elimination of excess reagents and by-products, as well as the solubility and purification of peptides, became much more simplified, as all unreacted products were easily washed away, and the growing peptide chain did not have to be soluble. The method was quickly improved (Merrifield, 1964a; Merrifield, 1964b) by changing the amino protecting groups to enable milder reaction conditions, and the resulting method was widely adopted to synthesise peptides successfully. The adoption of the Fmoc chemistry ensued to try and resolve some of the weaknesses of Merrifield's method, mainly by introducing milder reacting conditions and by enabling a monitoring of the process (Atherton and Sheppard, 1989). This technique is continuously being improved in several ways, through prediction of difficult couplings (Van Woerkom and Van Nispen, 1991), optimisation of cleavage cocktails (Solé and Barany, 1992), various methods for monitoring coupling (Flegel and Sheppard, 1990; Salisbury *et al.*, 1990; Young *et al.*, 1990) and use of more specific catalysts (Carpino, 1993).

A curious future alternative for synthesising peptides could be the use of proteases (Schellenberger and Jakubke, 1991) by adjusting appropriate physicochemical reaction parameters, thus exploiting the intrinsic reversibility of hydrolytic reactions.

2.4.1. Peptide Synthesis

All peptides were synthesised using the chemical method of fluorenyl-methoxycarbonyl-polyamide (Atherton and Sheppard, 1989), using a Cambridge Research Biochemicals' Pepsynthesiser II. The above is a two step method that consists of a N-terminal elongation of a peptide linked to a polyamide resin by addition of C-terminal activated, side chain protected amino acids, followed by the N-terminal deprotection of the newly attached amino acid. The N-terminal protection, whose purpose is to prevent cross-reactions, consists of a fluorenylmethoxycarbonyl (Fmoc) group (Atherton *et al.*, 1981). The activation at its C-terminal is achieved by means of a pentafluorophenyl ester (Opfp) group, which will react with the N-terminal of the peptide linked to the resin, catalysed by 1-hydroxybenzotriazole (HOBt) (Carpino, 1993):

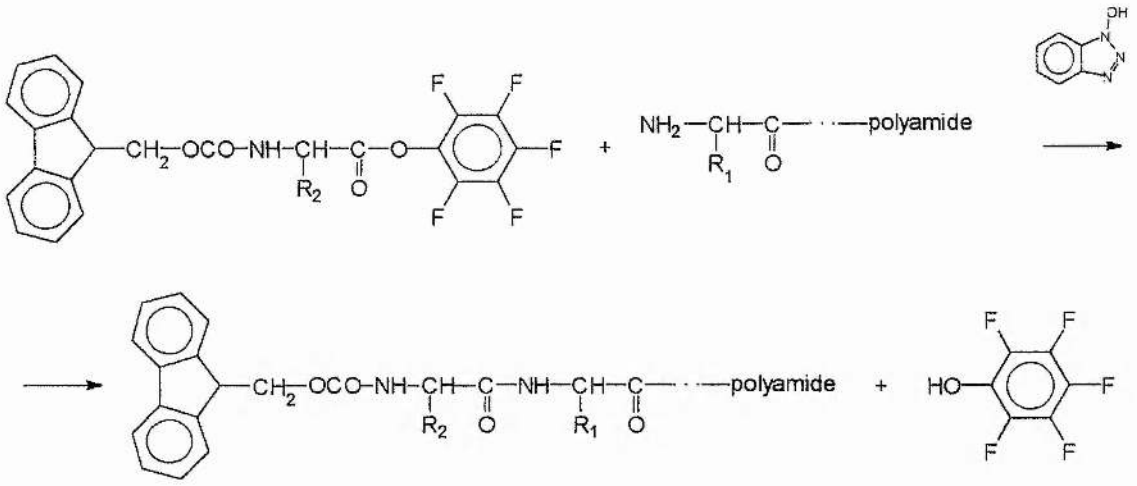


Figure 2.7 - Reaction of Fmoc protected amino acid, Opfp activated, with polyamide coupled peptide, catalysed by HOBt.

The polyamide resin holds the growing peptide chain, enabling the excess reagents to be washed from the reaction medium, which in turn allows the deprotection step to be performed, in which piperidine removes the Fmoc group from the N-terminal of the peptide chain:

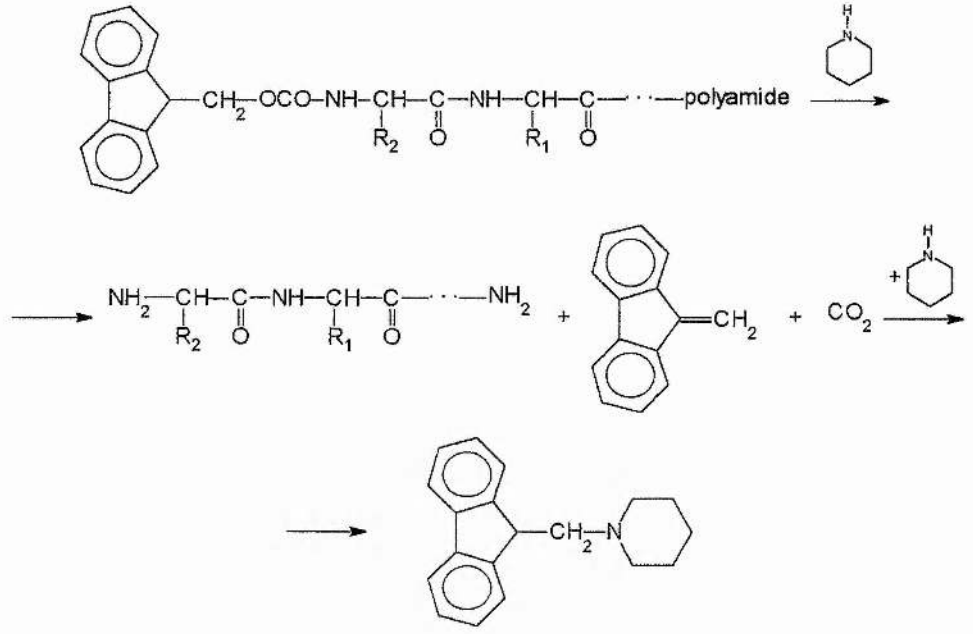


Figure 2.8 -Deprotection reaction of Fmoc protected peptide chain. The released dibenzofulvene-piperidine adduct absorbs at 330 nm, enabling its detection by spectrophotometry.

In certain peptides, the deprotected N-terminal is required to be acetylated, in order not to contain any free amino groups. That can be achieved by reacting acetic anhydride

with the peptide chain in the presence of pyridine (Van Woerkom and Van Nispen, 1991):

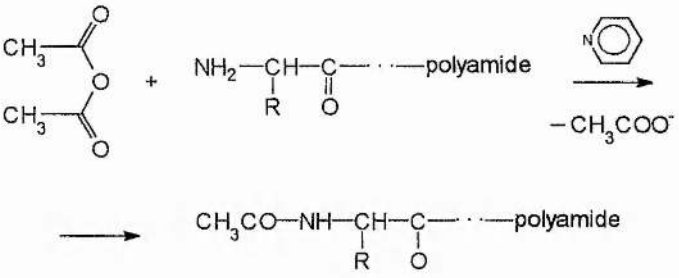


Figure 2.9 - Acetylation reaction of the N-terminal of a peptide chain with acetic anhydride

Dimethylformamide (DMF) is the solvent commonly used in solid phase peptide synthesis as it is a good solvating agent for both the resin and the Fmoc peptides, and it keeps both the resin and the growing peptide in an extended, mobile and unhindered conformation (Atherton and Sheppard, 1989). However, in order to minimise undesirable side reactions, this reagent must be reasonably pure, namely free of amines like dimethylamine, which can cause a significant loss of Fmoc protecting groups, leading to a double insertion of amino acids. In order to detect for the presence of this contaminant, Sanger's reagent (1-fluoro-2,4-dinitrobenzene, FDNB) (Sanger, 1945) is used (Figure 2.10), and its removal is achieved by using molecular sieves to trap it.

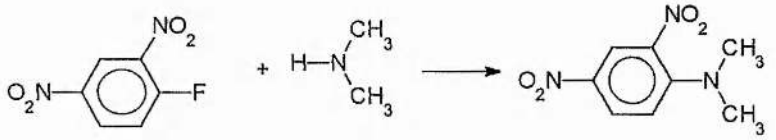


Figure 2.10 - Reaction of FDNB with dimethylamine. The product absorbs at 381 nm, enabling its detection spectrophotometrically.

In order to monitor the progress of the coupling reaction, the method of Counterion Distribution Monitoring (Salisbury *et al.*, 1990; Young *et al.*, 1990) was used. This method continuously monitors the total population of amino functions on the support resin using the distribution of a very small quantity of a reporter dye between these groups and cations in solution. As the medium is acidic, due to the presence of the catalyst HOBt, the amino groups bound to the resin are protonated, as well as a tertiary amine like diisopropylethylamine, and an inert anionic dye such as quinoline yellow will distribute itself among all available cations. However, as the coupling proceeds the number of free amines in the support resin decreases and the number of dye anions

bound as counterions is displaced into solution, which will increase the absorption of the solution. This will reach a theoretical maximum when the reaction is complete.

Therefore, both the coupling reaction and the deprotection step can be monitored spectrophotometrically, the first by measuring the increase in absorbance due to quinoline yellow (measured at 436 nm, (Salisbury *et al.*, 1990)) and the latter by measuring the absorbance of the released dibenzofulvene-piperidine adduct (measured at 330 nm, cf. Figure 2.8 (Atherton *et al.*, 1988)). Figure 2.11 shows a typical trace of the absorbance change for a complete cycle of coupling and deprotection.

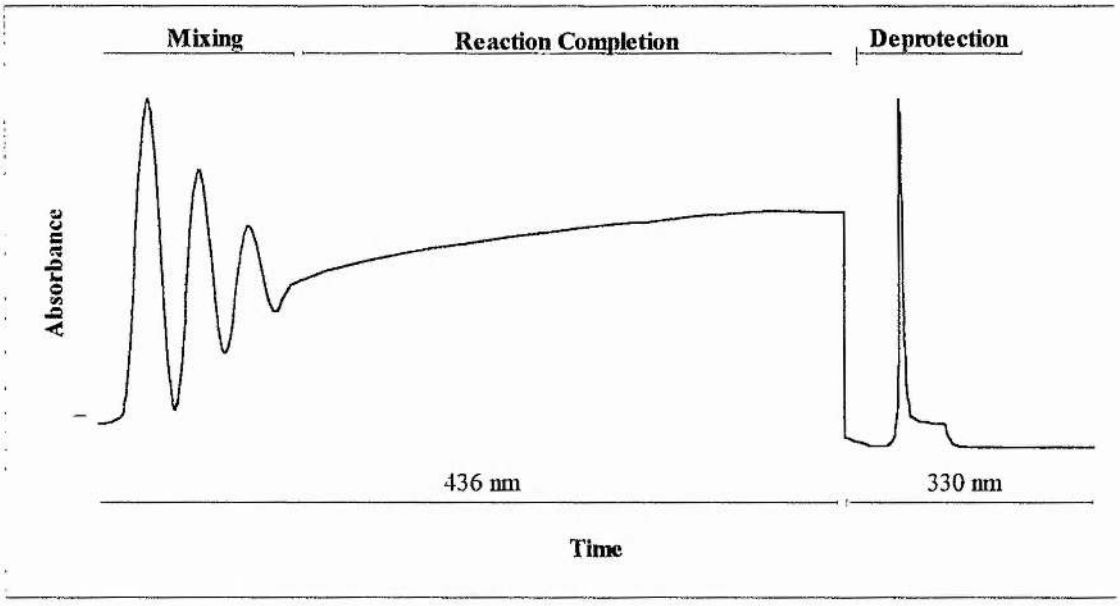


Figure 2.11 - Scheme of a typical trace of coupling and deprotection reactions, obtained spectrophotometrically.

Once the peptide chain is synthesised, the peptide must be cleaved from the resin, and the side chain protecting groups must be removed. All side chain protecting groups were designed to be cleaved by the same conditions, so that they can all be removed at the same time:

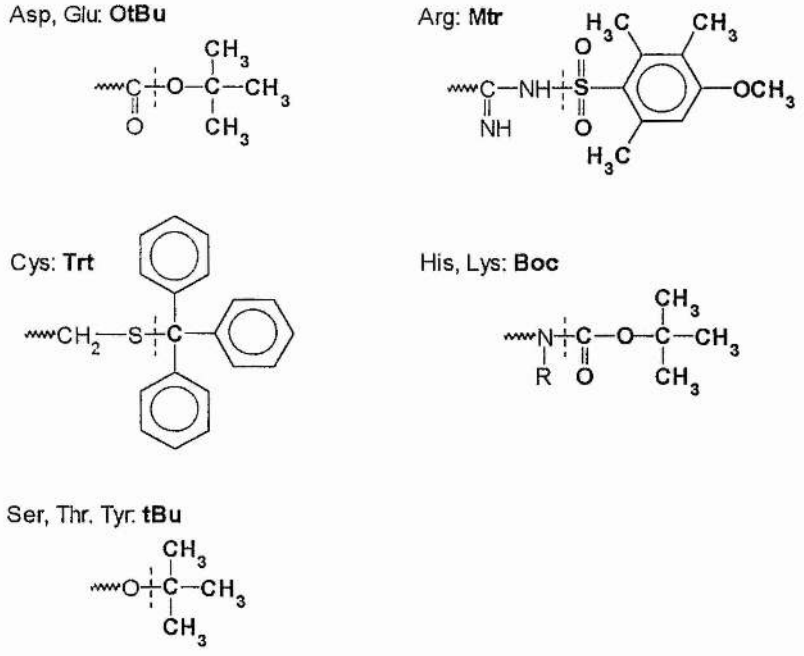


Figure 2.12 - Side chain protecting groups of amino acids used in peptide synthesis. The dashed line indicates the cleavage site.

Both the cleavage of the peptide from the resin and the removal of the side chain protecting groups are done at the same time by acidic cleavage, using trifluoroacetic acid (TFA) (Atherton *et al.*, 1988), after washing the resin with a mildly acidic reagent such as acetic acid, as any residual DMF has a pronounced inhibitory effect on acidolysis (Atherton and Sheppard, 1989). The advised solvent washing sequence is t-amyl alcohol, acetic acid, t-amyl alcohol and diethyl ether, the intermediary t-amyl alcohol being necessary to avoid the exothermic mixing of DMF with acetic acid (Arshady *et al.*, 1981). The final ether is used in order to easily dry the resin.

In the process of cleavage, the protecting groups released might react again with the free side chain groups. In particular, methionine, cysteine, tryptophan and possibly tyrosine appear to be most at risk of reacting with the released carbonium ions (Atherton and Sheppard, 1989). Therefore, in the presence of these groups scavengers should be added to the reacting medium, such as anisole (a good scavenger for tBu trifluoroacetate) and ethanedithiol (the most efficient scavenger for tBu trifluoroacetate, it also scavenges Mtr groups, minimising alkylation of Trp by Arg(Mtr)) (Applied Biosystems, 1990). Whereas all the protecting groups based on an alkoxybenzyl linkage are cleaved by acids relatively fast (trifluoroacetic acid, room temperature, 0.5-2 h (Atherton and Sheppard, 1989)), as well as the linkage of the peptide chain to the resin,

the Mtr protecting group of arginine (cf. Figure 2.12) requires longer cleavage times, as well as particular scavengers such as phenol in order to react with the released arylsulphonyl cleavage product (Atherton and Sheppard, 1989). Three or four arginines will typically require overnight reaction times, in which it is advisable to monitor the cleavage by removing samples to check whether all the Mtr groups are completely removed.

Once the cleavage is deemed to be finished, the peptide should be separated from the resin by means of filtration and the reaction mixture should be evaporated at reduced pressure, in order to wash and precipitate the peptide using ethyl ether. Several washes with ethyl ether are recommended so that all residual acid is removed. A final resuspension of the dried peptide in water followed by freeze-drying of the peptide should provide with a reasonably clean peptide.

2.4.1.1. Reagent Preparation

DMF from Rathburn was treated with 50 g of molecular sieves from Sigma per 2.5 l for 10-14 days. The sieves were activated by heating in a furnace at 200°C for 2 h. All DMF was filtered using a 0.2 µm pore nylon filter. It was then tested for the presence of amines by using 0.1% (w/v) FDNB in ethanol mixed in equal volumes with DMF, leaving at room temperature for 30 min, and then reading absorbance at 381 nm. All DMF with a reading above 0.15 was rejected, and treated further with sieves until a satisfactory test was obtained.

All syntheses used PEG-PS resins from Millipore with the first amino acid already coupled. Resin loadings ranged from 0.10-0.21 meq/g, and of these, 0.5-1 g was used depending on the amount of peptide required. These were suspended in DMF and inserted in Omnifit columns.

All Fmoc amino acid Opfp esters weights were calculated so as to be in a 4-fold excess of the resin loading, and were dissolved in 1 ml DMF, 1 M HOBt, 0.8 mM diisopropylamine and 0.1 mM quinoline yellow (quinoline yellow solution), which was kept as a stock at 4°C, and made fresh every second day. All reagents were from Aldrich.

2.4.1.2. Semi-Automated Peptide Synthesis

The Cambridge Research Biochemicals' Pepsynthesiser II is a machine controlled by computer (Apple IIe), in which both the steps of coupling and deprotection are done automatically, the new amino acid addition being the only manual step. It was connected

to a spectrophotometer which enabled the monitorisation of the coupling and deprotection procedures, as described above (cf. Figure 2.11). The program used to control of reagents flow was the following:

Step	Action	Time (min)	Description
1	FLOW A	5	Wash with DMF
2	FLOW B	15	Deprotection with 20% piperidine in DMF
3	FLOW A	10	Wash with DMF
4	CALL	0	Wait for operator
5	FLOW S	5	Add new amino acid in quinoline yellow solution
6	RECIRC	25	Coupling reaction (left on pause for at least 1 h)
7	FLOW S	1	Wash loading line
8	FLOW A	5	Wash with DMF
9	RECIRC	0.5	Clear lines and pump of excess reagents
10	FLOW A	5	Wash with DMF
11	REPEAT		Repeat the cycle

Every synthesis started on step 1, with each new amino acid being added at step 5 and left reacting for at least 1 h.

If acetylation was required, 0.53 M acetic anhydride, 12 mM pyridine in 1 ml of DMF were added at step 5 and left reacting for at least 1 h.

After the last amino acid was added and coupled, the program was left running until step 4, so that the last coupled amino acid was deprotected. The peptide coupled resin was then washed sequentially on a sintered filter with: ~25 ml DMF, ~10 ml t-amyl alcohol, ~10 ml glacial acetic acid, ~10 ml t-amyl alcohol and finally twice with ~25 ml of diethyl ether, and dried by using reduced pressure to pull all the washes through the filter.

2.4.1.3. Peptide Cleavage from Resin and Removal of Side Chain Protecting Groups

For each gram of resin, 40 ml of TFA from Rathburn, 1 ml of ethanedithiol and 1 ml of anisole both from Aldrich (0.66 ml of ethanedithiol, 0.66 ml of anisole and 0.66 ml of phenol in the case of a peptide containing Arg(Mtr)) were mixed gently together in a round bottom flask at room temperature.

In order to evaluate the progress of the cleavage and removal of the side chain protecting groups, 100 µl samples were taken from the reaction mixture and put into a 0.5 ml eppendorf, and the mixture was evaporated with the help of a gentle stream of nitrogen. The remaining residue was then dissolved in 100 µl of HPLC water and extracted four times with 200 µl of ether, the latter being discarded. The resulting sample was then analysed using Method 2.7.1. If only one peak (or one major peak) was observed, then it meant that the reaction had terminated.

After the completion of the reaction, the mixture was filtered to separate from the resin and washed twice with ~10 ml of TFA. The filtrate and washings were evaporated to approximately 1 ml of solution in a rotary evaporator at reduced pressure, with a water bath at a temperature of ~45°C. To the remaining solution ~20 ml of ether were added in order to precipitate the peptide, which was left to settle, and the supernatant ether was carefully discarded. This procedure was repeated about five times, to try and wash the peptide as thoroughly as possible from any remaining impurities and solvents. After the last extraction, the ether was completely dried off by gently stirring the flask in the water bath, and the resulting peptide powder was dissolved in 1-5 ml of water and freeze-dried overnight. This resulted in a reasonably clean peptide, which was then further purified.

2.4.1.4. Mini Method Peptide Synthesis

This method was designed to produce smaller amounts of peptide without recurring to the use of the Pepsynthesiser II. Basically, the chemistry used was the same as for the previous methods (Methods 2.4.1.1 to 2.4.1.3) as well as the reagents, the only difference lying on the amount of reagents used and the support where the peptide was synthesised. This consisted of a polyethylene box where some perforations were made in order to fit Mobicols (Bioscience Services) flow-through columns in which about 75 mg of resin could be introduced, with an extra hole to which a vacuum line could be fitted. The amino acid weights used were 30-40 mg per coupling.

Each coupling step consisted of:

1. Washing 3 times the resin with DMF, sucking the column dry after each wash.
2. Filling the column with 20% piperidine in DMF and sucking dry, with another filling of the same reagent being allowed to percolate through the column.
3. After sucking dry the column, washing 4 times with DMF, sucking the column dry after each wash.

4. Adding the amino acid dissolved in 300 µl DMF and 100 µl of a 1 M solution of HOBt in DMF to the resin, mixing it by using a plastic Pasteur pipette, allowing the peptide to react for at least 45 min at 35°C, and restarting a new step.

After all the amino acids were added, the last was deprotected by performing the above steps 1-3, and the resin was then washed twice with 0.5 ml of the following sequence of reagents: DMF, t-amyl alcohol, glacial acetic acid, t-amyl alcohol. It was then washed 4 times with 0.5 ml of ether and dried under vacuum.

The cleavage mixtures were the same as per Method 2.4.1.3, but a total amount of 200 µl was used, and the reaction times were also the same. The resin was then washed with 200 µl TFA, the washings being collected into an eppendorf tube and this evaporated with a stream of nitrogen. To the residual solution ~500 µl of ether were added to precipitate the peptide, which was then washed 5 times with ether, allowing the peptide to settle down and discarding the supernatant between each wash. After the final wash, the ether was vaporised using a stream of nitrogen and the peptide was dissolved in water and freeze-dried.

2.4.2. Peptide HPLC Purification

Preparative reverse-phase High-Performance (or High Pressure, as it is also known) Liquid Chromatography (HPLC) is a technique currently used to purify peptides (Atherton and Sheppard, 1989). The designation of reverse-phase is due to the stationary phase (the solid packing), being hydrophobic, presenting a lower polarity than the mobile phase, which is the reverse of the normal situation of a stationary phase with a higher polarity than the mobile phase (Pombeiro, 1983).

The eluent is usually a solvent of high polarity such as water in which low polarity solutes such as peptides tend to bind to the low polarity stationary phase, being eluted when increasing amounts of a lower polarity solvent such as acetonitrile are flowed through the column. The stationary phase usually consists of silica or polymer beads.

The high pressure has the advantage of enabling the separation of solutes with high resolution and in a relatively short time, in comparison to traditional chromatographic techniques (Pombeiro, 1983).

2.4.2.1. Reverse-phase HPLC

The HPLC system used was from Gilson, with two pumps model 303 delivering a pressure of up to ~400 bar, with a manometric module model 802 to keep the working

pressure range between 0-250 bar and a Holochrome UV detector selected for 226 nm. in order to detect for peptide absorbance. Pump A delivered a solution of 0.1% TFA in filtered water and pump B delivered 0.1% TFA in filtered acetonitrile.

The columns used were a Bio-Rad RSL C18 HL, with a length of 250 mm and an internal diameter of 4.6 mm, with a particle size of 5μm, for samples of a concentration of up to 10 mg/ml, and a Capital HPLC Ltd Spherisorb ODS2-Sb5-20061 with a length of 250 mm and an internal diameter of 10 mm, with a particle size of 5μm, for more concentrated samples.

Peptides were dissolved in 0.1% TFA, filtered (0.2 μm) and applied to the column equilibrated with 0.1% TFA in water. The maximum sample load was 2 ml. The gradient profile used to purify the peptides was as follows:

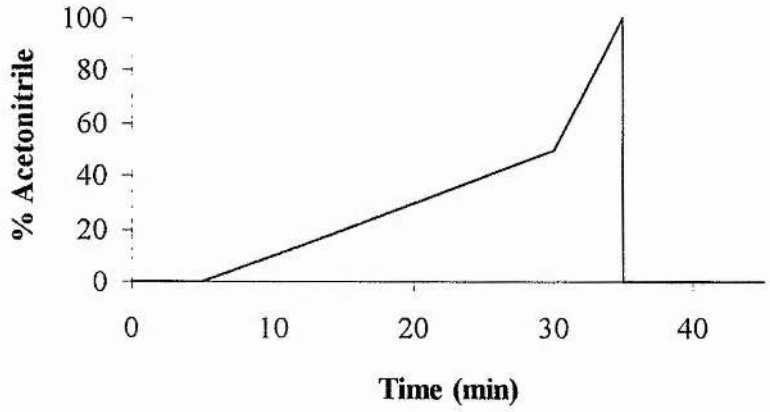


Figure 2.13 - Gradient profile of reverse-phase HPLC purification.

The flow rate was 1 ml/min for the Bio-Rad column and 4 ml/min for the Capital HPLC column. The peptides were eluted usually at 30-40% acetonitrile. The fractions collected were then analysed for purity using Method 2.7.1, and the appropriate fractions pooled and freeze-dried. The final peptides were considered to be virtually pure (only one peak shown on capillary electrophoresis analysis, Method 2.7.1).

2.5. Peptide and Protein Sequencing

The method commonly used for peptide and protein sequencing is based on the Edman degradation procedure (reviewed in (Han *et al.*, 1985)). This consists of three steps: the coupling of phenylisothiocyanate (PITC), which can be activated with methyl-piperidine, to the N-terminal of the peptide or protein (Figure 2.14), followed by the

removal of the first amino acid by acidic cleavage (Figure 2.15), and a final step of conversion of an unstable thiazolone to a more stable phenylthiohydantoin (PTH, Figure 2.16).

The first step should be performed under mild alkaline conditions (pH in the range 9-9.5), in order to avoid side reactions such as the hydrolysis of PITC with the consequent formation of aniline, which could react with excess PITC and form diphenylthiourea (DPTU), a compound that can interfere in the posterior identification of PTH-amino acids (Han *et al.*, 1985).

The buffer solution should also be aldehyde-free, as side reactions between aldehydes and α-amines tend to form Schiff bases and related products, which can reduce the yield of the expected phenylthiocarbamyl-peptide.

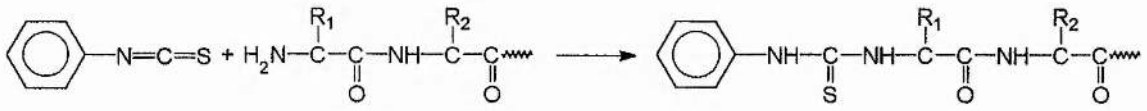


Figure 2.14 - Edman degradation: the coupling of PITC with the peptide chain.

After the coupling reaction, the excess PITC, solvents and by-products should be removed, to follow with the cleavage step.

The acidic cleavage is usually performed using anhydrous trifluoroacetic acid, and care should be taken to avoid the acid catalysed hydrolysis of the peptide bonds by ensuring the absence of water in the solvent.

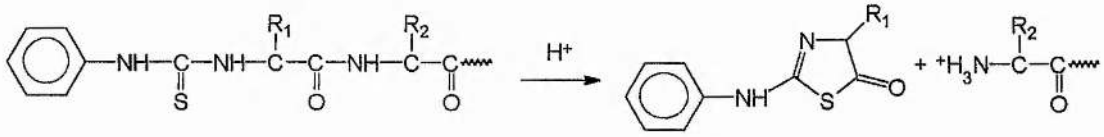


Figure 2.15 - Edman degradation: the cleavage of the phenylthiocarbamyl peptide bond via cyclisation in acidic media.

The final step ensures that the unstable anilino-thiazolinone derivatives of the amino acids are converted into more stable PTH-amino acids, and this is usually achieved by heating at 80°C for 10 min in 1 M HCl.

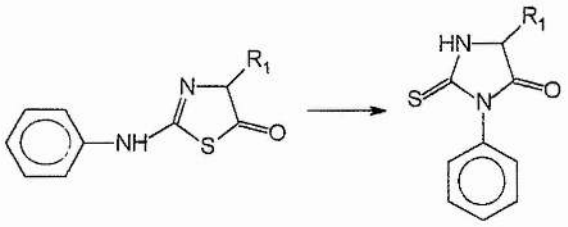


Figure 2.16 - Edman degradation: the conversion of the unstable thiazolinone derivative to a more stable PTH derivative.

The stable PTH derivative is then extracted and run on an HPLC system, where it elutes at a characteristic time which is matched with PTH-amino acid standards in order to determine which amino acid was in the original chain.

The most important side reactions consist of the amide groups of PTH-Asn and PTH-Gln being partially hydrolysed to PTH-Asp and PTH-Glu, and both PTH-Ser and PTH-Cys being decomposed during the conversion through a β-elimination, as well as PTH-Thr.

In automated Edman degradation, polybrene (1,5-dimethyl-1,5-diazaundeca-methylene-polymethobromide) is used as a film-stabilising carrier for the sequencing of proteins and peptides, which allows the reduction of wash-out of peptides.

2.5.1. Peptide Sequencing

Sample preparation consisted essentially on fixing the peptide to a membrane treated with polybrene (BioBrene Plus, Applied Biosystems) and loading it to the protein sequencer.

All the peptide sequencing was done at the University of St. Andrews using an Applied Biosystems Procise 491 Protein Sequencer.

2.5.2. Protein Sequencing

The protein sequencing methodology used involved separating the protein of interest by means of SDS-PAGE (similar to Method 2.2.1), followed by wet blotting (similar principle to Method 2.3) to transfer the protein to a PVDF membrane, and staining with amido black (similar principle to Coomassie Blue, Method 2.2.1.2 (Merril, 1990)) in order to visualise the protein band to sequence from the blot. The band can then be cut and loaded directly to the protein sequencer.

The gels are made essentially as per Method 2.2.1, but instead of using methylenebisacrylamide as a cross-linking agent, piperazine di-acrylamide (PDA) is used in order to reduce N-terminal blocking (Bio-Rad,).

Mini gels were prepared using a Bio-Rad Mini Gel kit. Separating gel solutions had a final concentration of 1.8 M acrylamide (12.5% (w/v)), 17 mM PDA (from a 30% (w/v) acrylamide, 41 mM PDA filtered (0.2 µm) stock solution, stored at 4°C in the dark), 0.375 M Tris-HCl, 3.5 mM SDS (from a 1.5 M Tris-HCl, 14 mM SDS, pH 8.8 stock solution), 3.3 mM TEMED and 1.6 mM $(NH_4)_2S_2O_8$ (from a 0.44 M (10% (w/v)) stock solution freshly prepared), which was topped with water saturated butanol until the gel set. Once set, the water saturated butanol was replaced by a 0.375 M Tris-HCl, 3.5 mM SDS solution and left at 4°C overnight.

Stacking gel solutions had a final concentration of 0.70 M acrylamide, 2.8 mM PDA (from a 30% (w/v) acrylamide, 41 mM PDA filtered (0.2 µm) stock solution, stored at 4°C in the dark), 0.375 M Tris-HCl, 3.5 mM SDS (from a 1.5 M Tris-HCl, 14 mM SDS, pH 8.8 stock solution), 6.6 mM TEMED and 3.3 mM $(NH_4)_2S_2O_8$ (from a 0.44 M (10% (w/v)) stock solution freshly prepared).

The gels were pre-run for 1 h at 15 mA per gel until sample buffer (25% (v/v) glycerol, 0.10 M SDS, 62.5 mM Tris-HCl, 0.05% bromophenol blue, 0.71 M β-MeOH (added just before use), pH 6.8) loaded in one lane reached the running gel, with a negative electrode buffer of 0.125 M Tris-HCl, 3.5 mM SDS, 0.05 mM glutathione (from a 10 mM stock solution freshly prepared), pH 6.8, and a positive electrode buffer of 75 mM Tris-HCl, 0.576 M glycine, 3.47 mM SDS.

Buffers were then replaced by 75 mM Tris-HCl, 0.576 M glycine, 3.47 mM SDS, 0.08 M sodium thioglycollate (from a 100 mM stock solution freshly prepared) for the negative electrode buffer, and 75 mM Tris-HCl, 0.576 M glycine, 3.47 mM SDS as the positive electrode buffer. Samples (15-20 µl) were added to 5 µl of sample buffer (see above), boiled for 2-5 min and loaded onto the gel, and a constant current of 20 mA was applied per gel, which ran for ~40 min.

A blotting sandwich was then assembled as follows: negative electrode, fibre pad soaked in transfer buffer (10 mM 3-[cyclohexylamino]-1-propanesulfonic acid (CAPS), 10% (v/v) methanol), one sheet of filter paper soaked in transfer buffer, polyvinylidene difluoride (PVDF) membrane (previously soaked in methanol for ~2 min, then in transfer buffer), the above gel (soaked in transfer buffer for at least 10 min with several buffer

changes), one sheet of filter paper soaked in transfer buffer and positive electrode. This was run at a constant current of 300 mA for 2 h, with a cooling container that was changed after one hour. The resulting blotted PVDF membrane was then stained for 30 s with amido black stain (0.1% (w/v) amido black, 40% (v/v) methanol and 1% (v/v) acetic acid) and destained with several washes of distilled water. The appropriate band was cut with the aid of a scalpel and loaded to the protein sequencer.

All the protein sequencing was done at the University of St. Andrews using an Applied Biosystems Procise 491 Protein Sequencer.

2.6. *Mass Spectrometry*

Mass spectrometry is a technique that enables the determination of the molecular weight of molecules.

Basically, a mass spectrometer consists of a ion source where ionisation of the sample occurs, followed by ion separation through an electric and/or magnetic field. The separated ions are then detected and its mass-to-charge ratio is recorded (Rose and Johnstone, 1982).

All mass spectrometry was carried out at the University of Aberdeen using a Vestec Lasertec benchtop laser desorption time-of-flight mass spectrometer. Laser desorption is an ionisation method in which polar non-volatile molecules (including salts) are coated onto a metal surface in a thin layer which is then subjected to a short but intense laser pulse that causes thermal desorption of alkali ions present as impurities in the metallic probe or added intentionally to the sample. These ions become attached to molecules to yield cationised species which are then accelerated out of the ion source. Little fragmentation of the molecules is observed, which enables the analysis of amino acids, peptides, nucleotides and oligosaccharides (Rose and Johnstone, 1982).

In ion separation by time-of-flight there is no magnetic separator and ions are separated by making use of their different velocities after acceleration through a potential. The ion velocity is mass dependent and therefore if ions are accelerated and then allowed to pass along a field-free region, the ions will arrive at the detector at different time intervals depending on their velocities, hence on their masses. Unlike other separation methods, this has no mass range limitations, thus allowing detection of ions of very high molecular weight (Rose and Johnstone, 1982).

Samples (typically 1 mg/ml for peptides and 50-200 µg/ml for protease) were diluted 1:10 with 0.1% (v/v) trifluoroacetic acid, and 0.5 µl was mixed with 0.5 µl of matrix solution (1% (w/v) sinapinic acid (3,5-dimethoxy-4-hydrocinnamic acid) in 0.05% (v/v) trifluoroacetic acid) in 50% (v/v) acetonitrile. Spectra were an average of 64 scans.

2.7. *Determination of Peptide Concentration*

Peptide concentrations were determined either by quantitation of N-terminal amines by reaction with fluorescamine and compared to standard curves of peptides of known concentrations or by comparison of electropherogram areas resulting from capillary electrophoresis, again compared to areas of peptides of known concentration.

2.7.1. Capillary Electrophoresis

Classical electrophoresis is a process in which charged molecules are separated based on their movement through a fluid under the influence of an applied electric field (Weinberger, 1993). In capillary electrophoresis, samples are separated into their components by their respective mass-to-charge ratios, size and chemical interaction (Bio-Rad, 1996). Each analyte migrates electrophoretically inside an electrolyte-filled capillary past an absorbance detector and its light absorption generates a signal proportional to its concentration:

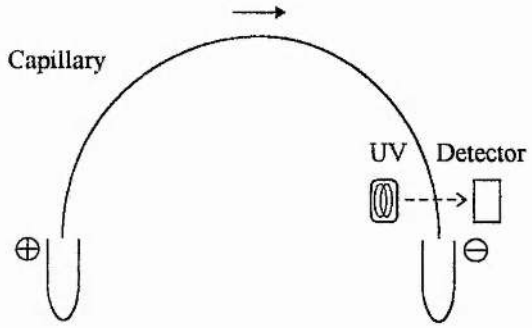


Figure 2.17 - Scheme of capillary electrophoresis. The sample in the left vial is inserted into the capillary and separated by an electric field. The separated analytes are detected at the other end of the capillary.

In order to achieve both a good separation and resolution in a short time, voltage field strengths of over a thousand volts per centimetre can be applied. Capillary electrophoresis combines the high resolution of traditional gel electrophoresis (cf.

Method 2.2.1) with the sensitivity, ease of use and high speed of HPLC (cf. Method 2.4.2.1).

The technique used to separate the analytes was Capillary Zone Electrophoresis (CZE). This technique consists of pre-concentrating the sample (sample stacking) by injecting it from a low-conductivity matrix into a run buffer which has a substantially higher conductivity (Bio-Rad, 1996; Weinberger, 1993). The samples must present a positive charge in order to migrate to the negative electrode, so they must be kept in a buffer at low pH. The output of capillary electrophoresis is called an electropherogram, which displays the absorbance of the outcoming analytes as a function of their elution time.

2.7.1.1. Sample Preparation and Analysis

All the samples had a volume of either 50 or 100 µl, and were composed 1:1:8 respectively of the actual sample, filtered (0.2 µm) phosphate buffer (Bio-Rad) at pH 2.5 and filtered (0.2 µm) HPLC water. These were run in a BioFocus 3000 capillary electrophoresis system, with a pressure injection of 20 PSI × s, at a constant voltage of 10 kV using coated capillary cartridges of either 17 cm × 25 µm or 24 cm × 25 µm, running from the positive to the negative electrode in a phosphate buffer, pH 2.5 (Bio-Rad). The absorbance was read at a single wavelength of 200 nm. All equipment was from Bio-Rad. All electropherograms were collected by an IBM compatible PC, using the Bio-Rad software package BioFocus V3.10 for Windows 3.1x, on .BFF format (BioFocus File) files.

2.7.1.2. Spectra Integration

All electropherograms (spectra) were analysed using the Bio-Rad software package Integrator V3.01, where the .BFF files were converted to .ACQ files (Acquisition) and integration was performed using manual baseline editing of the observed peaks. The output reports comprised the elution times of each peak together with its height and area, and the areas were given both in absolute and percentual forms.

2.7.2. Fluorescamine Assays

This assay is based on the fact that fluorescamine reacts with primary amines (Udenfriend *et al.*, 1972; Weigele *et al.*, 1972) (more recent reports show that it reacts

also with secondary amines, although not yielding a fluorescent compound (Nakamura *et al.*, 1980)):

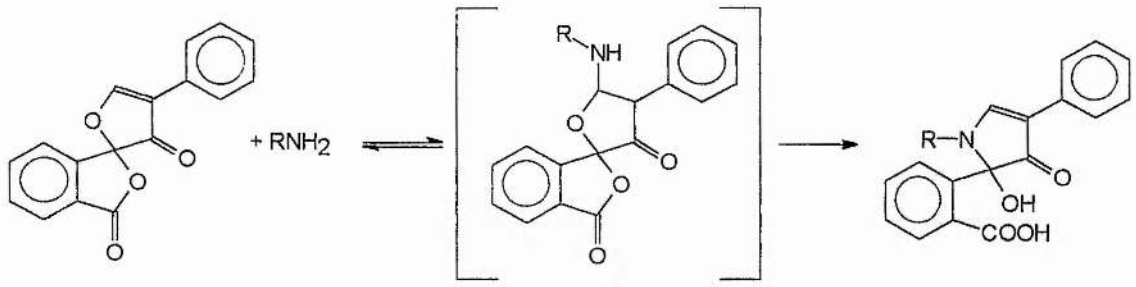


Figure 2.18 - Reaction of fluorescamine with a primary amine (Stein *et al.*, 1974).

Fluorescamine is not fluorescent, but the fluorescent reaction product with a primary amine has λ_{ex}=390 nm and λ_{em}=475 nm (Udenfriend *et al.*, 1972). Moreover, fluorescamine is readily hydrolysed in water and its hydrolysis products are not fluorescent. However, the reaction of fluorescamine with primary amines is much faster than its hydrolysis, allowing it to label the primary amines prior to the complete degradation of the excess reagent.

The observation that fluorescamine does not react with guanidine or creatine leads to the conclusion that arginyl groups do not contribute with amine groups for the fluorescamine reaction (Böhlen *et al.*, 1973). Instead, lysinyl groups do react with fluorescamine (Böhlen *et al.*, 1973; Schwabe, 1973; Singh and Sonar, 1988; Stein and Udenfriend, 1984) although in some circumstances this seems not to happen (Nakai *et al.*, 1974). This should depend on the pH conditions of the assay, as well as the pK_a of the amino groups (de Bernardo *et al.*, 1974), since unprotonated amines are expected to react due to its nucleophilic character whereas a protonated amine like ammonia is known not to react (Udenfriend *et al.*, 1972).

2.7.2.1. Assay for Peptide Concentration using Fluorescamine

A solution of 0.72 mM of fluorescamine (Sigma) in acetone was prepared and kept at room temperature in a flask covered by tin foil to protect from light. To 2 ml of PBS buffer (0.14 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4 (Sambrook *et al.*, 1989)) 10-20 µl of the peptide solution were added and as the solution was stirred 0.5 ml of fluorescamine solution were added. The fluorescence of the resulting solution was measured on a Hitachi F-1050 Fluorescence Spectrophotometer

with λ_{ex}=375 nm and λ_{em}=480 nm found to be the optimal working wavelengths (Cabrita, 1992).

2.8. Chemical Modifications to Peptides

The chemical modifications made to peptides consisted only on either fully reducing any disulphide bridge between cysteines of the peptide, or promoting the full dimerisation of those cysteines to cystines. These were based on pH control, necessary to ensure that the thiol group was ionised to thiolate, as this is the reacting species in oxidation-reduction interchange (Kolthoff *et al.*, 1955).

2.8.1. Oxidation (Dimerisation) of Peptide

The peptide was dissolved in 0.1 M $NaHCO_3$ and left stirring at room temperature overnight. It was then purified using reverse-phase HPLC (Method 2.4.2.1).

2.8.2. Reduction of Peptide

The peptide was dissolved in 0.1 M $NaHCO_3$ and a 3-fold excess of dithiothreitol (DTT) in regard to disulphide bridges was added. The solution was then bubbled with nitrogen for 2 min and stirred at room temperature for 45 min. 0.2 M acetic acid was then added to the solution until a pH of 4-5 was reached, and the peptide was purified by reverse-phase HPLC (Method 2.4.2.1).

2.9. Proteolysis Assays

All these assays were based on the digestion of synthetic peptides (Method 2.4.1) whose sequence mimics the established consensus sequence (Anderson, 1990; Webster *et al.*, 1989a).

2.9.1. Digestion of Peptides LSGAGFSW and Ac-LRGAGRSR

The design of these peptides was based on considerations of solubility and relative susceptibility to digestion. It is known (Webster *et al.*, 1989a) that the peptide MSGAGFSW is a good substrate for the protease, and the substitution of methionine to leucine was aimed at a more stable substrate (methionine can undergo chemical changes during synthesis), and further substitutions of serine, phenylalanine and tryptophan to arginine tried to improve solubility whilst still keeping the consensus sequence. The acetylation of the N-terminal was used to provide with a possible substrate for

fluorescamine assays (Method 2.7.2), since the result of digestion would release a new free amino group which could be titrated with the above reagent (Evans and Ridella, 1984).

The equations for substrate digestion by the protease are respectively:

$$LSGAGFSW \rightarrow LSGA + GFSW \qquad \text{Eqn. 2.9.1}$$

and:

$$Ac\text{-}LRGAGRSR \rightarrow Ac\text{-}LRGA + GRSR \qquad \text{Eqn. 2.9.2}$$

As the protease is not active at low pH (Webster *et al.*, 1989b), dropping the pH of the assay mixture was the method used to stop the reaction in order to take time samples.

The reaction conditions are a modification of a previous assay (Webster *et al.*, 1989b). In a typical assay, the recombinant protease (10 µl of 0.02-0.06 mg/ml) was incubated with a solution of the activating peptide or its equivalent (10 µl of 0.25-1 mg/ml) and buffer (25 µl of 50 mM Tris-HCl, 10 mM EDTA, pH 8, 2 mM β-MeOH) at 37°C for 15 min, followed by the addition of the substrate (5 µl of 2 mg/ml of synthetic peptide) equilibrated at 37°C. Samples (5 µl) were taken at required time points and analysed using Method 2.7.1.

2.9.2. Digestion of $(LRGG)_2$-Rhodamine

This substrate was based on a previously reported method (Mangel *et al.*, 1993). The substrate digestion by the protease can be described as follows:

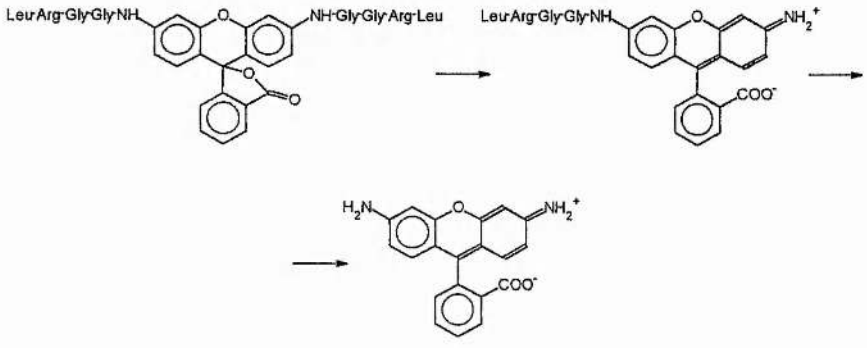


Figure 2.19 - Protease cleavage of $(LRGG)_2$-rhodamine. Both the intermediary monoamide and the final rhodamine 110 are fluorescent.

This assay provided a continuous measurement of digestion kinetics, quite useful for following kinetics of activation of the protease. Rather than the kinetics of cleavage of

substrate by a fully activated protease, the aim of this assay was to follow the kinetics of activation of the protease

In a typical assay, 100 µl of $(LRGG)_2$-rhodamine substrate solution (0.03 mM) were mixed with 5 µl of solution of the activating peptide or its equivalent (1 mg/ml), 300 µl of buffer (25 µl of 50 mM Tris-HCl, 10 mM EDTA, pH 8, 2 mM β-MeOH) and 190 µl of HPLC water, and incubated at 37°C for 10 min, and 5 µl of protease (0.02-0.06 mg/ml) were added, monitoring the fluorescence with a Perkin-Elmer Luminescence Spectrometer LS 50 B, with λ_{ex}=492 nm and λ_{em}=523 nm, with a 5 nm slit width.

2.9.3. Inhibition Assays

Inhibition assays were intended to assess the ability of several peptides to inhibit activation of the protease with the activating peptide, rather than being directed at the catalytic site of the protease. Therefore, the assays consisted of adding a possible inhibitor to the protease followed by the addition of the activating peptide, and assaying for protease activity by digestion of a synthetic peptide, as in Method 2.9.1.

In a typical assay, the recombinant protease (10 µl of 0.02-0.06 mg/ml) was incubated with a solution of the possible inhibiting peptide (10 µl of 0.25-1 mg/ml) and buffer (25 µl of 50 mM Tris-HCl, 10 mM EDTA, pH 8, 2 mM β-MeOH) at 37°C for 15 min, followed by the addition of the activating peptide (10 µl of 0.25-1 mg/ml) and incubation of another 15 min at 37°C. The substrate (5 µl of 2 mg/ml of synthetic peptide) equilibrated at 37°C was then added and samples (5 µl) were taken at required time points and analysed using Method 2.7.1.

2.10. Determination of the pK_a values of the Activating Peptide

A diprotic acid dissociates its protons according to the following scheme (Edsall *et al.*, 1958):

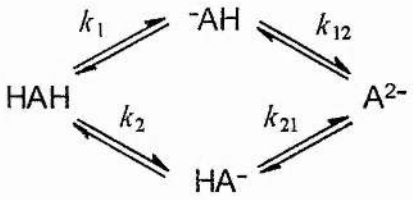


Figure 2.20 - Scheme of the ionisation of a diprotic acid, illustrating the meaning of the four "microscopic" constants. The symbol H on the left of HAH is the acidic proton denoted as (1), the one on the right is (2).

Each of the above constants, such as k_1, is defined by the relation $k_1 = (H^+)(^-AH)/(HAH)$, etc., where (H^+) denotes the hydrogen ion activity, etc. The notation used is that of Hill (quoted in (Edsall *et al.*, 1958)), in which the last numeral in the subscript denotes the group that donates a proton in the process, while the numerals that precede it indicate the groups already deprotonated, if any.

In an ordinary titration, in which pH is determined as a function of the mean number of protons removed from the diprotic acid by addition of a base, the values of the microscopic constants can not be determined directly. However, two "macroscopic" constants can be determined, K_1 and K_2, according to the scheme:

$$HAH \overset{K_1}{\rightleftharpoons} (^-AH + HA^-) \overset{K_2}{\rightleftharpoons} A^{2-}$$

Figure 2.21 - Scheme of the ionisation of a diprotic acid, illustrating the "macroscopic" constants.

There are several methods available for determining these macroscopic constants (see, for instance (Parke and Davis, 1954; Simms, 1926a; Simms, 1926b)). Furthermore, relations have been described between the macroscopic and microscopic constants (Ryklan and Schmidt, 1944):

$$K_1 = \frac{(H^+)\left[(^-AH) + (HA^-)\right]}{(HAH)} = k_1 + k_2 \qquad \text{Eqn. 2.10.1}$$

$$K_2 = \frac{(H^-)(A^{2-})}{\left[(^-AH) + (HA^-)\right]}; \quad K_2^{-1} = k_{12}^{-1} + k_{21}^{-1} \qquad \text{Eqn. 2.10.2}$$

Another important relation comes from Eqn. 2.10.1 and Eqn. 2.10.2:

$$k_1 k_{12} = k_2 k_{21} = K_1 K_2 = \frac{(H^-)^2 (A^{2-})}{(HAH)} \qquad \text{Eqn. 2.10.3}$$

When the two ionisable protons are indistinguishable, $k_1 = k_2$ and the macroscopic constants (also denominated "apparent dissociation constants") suffice to describe the acid/base behaviour of the molecule. If, however, those protons come from distinct groups, their lability will be different and macroscopic constants can only be used if one of the protons is much more ionisable than the other, so that $k_1 >> k_2$ and $K_1 \cong k_1$. In the opposite case, when k_1 and k_2 are within the same order of magnitude, the microscopic constants have to be used.

In the latter situation, the above equations are insufficient to determine the four microscopic constants, one additional relation being needed in order to obtain independent checks of the values for the microscopic constants. Two main approaches can be employed: (a) convert one of the acidic groups in HAH to an alkyl derivative RAH, or HAR', where R and R' generally denote methyl or ethyl groups, or (b) using a method that enables the determination of the extent of ionisation of only one of the groups, be it by spectrophotometry, nuclear magnetic resonance (NMR), Raman spectroscopy, etc.

Using the first approach, it is assumed that the effect of the group RA- on the ionisation of hydrogen is equivalent to that of HA$^-$. If the hydrogen **H**AH is the acidic hydrogen of a carboxyl group and RAH its monoester, this is known as the principle of Wegscheider (referenced in (Edsall *et al.*, 1958)). Therefore, when this equivalence applies, the pK value of HAR' is equal to that of pk_1 in Figure 2.20, and that of RAH is equal to pk_2. The other values can then be determined with the aid of the above equations.

In the second approach, if the extent of ionisation of only one group can be determined by some method, postulating that it is the hydrogen **H**AH of Figure 2.20 a function α_1 can be defined that represents the fraction of ionised **H** as follows (Edsall *et al.*, 1958):

$$\alpha_1 = \frac{(^-AH) + (A^{2-})}{(HAH) + (^-AH) + (HA^-) + (A^{2-})}$$

$$= \frac{k_1 \cdot (H^+) + k_1 k_{12} \cdot (H^+)^2}{1 + (k_1 + k_2) \cdot (H^+) + k_1 k_{12} / (H^+)^2} \qquad \text{Eqn. 2.10.4}$$

The following function M_1 can then be readily defined as:

$$M_1 \equiv \frac{(H^+)\alpha_1}{1-\alpha_1} = \frac{k_1(H^+) + k_1 k_{12}}{(H^+) + k_2} = \frac{k_1(H^+) + k_2 k_{21}}{(H^+) + k_2} \qquad \text{Eqn. 2.10.5}$$

or, in logarithmic form,

$$pM_1 = pH - \log\frac{\alpha_1}{1-\alpha_1} = -\log\frac{k_1(H^+) + k_1 k_{12}}{(H^+) + k_2} = -\log\frac{k_1(H^+) + k_2 k_{21}}{(H^+) + k_2} \qquad \text{Eqn. 2.10.6}$$

It follows from Eqn. 2.10.4 and Eqn. 2.10.6 that when $(H^-) \rightarrow +\infty$, $\alpha_1 \rightarrow 0$ and $pM_1 \rightarrow pk_1$, and when $(H^-) \rightarrow 0$, $\alpha_1 \rightarrow 1$ and $pM_1 \rightarrow pk_{21}$. The function pM_1 is then completely determined by the ratios k_1/k_{21} and k_2/k_1. The character of the function defined by Eqn. 2.10.6 is illustrated in detail by the curves shown in Figure 2.22. These curves are calculated taking $pk_{21} - pk_1 = 1$ and therefore the intercepts at $\alpha_1 = 0$ and $\alpha_1 = 1$ coincide in all the curves. The different curves correspond to different values of the ratio k_2/k_1. Similar families of curves are obtained for other values of $pk_{21} - pk_1$. It is apparent from Figure 2.22 that, if $k_2 >> k_1$, the value of pM_1 rapidly rises to a value near to pk_{21} with the increase of α_1. Therefore, in this case the determination of pk_{21} by extrapolation is straightforward but the converse for pk_1 is not true, the extrapolation being reasonably uncertain. In the same manner, if $k_2 << k_1$, the values of pM_1 remain close to pk_1 until α_1 approaches 1, when the curve suddenly raises to reach pk_{21}. Hence, in this case the extrapolation for the determination of pk_1 is accurate but the estimation of pk_{21} remains difficult. However, if k_1 and k_2 are within the same order of magnitude (the ratio k_2/k_1 lying between 5 and 1/5) both the above constants can be determined with reasonable accuracy.

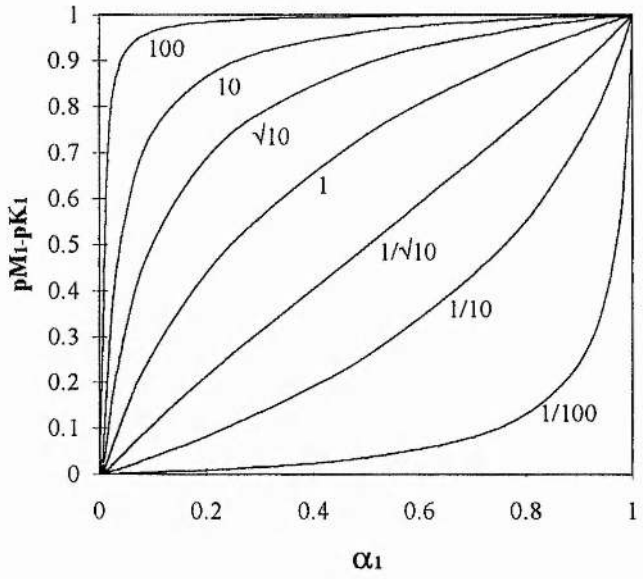


Figure 2.22 - The function pM_1 represented as a function of the fractional ionisation α_1. The zero of the ordinate scale is taken as $pM_1 = pk_1$ and all the curves correspond to an assumed value of $pk_{21} - pk_1 = 1$ pH unit. The number attached to each curve represents the corresponding value of the ratio k_2/k_1 (after (Edsall *et al.*, 1958)).

Once k_1 and k_{21} have been determined, it is possible to take these two constants as known quantities and to solve for k_2. There are several ways of estimating this (Edsall *et al.*, 1958). The approach taken was to computer generate a family of curves for several values of k_2, and evaluate visually which is the best fit, as the equations are non-linear and a least-squares fit is not possible, and the use of more complex fitting algorithms seemed unnecessary, given the required accuracy and the experimental error involved. Furthermore, a visual evaluation already enables an accuracy of 0.1 pH units.

2.10.1. Experimental Determination of the pK_a values

There are several ways of detecting the fraction of ionised cysteines: by spectrometry, as the thiolate anion absorbs at a wavelength of 230-240 nm whereas the thiol group does not (Benesch and Benesch, 1955; Coates *et al.*, 1969; Gorin and Clary, 1960); by NMR, by observing the change in the chemical shift of the thiol group with its ionisation (Rabenstein, 1973; Shrager *et al.*, 1972), etc. The spectrophotometric method was the one chosen, due to its simplicity.

Peptides were dissolved in a constant (I = 0.1) ionic strength buffer (0.05 M acetic acid, 0.05 M 2-(N-morpholino)-ethanesulphonic acid (MES), 0.1 M Tris, (Ellis and Morrison, 1982)), the pH lowered with the addition of 1 M HCl, and this solution was titrated with either 0.5 or 1 M NaOH with constant mixing, with measured values of pH

and absorbance at 238 nm being recorded for several volumes of added NaOH. Values of pH were measured at room temperature using a Jenway 3010 pH Meter, with the electrode being calibrated with buffers of pH 4.0 and 7.0 (Russell), and absorbance was measured using a LKB Ultraspec II Biochrom spectrophotometer, readings being corrected for the increase in volume.

2.11. Determination of the Oxidation State of the Activating Peptide

Essentially, two methods were used: capillary electrophoresis (Method 2.7.1) and Ellman's reagent. The first method is based on the observation that the oxidised (dimeric) form of the peptide has a higher electrophoretic mobility than its reduced counterpart, thus appearing first on an electropherogram. The second method is described below.

2.11.1. Ellman's Reagent

Ellman's reagent 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) (Ellman, 1959; Jocelyn, 1987) reacts with thiols, as aromatic disulphides are easily reduced by thiol-disulphide interchange:

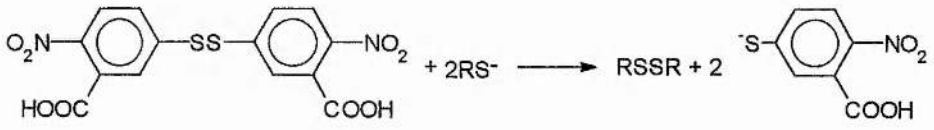


Figure 2.23 - Reaction of DTNB with thiolate anions. The resulting aromatic thiol absorbs at 412 nm.

Though with protein thiols the mixed disulphides are stable products, non protein thiols are themselves quickly reduced. In any circumstance, the overall stoichiometry is one aromatic thiol generated for each biological thiol originally present.

Care should be taken when handling DTNB, as although it is stable indefinitely when kept in the dark at 0°C, it rapidly develops a yellow colour if exposed to light at room temperature. Also, if the pH locally reaches 9 the disulphide is hydrolytically cleaved with formation of yellow thiol (Jocelyn, 1987).

Once it is allowed to react with the biological thiol, the absorbance at 412 nm should be read after at least 2 minutes, and this remains constant over the next 8 minutes. Reasons for decline of absorption include autoxidation of the aromatic thiol, and increase is due to slow hydrolysis (Jocelyn, 1987).

2.11.1.1. Procedure for Determination of Thiol Groups with DTNB

From a stock solution of 10 mM DTNB in PBS (0.14 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4 (Sambrook *et al.*, 1989)), 6.7 µl were added to 1 ml of protein/peptide solution, and absorbance was read at 412 nm after 2-3 min, against a blank of DTNB and buffer.

2.12. Determination of Initial Rates of Digestion

Usually, for the determination of the initial rate of digestion, this rate is commonly considered to be approximately linear until about 10% of the substrate is digested (Eisenthal and Danson, 1993) (although some consider that the ideal situation would be less than 1% (Cornish-Bowden, 1995)), and a plot of number of molecules of substrate S digested against time t of digestion will yield a linear relationship between this two variables. The initial rate v of cleavage will then be the slope of that line.

However, the method used to determine the percentage of substrate digestion S/S_0 (Method 2.7.1) was not sensitive enough to determine accurately little amounts of digestion such as less than 10%. Therefore, an analytical equation which describes a general digestion curve of a substrate had to be used. The following was obtained by manipulating mathematically simple kinetic equations, though there are several ways of obtaining the same result (Cornish-Bowden, 1995).

All digestions of substrate in Method 2.9.1 can be described with the following equation:

$$S \rightarrow P \qquad \text{Eqn. 2.12.1}$$

where S is the substrate and P are the products of digestion. Although in a digestion there are two molecules of products from the cleavage of each molecule of substrate (see Method 2.9.1), that is irrelevant for the following determinations if we consider it to be an irreversible reaction. Furthermore, although these are hydrolysis reactions, as water is in large excess, this results in an overall pseudo first order reaction.

So, for an first order reaction (Perry and Green, 1984), its rate v is defined as the variation of substrate concentration dS that is consumed in the time interval dt. This is described mathematically by:

$$v = -\frac{dS}{dt}$$ Eqn. 2.12.2

On the other hand, the rate v of an irreversible first order reaction is only proportional to the substrate concentration:

$$v = kS$$ Eqn. 2.12.3

The constant k is termed **rate constant** and it is independent of the substrate concentration. The initial rate will then be the product of the rate constant by the initial concentration of substrate S_0. In order to determine k, Eqn. 2.12.2 and Eqn. 2.12.3 can be equalled (Wesley, 1969):

$$-\frac{dS}{dt} = kS$$ Eqn. 2.12.4

Integration of Eqn. 2.12.4 yields:

$$\ln\left(\frac{S}{S_0}\right) = -kt$$ Eqn. 2.12.5

which is an equation of the form $y = Ax$, with $y = \ln (S/S_0)$ and $x = t$. So, a plot of $\ln (S/S_0)$ against t will yield a straight line of slope $-k$ that, when replaced in Eqn. 2.12.3 together with the initial substrate concentration, will calculate the initial rate of digestion.

Finally, by rewriting Eqn. 2.12.5 we can get an analytical equation that describes the dependence of percentage of substrate digestion with time:

$$\frac{S}{S_0} = e^{-kt}$$ Eqn. 2.12.6

2.12.1. Experimental Determination of Initial Rates

Digestion assays were performed as described in Method 2.9.1, and analysed using Method 2.7.1. The integrations obtained by Method 2.7.1.2 were then treated with Microsoft Excel 5.0 as illustrated in the following example:

Table 2.2 - Example of data treatment of digestion assays.

	A	B	C
1	Time (min)	S/S0	ln(S/S0)
2	5	0.9083	-0.096181
3	10	0.8228	-0.195042
4	15	0.7412	-0.299485
5	20	0.6645	-0.40872
6	30	0.5251	-0.644167

Cells C2 to C6 contain the formula `=ln(B2)` to `=ln(B6)` respectively. A linear regression was then performed:

Table 2.3- Example of calculation of a linear regression in order to calculate the rate constant.

		G
slope (-*k*):	**2**	-0.020862
standard error:	**3**	0.000357
correlation coefficient $\mathbf{r}^2$:	**4**	0.995305
F statistic:	**5**	848.0609
regression sum of squares:	**6**	0.177867

where cells G2 to G6 contain the Excel array formula `{=LINEST(C2:C6,A2:A6,FALSE,TRUE)}`. The slope in G2 is equal to -*k*, so $k = 0.021$ min^{-1}. By using this value in Eqn. 2.12.6, we can plot the analytical curve that fits best the experimental data:

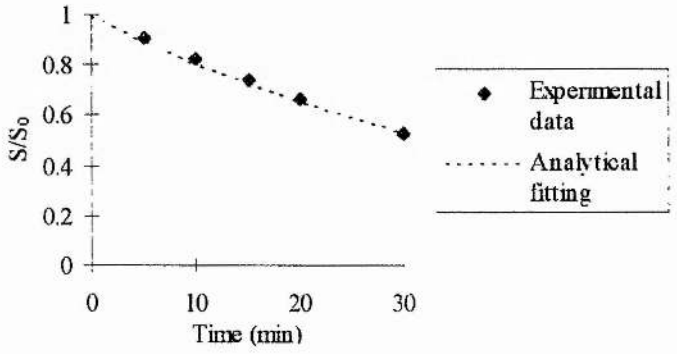


Figure 2.24- Graphical result of the fitting of experimental data with Eqn. 2.12.6.

3. Results and Discussion

The main objective of this work was to try and understand how the activating peptide GVQSLKRRRCF converts the adenovirus protease from an inactive to an active form. In order to achieve this purpose, several characteristics of the activating peptide were studied.

This chapter starts with some general experimental observations made throughout all the work regarding certain characteristics of the recombinant protease (3.1). It then follows to the activating peptide, in trying to assess what are the relevant properties of the peptide which enable it to somehow render the protease active: the importance of the cysteine residue (3.2), the effect of N-terminal truncations in the peptide on its activation properties (3.3), which led to an assessment of the importance of the N-terminal residues (3.4), and the influence of the environment around the cysteine (3.5). From the results obtained by these investigations, the idea of inhibition of activation arose and was followed further (3.6). Finally, an attempt was made to clone a protease from an avian adenovirus in *E. coli* (3.7).

Also, it was deemed necessary to include a few considerations on some of the methodologies used (3.8), in order to evaluate their validity and applicability.

3.1. The Adenovirus Recombinant Protease: General Features

While working with the recombinant protease *in vitro*, several characteristics of this enzyme were observed that made it a difficult protein to work with, and these are discussed.

The recombinant protease is very sensitive to temperature changes, and repeated freeze-thawing turns it inactivatable, that is, the addition of the activating peptide GVQSLKRRRCF is then unable to render the protease active. Also, if left at room temperature, it gradually loses the ability to be activated. This inactivation seems to depend on the concentration of mercaptoethanol present, which would suggest that this process is associated with some oxidation of the protease. Furthermore, the mere mixing of the protease after thawing in order to homogenise the solution does affect its ability to be activated, another fact that could be explained in terms of oxidation. For all of the above reasons, the attainment of a constant and reproducible digestion rate for the protease proved to be an almost impossible task.

However, once the protease is allowed to incubate with the activating peptide, it then becomes reasonably stable to any of the above perturbations and remains active for at least 8 hours (maximum period observed) without showing any signs of loss of activity. Interestingly enough, it has been observed (Jones *et al.*, 1996) that increasing concentrations of activating peptide allow a more efficient activation of the protease, as well as that a subsequent reduction of the activating peptide concentration reduces the activity of the protease once this is activated (Iqbal, unpublished results). However, after the peptide is allowed to incubate with the protease, the addition of further peptide does not seem to enhance the activity of an already activated protease. In other words, before activation the protease is sensitive to the concentration of activating peptide, whereas after it is activated it is sensitive only to a reduction of the activating peptide concentration.

Another interesting observation is that dithiothreitol (DTT) inhibits the recombinant protease activation when added before, whereas if DTT is added after the activation, it actually stimulates activity (Webster *et al.*, 1993). On the other hand, viral protease (already activated) seems to be stimulated by both DTT and cysteine up to an optimal concentration of 0.5-2 mM, followed by a decrease in activation for higher concentrations of these reagents (Webster *et al.*, 1993). Certainly, different protease preparations seemed to show a different susceptibility to reducing reagents, as well as stability at room temperature. Curiously, whereas some protease preparations seemed to present an activity (after activation) directly proportional to their protease concentration, some others did suffer a dramatic decrease in activity once diluted.

Finally, the recombinant protease was also observed to form dimers (Vaughan, personal communication), in a fixed proportion to the amount of monomer present, the two forms liable of being separated by a non-reducing SDS-PAGE. Attempts made to try and identify via which cysteine(s) these dimers occur have been unsuccessful, and it is possible that dimerisation occurs through several combinations of available cysteines. An interesting observation (Vaughan, unpublished results) is that the incubation of a protease dimer with the activating peptide reduces the amount of dimerised form to what seems to be the monomer possibly bound to the activating peptide.

3.2. *Importance of Cys-10 in the Native Activating Peptide*

The observation that the peptide GVQSLKRRRAF did not activate the adenovirus protease led to the assumption that the cysteine residue in position 10 of the activating peptide could be involved in a thiol-disulphide interchange with the protease (Webster *et al.*, 1993). In order to try and elucidate what was the nature of this interaction, several peptides were synthesised in which the cysteine was substituted by other amino acids. Three amino acid substitutions were tried: alanine, providing a side chain similar in length to a cysteine, though apolar, which was the original substitution that rendered the activating peptide non functional (Webster *et al.*, 1993); serine, quite similar to cysteine in both polarity and structure, although not as reactive and not allowing the formation of disulphide bridges; aspartate, a residue analogous in structure and function to a sulphenic acid (RSOH), which has been observed to contribute to redox control involving cysteine residues (Abate *et al.*, 1990).

3.2.1. Residue Substitutions of Cys-10

Two peptides were synthesised, GVQSLKRRRSF and GVQSLKRRRDF as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2; GVQSLKRRRAF had been synthesised previously (Webster *et al.*, 1993).

Preliminary experiments were made with non-purified peptides (used immediately after the freeze-drying that follows synthesis), attempting to activate the protease for 10 min before the addition of substrate by incubating the protease (0.02 mg/ml) with each of the peptides at 37°C (Method 2.9.1). The substrate LSGAGFSW was then allowed to digest for 30 min at several concentrations (0.15, 0.12, 0.089, 0.059, 0.030 and 0.015 mM), of each of the peptides. Time samples were taken and analysed using capillary electrophoresis (Method 2.7.1.1). The results are shown in Figure 3.1:

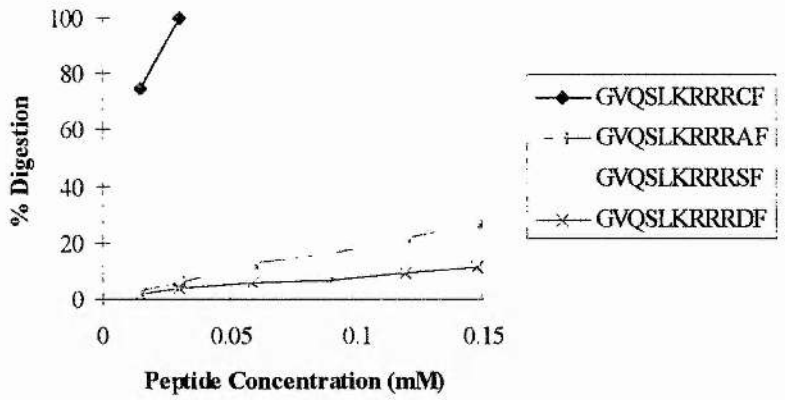


Figure 3.1- Effect of cysteine substitution in the activating peptide on activation of the adenovirus protease. Curves correspond to a 30 min digestion of the substrate LSGAGFSW. The peptides are not purified.

These results show that cysteine is the only residue that effectively enables the activation of the protease. However, in order to try and explain the reasonably high activation with some of the peptides at higher concentration they were further purified by HPLC (Method 2.4.2.1), to discard any contaminants remaining after the synthesis and cleavage.

After purification, peptide samples were analysed with capillary electrophoresis (Method 2.7.1.1) in order to check for purity, as the presence of more than one peak would indicate evidence of a contaminant. As a reference, typical percentages of acetonitrile for migration of the peptides (4.6 mm × 25 cm C18 column), as well as migration times on capillary electrophoresis (CE) are shown in Table 3.1.

Table 3.1 - Typical values of purification parameters for several peptides. Migration times for CE are given for a 24 cm × 25 μm capillary.

Peptide	Percentage of acetonitrile for peptide elution	Migration time on CE (min)
GVQSLKRRRCF	35-45	3.3
GVQSLKRRRAF	40-50	3.5
GVQSLKRRRSF	40-50	3.5
GVQSLKRRRDF	40-50	3.8

When these purified peptides were used to activate the protease in the same manner as described above, the results obtained were as follows:

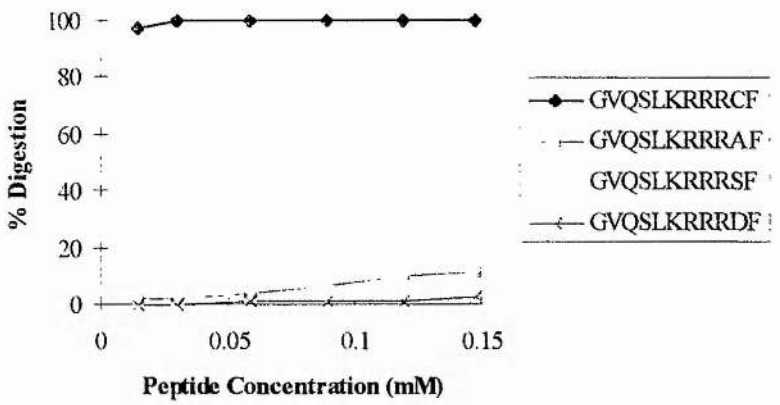


Figure 3.2 - Effect of cysteine substitution in the activating peptide on activation of the adenovirus protease. Curves correspond to a 30 min digestion of the substrate LSGAGFSW. Purified peptides.

As well as showing a more distinctive difference between GVQSLKRRRCF and the remaining peptides, stressing the importance of the cysteine in activation, these results also showed that impure peptides must contain a common contaminant which seems to help in activation. One possibility might be ethanedithiol, a scavenger used to remove tBu and Mtr groups, as it also contains thiol groups. All non-purified peptides present a characteristic odour which can be ascribed to this contaminant.

3.2.2. Oxidation State of Cysteine

Previous results indicated that the activating peptide needs to be in the oxidised (dimeric) form in order to efficiently activate the protease (Webster *et al.*, 1993). Measurements of initial rates showed a 10 fold increase of that value when oxidised activating peptide was used instead of the reduced version, which enabled the conclusion that the oxidised form should be the active molecule.

The above observation led to the assumption that the activating peptide was catalysing the oxidation/reduction of specific cysteine/cystine residues in the protease by a mechanism of thiol-disulphide interchange. A model was devised where the protease was synthesised as an inactive form in which the active site cysteine was involved in an intramolecular disulphide bridge, and the role of the activating peptide was to form a new disulphide bond leaving the active site cysteine as a free thiol (Webster *et al.*, 1993):

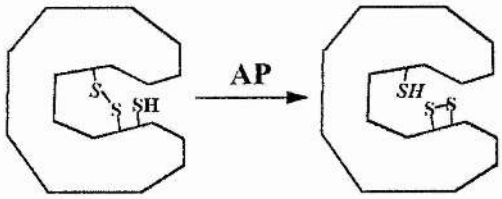


Figure 3.3 - The proposed mechanism of activation of the protease via a thiol-disulphide interchange. The activating peptide (AP) would promote the rearrangement of an internal disulphide bridge within the protease molecule which would release the active site cysteine thiol (*S*).

In the above model, the activating peptide would act presumably by setting the appropriate redox potential to allow for the interchange to occur.

In order to evaluate how determinant the oxidation state is, and to provide with some better insight into the mechanism of activation, some experiments were conducted with oxidised and reduced versions of the activating peptide, in which its ability to activate the protease was assessed.

The activating peptide was prepared in the reduced (Method 2.8.2) or oxidised form (Method 2.8.1). The presence of free thiols in each preparation was assayed (Method 2.11.1.1) to check that the peptides remained in the appropriate redox state, and this was verified by CE (Method 2.7.1.1), as the oxidised activating peptide migrates faster than its reduced counterpart. The two forms were then dissolved in either 5 mM β-MeOH (to prevent the reduced peptide from oxidation) or in HPLC water (for the oxidised form) at a concentration of ~1 mM. These solutions were then used in activity assays (Method 2.9.1), with 10 µl of protease (0.05 mg/ml) being activated with 10 µl of either of the above activating peptide solutions, and 25 µl of assay buffer (without β-MeOH in the case of the reduced peptide, in order to keep the β-MeOH concentration constant in both assays). The substrate Ac-LRGAGRSR was then used to determine initial rates of digestion (Method 2.12.1).

It was found (Hawkins, 1997) that the initial rate using the oxidised peptide relative to that found using the reduced peptide was 1.15 ± 0.16, that is, the difference in activation between the two forms is not significantly different ($p < 0.05$). The cause for this was later found out to be that after the 15 min incubation period, both peptides were found mainly in the reduced form, most certainly due to the presence of β-MeOH.

More attempts were made (Reid, 1997), this time also trying to assess the influence of the peptide concentration in its activating properties. Again, the preparation of both

the reduced and oxidised forms were as described above, although it was chosen to dissolve the purified peptides only immediately prior to assaying, in order to try and keep their redox state unchanged. The results are shown in the following graph (protease concentration of 0.05 mg/ml):

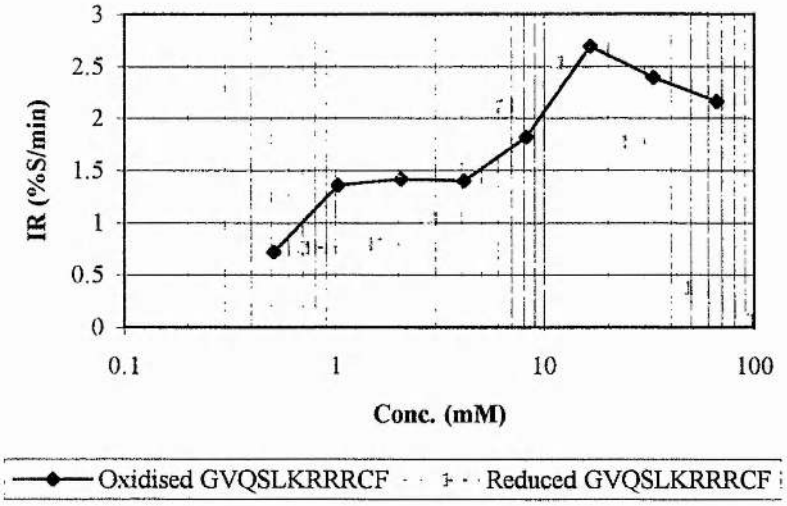


Figure 3.4 - The influence of the activating peptide (GVQSLKRRRCF) concentration and redox state in the activation of the adenovirus protease.

These results show again that there seems not to be any definite difference between the two forms within the peptide concentration range of 0.5-11 mM, considering that the results were not repeated in order to evaluate their statistical significance. However, the decrease in activation observed at higher concentrations was unexpected and could possibly have been caused by the presence of an impurity in the peptide solutions that only becomes apparent at higher concentrations.

Therefore, from the above results no difference between the oxidised and reduced forms of the peptide were found. The fact that the assay buffer contained β-MeOH must have contributed to this apparent lack of difference between the oxidised and reduced peptides. However, if β-MeOH was removed from the assay buffer, the protease presented a very variable degree of activation. Therefore the evaluation of which redox form of the peptide is the most effective activator of the protease was inconclusive.

3.2.3. pK_a of Cysteine

A striking feature of the native activating peptide (GVQSLKRRRCF) is the presence of three consecutive arginines to the left of the cysteine presumed to be

involved in the interaction to the protease. Furthermore, this sequence is reasonably well conserved within the several known serotype sequences of pVI (cf. Figure 1.9).

It has been reported that positively charged groups adjacent to a cysteine can enhance the formation of a specific reversible disulphide bond (Snyder *et al.*, 1981; Snyder *et al.*, 1983). This finding, together with the aforementioned observation of a high conservation of arginines nearby the cysteine, lead to the possibility that these arginines were responsible for a reactivity enhancement of the cysteine, eventually by lowering its pK_a below the typical value of 8.5 (Stryer, 1988).

The formation of a disulphide bond, concomitant with the oxidation of the thiol group, is thought to proceed in the following two steps (Benesch and Benesch, 1955):

$$RS^- \rightarrow RS^\bullet + e^- \qquad \text{Eqn. 3.2.1}$$

$$RS^\bullet + RS^\bullet \rightarrow RSSR \qquad \text{Eqn. 3.2.2}$$

A lower pK_a would facilitate the formation of the bond by increasing the rate of the reaction in Eqn. 3.2.1, causing the reaction in Eqn. 3.2.2 to be the limiting step.

Therefore, the determination of the pK_a of the cysteine was the consequent step, in order to try and evaluate whether the cysteine had an unusually low pK_a value which could explain the role of so many conserved arginines.

The determination of the pK_a value for the cysteine thiol is not straightforward in the case of the activating peptide, due to the presence of other ionisable groups, such as the carboxyl at the C-terminal, the three arginines, the serine hydroxyl group, the amino groups at the N-terminal and in lysine-6. Although the carboxyl group loses its proton at such a low pH (typically with a pK_a of 3.1 (Stryer, 1988)) that it should ionise completely prior to any other group, and the arginines deprotonate at such a high pH (typically with pK_a values of 12.0 (Stryer, 1988)) that they should ionise only after the complete deprotonation of amines and thiols, the situation is not so distinct between amino and sulfhydryl groups (Benesch and Benesch, 1955). The ionisation equilibria can be best described as follows:

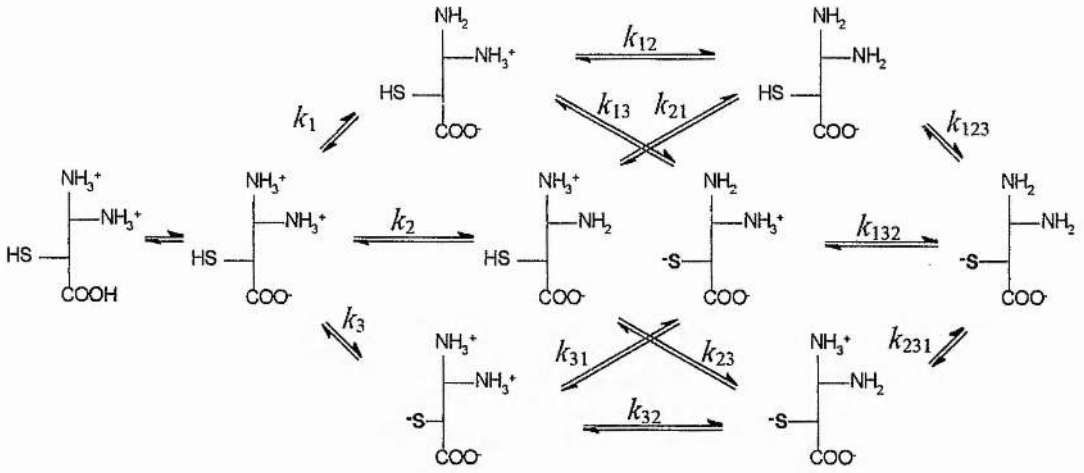


Figure 3.5 - The ionisation equilibria involved in the deprotonation of the activating peptide (GVQSLKRRRCF). Only the ionisable groups are represented, for simplicity. The numbering of the microscopic equilibrium constants is based on the notation of Hill, cited in (Edsall *et al.*, 1958). The amino terminal is group 1, the amino lysine is group 2 and the cysteine thiol is group 3. The carboxyl group was not numbered for simplicity.

The system described in Figure 3.5 is too complex for the equilibrium constants involved to be determined using Method 2.10 directly. Therefore, use was made of both the principle of Wegscheider and thiolate absorption measurements (cf. Method 2.10) in order to measure some of the constants involved so that the remaining unknowns could then be determined.

The approach taken was then to titrate peptides in which one of the amino groups was blocked and following their absorbance at 238 nm so that the ionisation of the thiol group could be monitored. It can be seen from Figure 3.5 that if the peptide Ac-GVQSLKRRRCF is titrated, the constants k_{12}, k_{13}, k_{123} and k_{132} can be determined by using Method 2.10, as the acetylated N-terminal should be equivalent to a non-protonated amino group (principle of Wegscheider). By a similar reasoning, the titration of the peptide GVQSLARRRCF should yield the constants k_{21}, k_{23}, k_{123} and k_{231}. These two determinations yield a value for k_{123}, which can be used as a guideline for the reliability of the results, or in case of experimental error it can provide with an estimate of what the real value should be, by matching the two values obtained from the titrations.

Using this approach, seven of the twelve equilibrium constants can be determined. In order to obtain the remaining five constants, an extension of Method 2.10 was devised, which can be applied to the titration of GVQSLKRRRCF, provided that values for the above constants are defined.

Equations corresponding to Eqn. 2.10.4 and Eqn. 2.10.6 can be written for the system described in Figure 3.5 as follows:

$$\alpha_3 = \frac{k_3/(H^-) + (k_1k_{13} + k_2k_{23})/(H^+)^2 + k_1k_{12}k_{123}/(H^+)^3}{1 + (k_1 + k_2 + k_3)/(H^-) + (k_1k_{12} + k_1k_{13} + k_2k_{23})/(H^+)^2 + k_1k_{12}k_{123}/(H^+)^3}$$

Eqn. 3.2.3

$$pM_3 = -\log\frac{k_3 + (k_1k_{13} + k_2k_{23})/(H^+) + k_1k_{12}k_{123}/(H^+)^2}{1 + (k_1 + k_2)/(H^+) + k_1k_{12}/(H^+)^2}$$

Eqn. 3.2.4

It then follows from Eqn. 3.2.3 and Eqn 3.2.4 that when $(H^+) \rightarrow +\infty$, $\alpha_3 \rightarrow 0$ and $pM_3 \rightarrow pk_3$, and when $(H^-) \rightarrow 0$, $\alpha_3 \rightarrow 1$ and $pM_3 \rightarrow pk_{123}$. From these equations and a plot of pM_3 vs. α_3, the values of k_3 and k_{123} can be determined, and the latter can be made to adjust to the previously obtained values for this constant. The unknown values in the above equations are then k_1 and k_2, but these are related by the equation $k_1 k_{12} = k_2 k_{21}$ and therefore assigning a value to one of them will define the other. The remaining two constants, k_{31} and k_{32}, can also be calculated by using similar equations, such as $k_1 k_{13} = k_3 k_{31}$ and $k_2 k_{23} = k_3 k_{32}$, which means that all the constants can be determined, using a process similar to that described in Method 2.10. Eventually, for a better agreement of all the constants, several iterations had to be made, so that all the values coincided for the three different titrations.

After these constants were determined, some other peptides were also titrated to assert about the influence of the residues around the cysteine on its acidity. For this purpose, the peptides GVQSLKRRRCA, GVQSLKRRCRF and GVQSLKRCRRF were also titrated and equilibrium constants determined.

All peptides were synthesised as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2, and purified by HPLC (Method 2.4.2.1).

3.2.3.1. The Titration of the Buffer Solution

When titrations were performed, an unexpected decrease of absorbance was observed between pH values of 3 and 6 (cf. Figure 3.6) which could not be explained in terms of thiol ionisation.

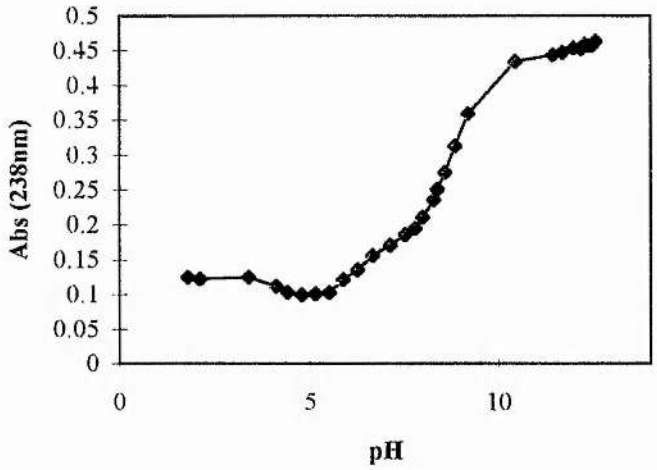


Figure 3.6 - Example of titration curve of a peptide (Ac-GVQSLKRRRCF), illustrating the unexplained decrease in absorbance between pH 3 and 6. Absorbance values are corrected for the increase in volume.

A search for the source of this behaviour led to the titration of the buffer solution alone. This revealed to be the cause of the decrease in absorbance, which in fact was due to the buffer also absorbing at 238 nm. The absorbance profile of the buffer at a pH ranging from 2 to 12 is shown in Figure 3.7.

It then became obvious that the buffer absorbance profile had to be subtracted from every peptide titration absorbance curve. For this purpose, an analytical description of the buffer absorbance profile had to be obtained, in order to be able to subtract the buffer absorbance at any pH value. However, the profile shape did not resemble any easily describable curve, and a choice was made to adjust separate curves for different pH ranges. Each of these separate curves was adjusted using least-squares fitting, and the following analytical curves were obtained for the several pH ranges:

pH: 1.75 - 4.22	$A = 0.1413 - 0.1396/pH - 0.02317 \times pH$	Eqn. 3.2.5
pH: 4.23 - 6.02	$A = 0.1530 - 0.03653 \times pH + 1.965 \times 10^{-6} \times pH^{6}$	Eqn. 3.2.6
pH: 6.03 - 11.37	$A = 0.07509 - 14309/pH^{7}$	Eqn. 3.2.7
pH: 11.38 - 12.5	$A = 0.1101 - 0.003541 \times pH + 3.631 \times 10^{-24} \times pH^{20}$	Eqn. 3.2.8

Each curve is represented in Figure 3.7, with Eqn. 3.2.5 and Eqn. 3.2.7 represented in red and Eqn. 3.2.6 and Eqn. 3.2.8 represented in black.

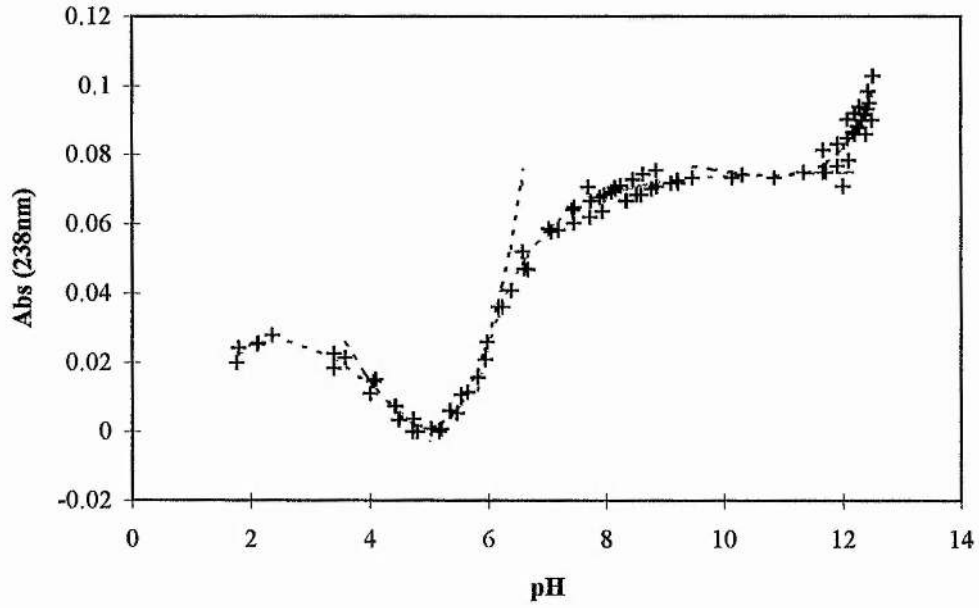


Figure 3.7 - Titration of the buffer solution. Crosses (+) represent experimental readings for absorbance corrected for increase in volume. The dashed lines represent the analytical fitting curves for various pH ranges (see text for explanation).

When the buffer absorbance profile is subtracted, the resulting curve has a substantially different aspect (cf. Figure 3.8), with the absorbance decrease at the pH range of 3 to 6 disappearing, together with a levelling of the right end of the titration curve, at high pH.

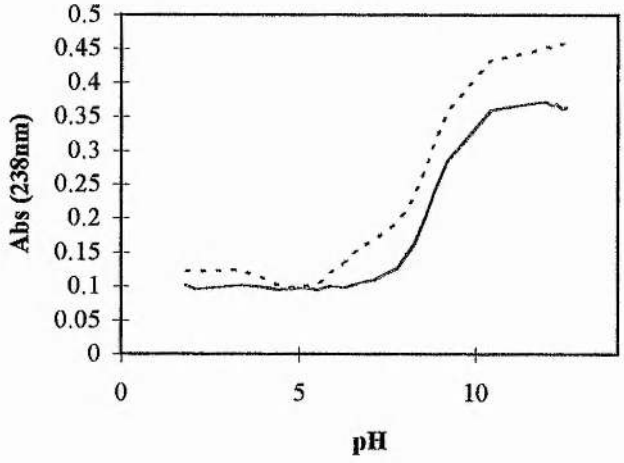


Figure 3.8 - Titration curve from Figure 3.6 before (dashed line) and after subtraction (solid line) of buffer absorbance. Notice the disappearance of the decrease in absorbance between pH 3 and 6.

The subtraction of the buffer absorbance did not always produce a perfect levelling of both ends of the titration curve, and this made it more difficult to attribute a maximum

and a minimum required for the calculations described in Method 2.10, which then had to be chosen as to produce the best fitting among the several curves obtained. This contributed in some instances to minimise the scattering of the experimental data.

3.2.3.2. The Titration of Ac-GVQSLKRRRCF

The peptide Ac-GVQSLKRRRCF was titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using the following equation:

$$\alpha_{3i} = (A_i - A_{min})/(A_{max} - A_{min}) \qquad \text{Eqn. 3.2.9}$$

in which A_i is the absorbance at 238 nm, A_{min} is the minimum absorbance of the titration curve and A_{max} the absorbance maximum. These limit values were in some instances adjusted in order to obtain a better fitting of the several curves. The function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points as described in Method 2.10, also taking into consideration the titration curves obtained in 3.2.3.3 and 3.2.3.4. The result of the fitting is shown in Figure 3.9. This fitting yielded the values $pk_{12} = 9.0$, $pk_{13} = 8.65$, $pk_{123} = 9.55$ and $pk_{132} = 9.9$ for the constants of the equilibria in Figure 3.5.

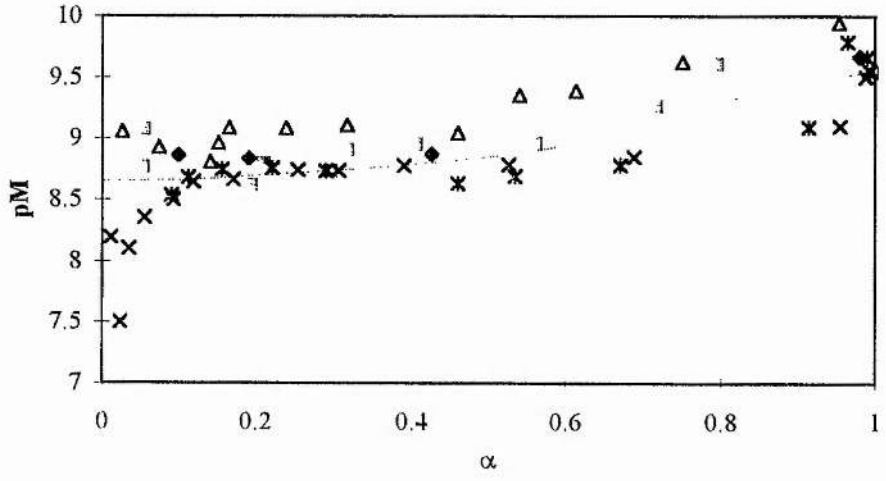


Figure 3.9 - The function pM vs. α for the peptide Ac-GVQSLKRRRCF. The different symbols represent experimental data from 5 titrations and the dashed line is the theoretical curve obtained using Eqn. 2.10.6.

Although the experimental data are quite scattered, the theoretical curve for this situation seems to describe the general trend. Using Eqn. 2.10.4, a theoretical curve

could be fitted to the experimental data for titration to check for the goodness of fitting, as shown in Figure 3.10.

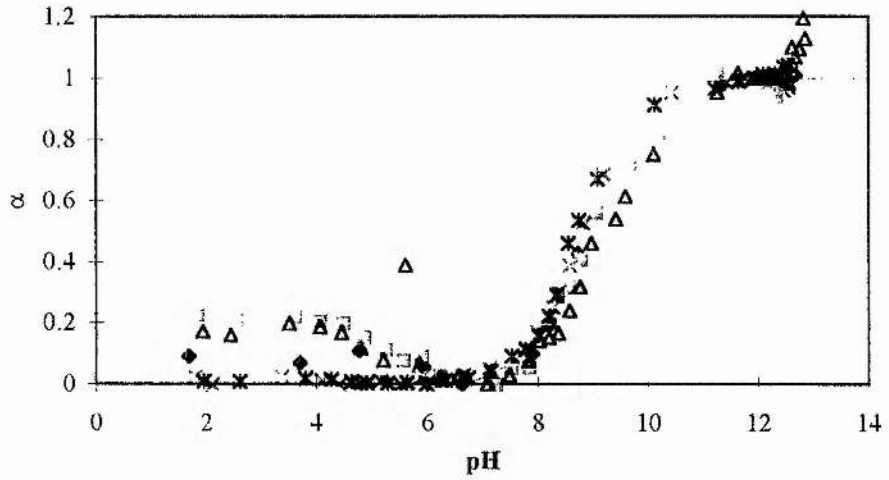


Figure 3.10 - Titration curve for the peptide Ac-GVQSLKRRRCF. The different symbols represent experimental data from 5 titrations and the dashed line is the theoretical curve obtained using Eqn. 2.10.4.

Again in the above plot the scattering of experimental data is evident, but the theoretical curve seems to describe the trend reasonably well. From the constants obtained above, the deprotonation profile for the lysine amino group can also be calculated and compared to that of the cysteine:

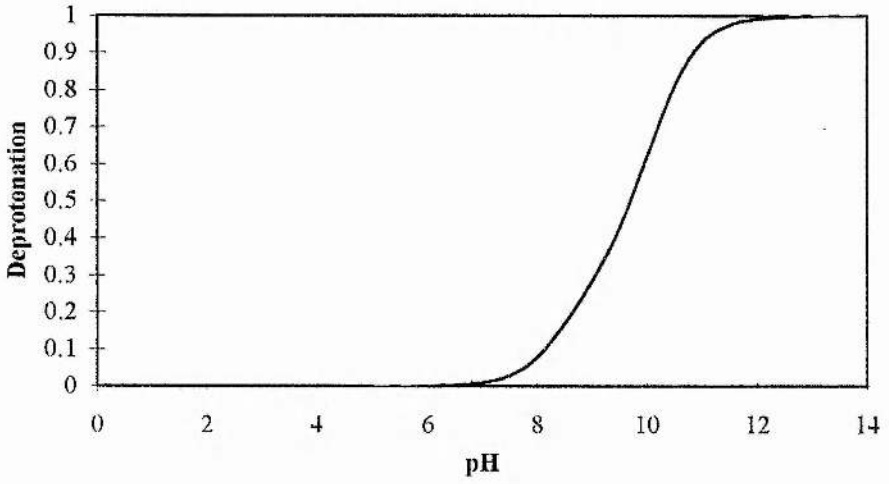


Figure 3.11 - Deprotonation profiles of the cysteine thiol (dashed line) and the lysine amino group (solid line) for the peptide Ac-GVQSLKRRRCF.

From these curves it can be observed that the cysteine thiol does not ionise at an unusually low pH, although it is somewhat less basic than the lysine amino group. In fact,

the corresponding values of pk_{12}, pk_{13}, pk_{123} and pk_{132} for cysteine, which also contains an amino group as well as the thiol group, have been determined (Benesch and Benesch, 1955), and are reproduced in the following table:

Table 3.2 - p*k* values for the peptide Ac-GVQSLKRRRCF and cysteine. Values for cysteine taken from (Benesch and Benesch, 1955).

p*k*	Ac-GVQSLKRRRCF	Cysteine
pk_{12}	9.0	8.86
pk_{13}	8.65	8.53
pk_{123}	9.55	10.03
pk_{132}	9.9	10.36

The above table shows a remarkable similarity between the values, which seems to indicate that the cysteine has an ionisation pattern similar in both compounds, that is, with a thiol group not unusually acidic, although with an amino group more basic in cysteine than in the peptide Ac-GVQSLKRRRCF.

3.2.3.3. The Titration of GVQSLARRRCF

The peptide GVQSLARRRCF was also titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using Eqn. 3.2.9. In the same manner described above (cf. 3.2.3.2), the function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points as described in Method 2.10, also taking into consideration the titration curves obtained in 3.2.3.2 and 3.2.3.4. The result of the fitting is shown in Figure 3.12. This fitting yielded the values $pk_{21} = 10.7$, $pk_{23} = 8.85$, $pk_{123} = 9.55$ and $pk_{231} = 11.4$ for the constants of the equilibria in Figure 3.5.

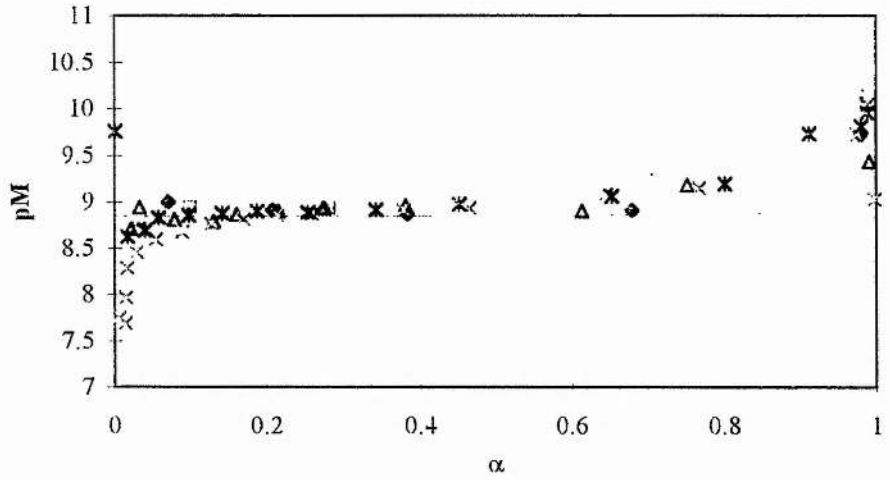


Figure 3.12 - The function pM vs. α for the peptide GVQSLARRRCF. The different symbols represent experimental data from 5 titrations and the dashed line is the theoretical curve obtained using Eqn. 2.10.6.

The above plot shows a good agreement between the experimental data and the theoretical curve for $\alpha_3 < 0.6$, although it then deviates for higher values. This is a consequence of keeping the values of the equilibrium constants for the three titrations coherent. Again, using Eqn. 2.10.4, a theoretical curve could be fitted to the experimental data for titration to check for the goodness of fitting, and this is shown in Figure 3.13.

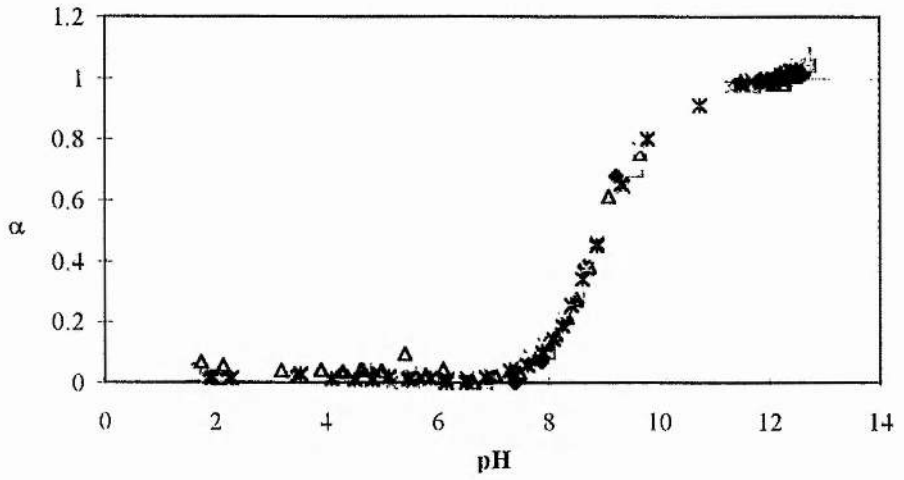


Figure 3.13 - Titration curve for the peptide GVQSLARRRCF. The different symbols represent experimental data from 5 titrations and the dashed line is the theoretical curve obtained using Eqn. 2.10.4.

In this representation, the theoretical curve does seem to describe the general trend, although the collection of more experimental points in the pH range between 10 and 11.5

could have benefited the determination of these constants. From the constants obtained above, the deprotonation profile for the N-terminal amino group can also be calculated and compared to that of the cysteine:

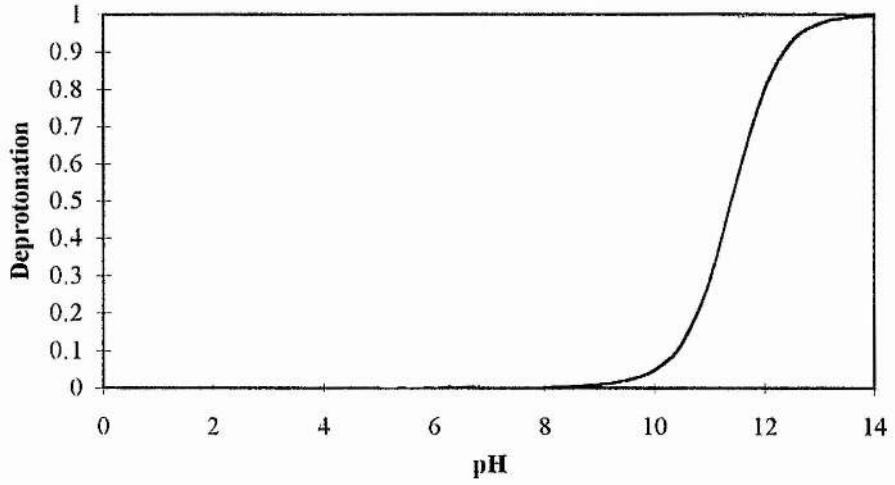


Figure 3.14 - Deprotonation profiles of the cysteine thiol (dashed line) and the N-terminal amino group (solid line) for the peptide GVQSLARRRCF.

From the curves above it can again be noted that the cysteine thiol does not ionise at an unusually low pH. Unexpectedly, the N-terminal amino group is quite basic, only starting to deprotonate when the thiol is almost completely ionised. Using the same reasoning as in 3.2.3.2, the corresponding values of pk_{21}, pk_{23}, pk_{123} and pk_{231} can also be compared to those for cysteine in the following table:

Table 3.3 - p*k* values for the peptide GVQSLARRRCF and cysteine. Values for cysteine taken from (Benesch and Benesch, 1955).

p*k*	GVQSLARRRCF	Cysteine
pk_{21}	10.7	8.86
pk_{23}	8.85	8.53
pk_{123}	9.55	10.03
pk_{231}	11.4	10.36

This time, the table shows some similarity only for the value of pk_{23}, that corresponds to the ionisation of the thiol group when the amino group is still protonated, which still shows that the thiol group in the peptide is not unusually acidic.

3.2.3.4. The Titration of GVQSLKRRRCF

The peptide GVQSLKRRRCF was also titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using Eqn. 3.2.9.

However, as this peptide has three ionisable groups, the equations developed in 3.2.3 had to be used. The function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points using Eqn. 3.2.4 as described in 3.2.3, also taking into consideration the titration curves obtained in 3.2.3.2 and 3.2.3.3. The result of the fitting is shown in Figure 3.15. This fitting yielded the remaining values $pk_1 = 7.2$, $pk_2 = 5.5$, $pk_3 = 7.15$, $pk_{31} = 8.7$ and $pk_{32} = 7.2$ for the constants of the equilibria in Figure 3.5.

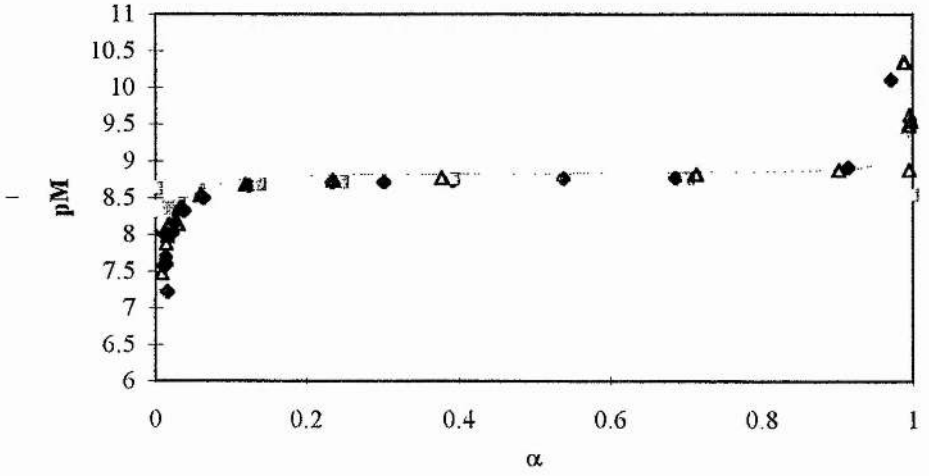


Figure 3.15 - The function pM vs. α for the peptide GVQSLKRRRCF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.4.

It can be seen from the above figure that a very good agreement between the experimental data and the theoretical curve was obtained. Therefore, the constants obtained using the method described in 3.2.3 seem to be reasonably reliable. Using Eqn 3.2.3, a theoretical curve could be fitted to the experimental data for titration to check again for the goodness of fitting, and this is shown in Figure 3.16.

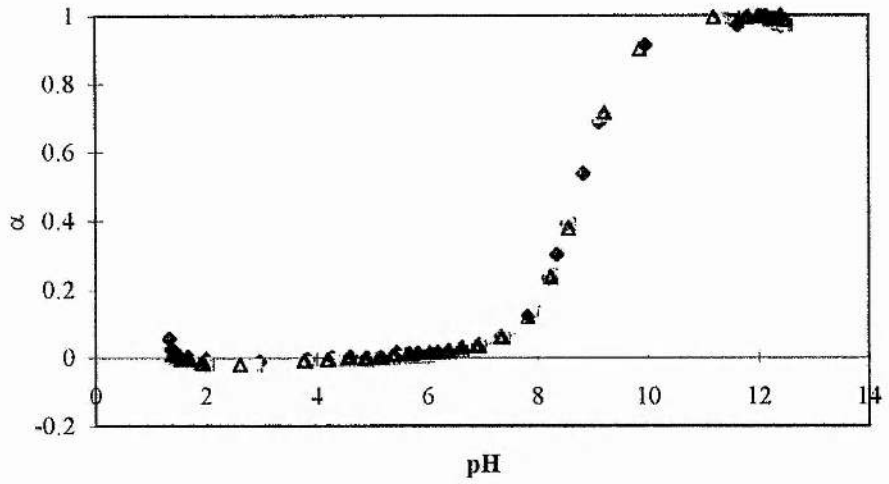


Figure 3.16 - Titration curve for the peptide GVQSLKRRRCF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.3.

The above graph shows a tight fitting of the theoretical curve to the experimental points, thus providing with another way of confirming the validity of the constants obtained. Again from these values deprotonation profiles can be calculated for the N-terminal and lysine amino groups, as well as for the cysteine thiol:

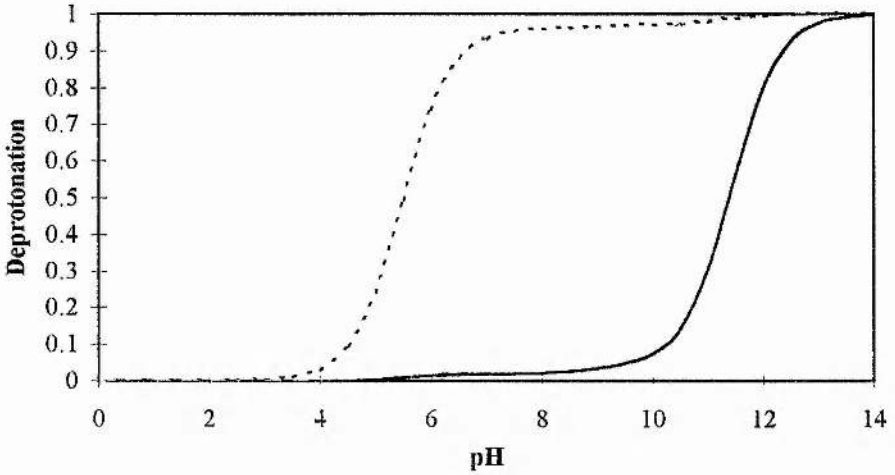


Figure 3.17 - Deprotonation profiles of the lysine amino group (---), the cysteine thiol (– – –) and the N-terminal amino group (——) for the peptide GVQSLKRRRCF.

The curves above indicate three interesting results: (a) the cysteine thiol once more does not seem to present an unusual pattern of ionisation; (b) the high basicity of the N-terminal amino group does not seem to be affected by the presence of another ionisable group, as compared to its behaviour in the peptide GVQSLARRRCF (cf.

Figure 3.14), this leading to this amino group being always protonated at physiological pH; (c) on the other hand, the lysine amino group suffers a remarkable increase in acidity when the N-terminal amino group is present, as compared to its behaviour in the peptide Ac-GVQSLKRRRCF (cf. Figure 3.11).

The whole set of constants determined for the peptide GVQSLKRRRCF is presented in the following table:

Table 3.4 - p*k* values for the peptide GVQSLKRRRCF (cf. Figure 3.5).

p*k*	GVQSLKRRRCF
pk_1	7.2
pk_2	5.5
pk_3	7.15
pk_{12}	9.0
pk_{13}	8.65
pk_{21}	10.7
pk_{23}	8.85
pk_{31}	8.7
pk_{32}	7.2
pk_{123}	9.55
pk_{132}	9.9
pk_{231}	11.4

3.2.3.5. The Titration of GVQSLKRRRCA

The determination of the equilibrium constants for this peptide had to rely on certain assumptions. As the only difference between the peptide GVQSLKRRRCF and this peptide rests on the C-terminal residue, a replacement that does not add any charge to the molecule, it was assumed that only the thiol involving constants (k_3, k_{13}, k_{23} and k_{123}, cf. Figure 3.5) would change, all the others remaining constant. By using this hypothesis, only these constants had to be recalculated, all the remaining being equal to the corresponding constants in the peptide GVQSLKRRRCF. This implied only one degree of freedom in the determination, which meant that if a value was attributed to one of the equilibrium constants, the others would be consequently defined.

Taking k_3 as the independent equilibrium constant, the other constants could be calculated using the following equations:

$$k_{13} = k_3 k_{31} / k_1 \qquad \text{Eqn. 3.2.10}$$

$$k_{23} = k_3 k_{32} / k_2 \qquad \text{Eqn. 3.2.11}$$

$$k_{123} = k_{13} k_{132} / k_{12} \qquad \text{Eqn. 3.2.12}$$

The peptide GVQSLKRRRCA was also titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using Eqn. 3.2.9. In the same manner described above (cf. 3.2.3.4), the function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points using Eqn. 3.2.4 as described in 3.2.3. The result of the fitting is shown in Figure 3.18 and the resulting constants are presented in Table 3.5.

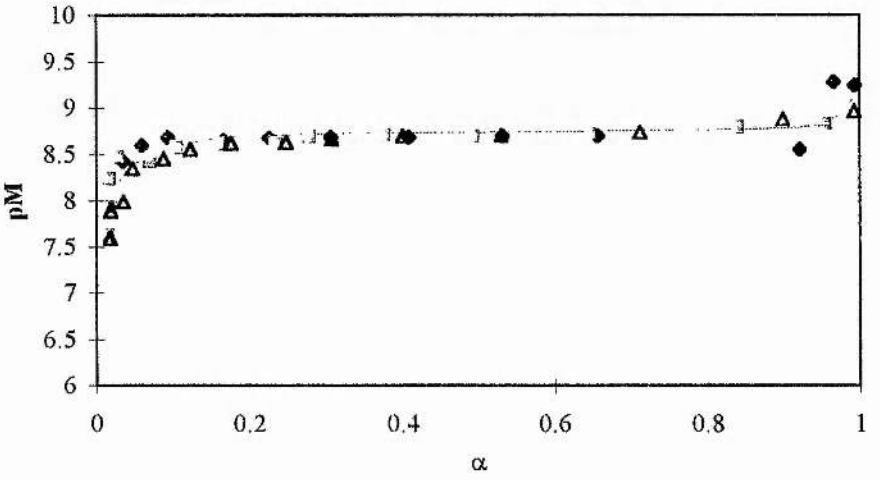


Figure 3.18 - The function pM vs. α for the peptide GVQSLKRRRCA. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.4.

A very good agreement between the experimental data and the theoretical curve is apparent from the above figure. This seems to support the assumptions made above. Using Eqn. 3.2.3, a theoretical curve could be fitted to the experimental data for titration to provide with an alternative test for the goodness of fitting, and this is shown in Figure 3.19.

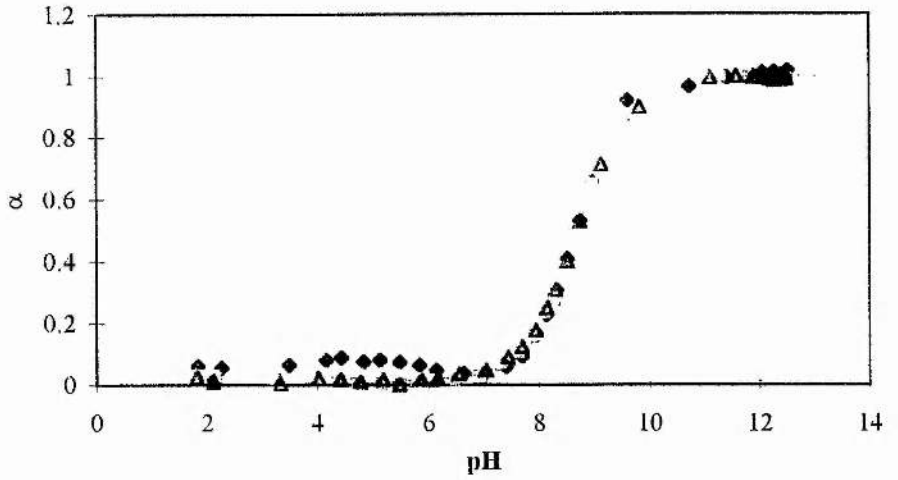


Figure 3.19 - Titration curve for the peptide GVQSLKRRRCA. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.3.

Having confirmed with the above graph a remarkable fitting of the theoretical curve to the experimental points, deprotonation profiles can again be calculated for the N-terminal and lysine amino groups, as well as for the cysteine thiol:

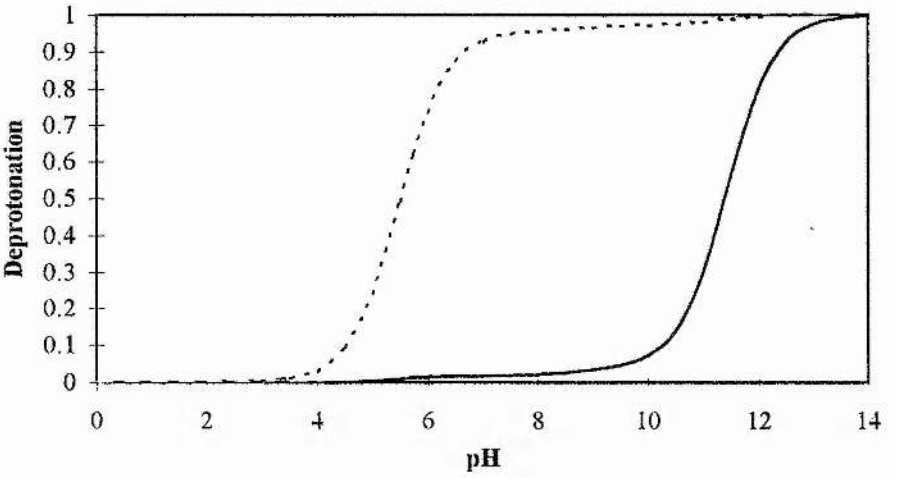


Figure 3.20 - Deprotonation profiles of the lysine amino group (---), the cysteine thiol (– – –) and the N-terminal amino group (——) for the peptide GVQSLKRRRCA.

As expected from the assumptions made above, the deprotonation curves for both the amino groups remained constant as compared to those corresponding to the peptide GVQSLKRRRCF. Even the curve corresponding to the ionisation of the cysteine thiol is only slightly shifted towards lower pH values, indicating that the substitution of the phenylalanine for an alanine in the activating peptide makes the thiol slightly less basic.

Table 3.5 - p*k* values for the peptide GVQSLKRRRCA (cf. Figure 3.5).

p*k*	GVQSLKRRRCA
pk_1	7.2
pk_2	5.5
pk_3	7.05
pk_{12}	9.0
pk_{13}	8.55
pk_{21}	10.7
pk_{23}	8.75
pk_{31}	8.7
pk_{32}	7.2
pk_{123}	9.45
pk_{132}	9.9
pk_{231}	11.4

3.2.3.6. The Titration of GVQSLKRRCRF

As with the previous peptide, also the determination of the equilibrium constants for the peptide GVQSLKRRCRF had to rely on certain assumptions. However, in this case the difference between this and the peptide GVQSLKRRRCF is a shift of one of the ionisable groups (the thiol) within the molecule, which will be expected to affect most equilibrium constants, as it has already been observed previously that the insertion of an ionisable group is liable to affect the ionisation of other groups. Therefore, it was assumed that only constants k_1, k_2, k_{12} and k_{21} would not change as compared to those of the peptide GVQSLKRRRCF, since these are the only equilibria which involve species with a non ionised thiol group (cf. Figure 3.5), that should therefore have an equivalent behaviour regardless of the relative position of the cysteine within the molecule. By using this hypothesis, eight constants had to be recalculated, with four degrees of freedom in the determination.

The choice of four independent equilibrium constants with which to attribute values was based on the observation of Eqn. 3.2.4. As mentioned before (cf. 3.2.3), when $(H^-) \rightarrow +\infty$, $\alpha_3 \rightarrow 0$ and $pM_3 \rightarrow pk_3$, and when $(H^-) \rightarrow 0$, $\alpha_3 \rightarrow 1$ and $pM_3 \rightarrow pk_{123}$, so the plot of the experimental pM values will provide with an estimate for values of k_3 and k_{123}. Thus, these two could be used as independent constants, and the two remaining necessary constants were arbitrarily chosen to be k_{23} and k_{31}.

The dependent constants were then calculated using the following equations:

$$k_{13} = k_3 \, k_{31} / k_1 \qquad \text{Eqn. 3.2.13}$$

$$k_{132} = k_{12} \, k_{123} / k_{13} \qquad \text{Eqn. 3.2.14}$$

$$k_{231} = k_{21} \, k_{123} / k_{23} \qquad \text{Eqn. 3.2.15}$$

$$k_{32} = k_{31} \, k_{132} / k_{231} \qquad \text{Eqn. 3.2.16}$$

Also the peptide GVQSLKRRCRF was titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using Eqn. 3.2.9. In the same manner described above (cf. 3.2.3.4), the function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points using Eqn. 3.2.4 as described in 3.2.3. The result of the fitting is shown in Figure 3.21 and the resulting constants are presented in Table 3.6.

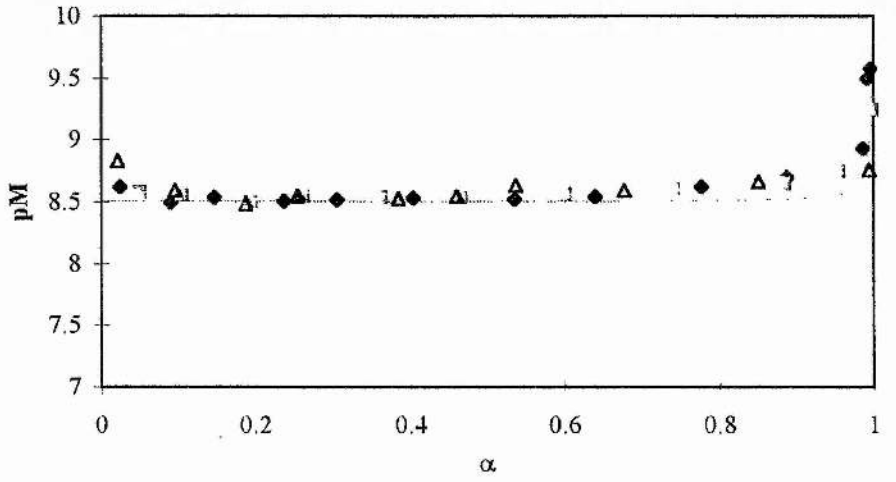


Figure 3.21 - The function pM vs. α for the peptide GVQSLKRRCRF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.4.

This figure allowed the confirmation that the experimental data and the theoretical curve are closely related. Using Eqn. 3.2.3, a theoretical curve could also be fitted to the experimental data for titration, enabling a second confirmation of the goodness of fitting, and this is shown in Figure 3.22.

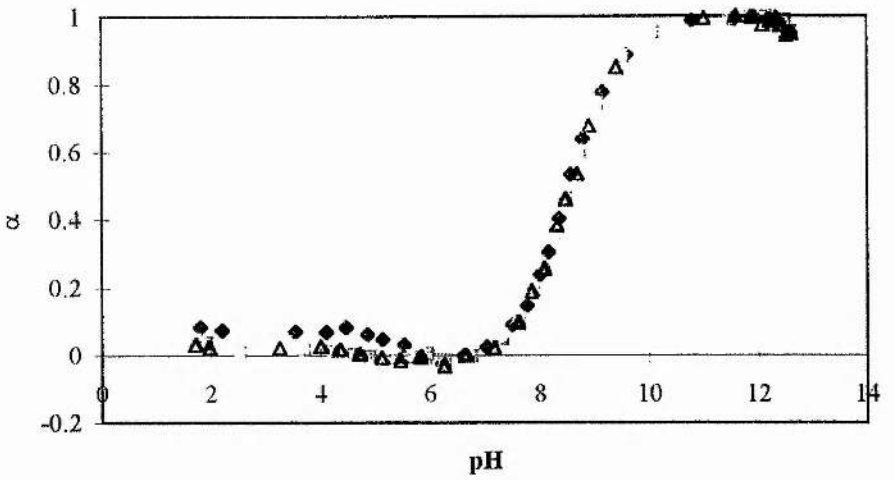


Figure 3.22 - Titration curve for the peptide GVQSLKRRCRF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.3.

With the constants obtained, deprotonation profiles could be calculated for all the ionisable groups, as depicted below:

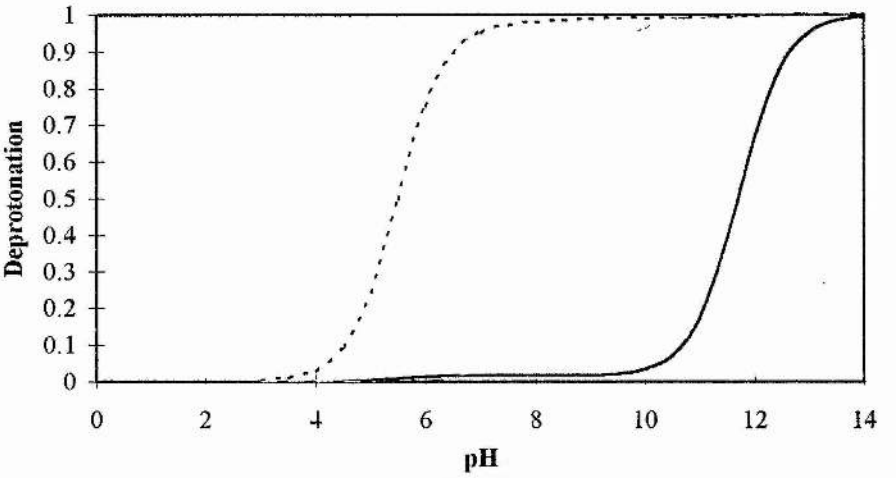


Figure 3.23 - Deprotonation profiles of the lysine amino group (---), the cysteine thiol (– – –) and the N-terminal amino group (——) for the peptide GVQSLKRRCRF.

When compared to the peptide GVQSLKRRRCF (cf. Figure 3.17), the above figure shows that, whereas the lysine amino group retains its acidity unchanged, the N-terminal amino group actually becomes more basic, albeit slightly, and the cysteine thiol becomes slightly less basic.

Table 3.6 - p*k* values for the peptide GVQSLKRRCRF (cf. Figure 3.5).

p*k*	GVQSLKRRCRF
pk_1	7.2
pk_2	5.5
pk_3	8.5
pk_{12}	9.0
pk_{13}	8.8
pk_{21}	10.7
pk_{23}	8.5
pk_{31}	7.5
pk_{32}	5.5
pk_{123}	9.5
pk_{132}	9.7
pk_{231}	11.7

3.2.3.7. The Titration of GVQSLKRCRRF

The assumptions made and method used for the calculation of the equilibrium constants for this peptide were identical to those for peptide GVQSLKRRCRF (cf. 3.2.3.6).

This peptide was also titrated as described in Method 2.10.1, the buffer absorbance was subtracted using the equations stated in 3.2.3.1 and the ionisation fraction α_3 was calculated for each experimental point using Eqn. 3.2.9. In the same manner described above (cf. 3.2.3.4), the function pM_3 was then calculated using Eqn. 2.10.6 for each value of α_3 and plotted against α_3, and a theoretical curve was fitted to the points using Eqn. 3.2.4 as described in 3.2.3. The result of the fitting is shown in Figure 3.24 and the resulting constants are presented in Table 3.7.

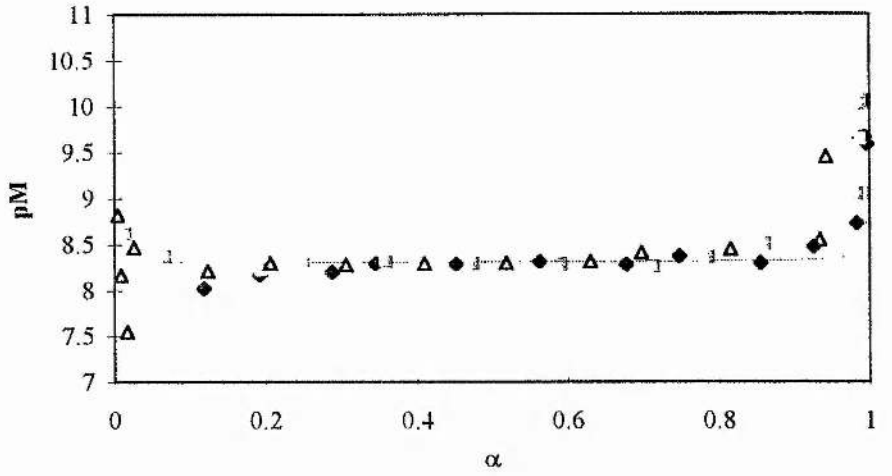


Figure 3.24 - The function pM vs. α for the peptide GVQSLKRCRRF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.4.

As with the previous peptide, again from the above figure a good agreement between the experimental data and the theoretical curve can be observed. Using Eqn. 3.2.3, a theoretical curve could be fitted to the experimental data for titration to allow for another judgement on the goodness of fitting, and this is shown in Figure 3.25.

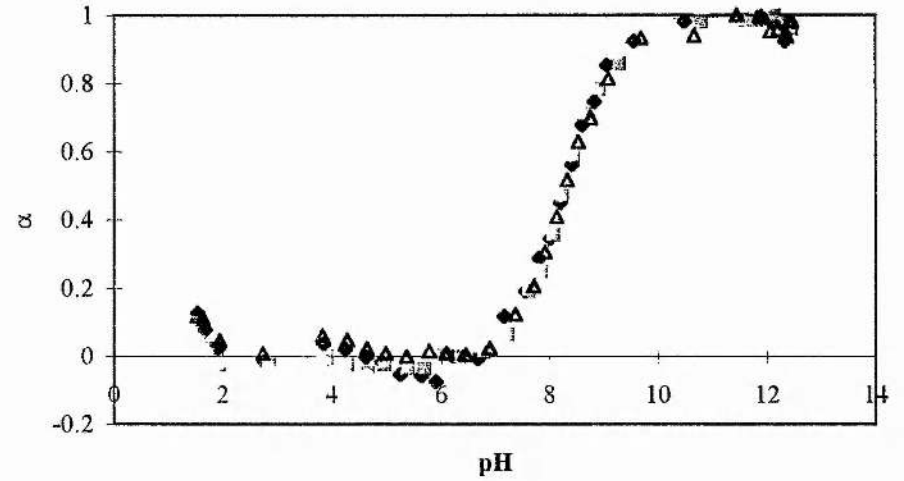


Figure 3.25 - Titration curve for the peptide GVQSLKRCRRF. The different symbols represent experimental data from 3 titrations and the dashed line is the theoretical curve obtained using Eqn. 3.2.3.

The above graph shows a good fitting of the theoretical curve to the experimental points, although the points for pH below 6 show some considerable scattering, which imposed a careful assignation of the curves minima. Again from these values

deprotonation profiles could be calculated for the N-terminal and lysine amino groups, as well as for the cysteine thiol:

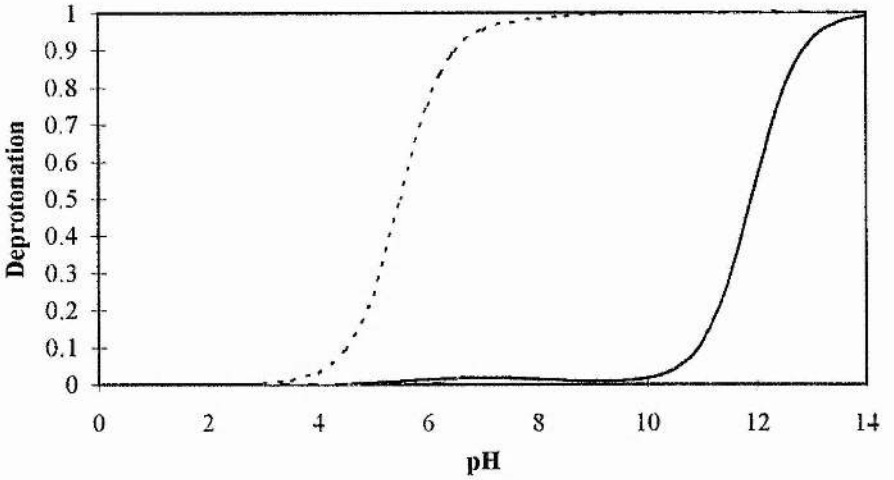


Figure 3.26 - Deprotonation profiles of the lysine amino group (---), the cysteine thiol (– – –) and the N-terminal amino group (——) for the peptide GVQSLKRCRRF.

Again in comparison to the peptide GVQSLKRRRCF (cf. Figure 3.17), the above figure shows that, whereas the lysine amino group retains its acidity unchanged, the N-terminal amino group actually becomes even more basic (increasing even in relation to the peptide GVQSLKRRCRF), and the cysteine thiol becomes slightly less basic (again to a greater extent than in the peptide GVQSLKRRCRF).

Table 3.7 - p*k* values for the peptide GVQSLKRCRRF (cf. Figure 3.5).

p*k*	GVQSLKRCRRF
pk_1	7.2
pk_2	5.5
pk_3	9.0
pk_{12}	9.0
pk_{13}	9.3
pk_{21}	10.7
pk_{23}	8.3
pk_{31}	7.5
pk_{32}	4.8
pk_{123}	9.5
pk_{132}	9.2
pk_{231}	11.9

3.2.3.8. Final Considerations on the Titration Results

The titration results provide a better understanding of the acid-base behaviour of the activating peptide. The initial assumption that the array of arginines close to the peptide would alter significantly the acidic character of the cysteine thiol group was shown not to be true. Also, although the equilibria description in Figure 3.5 provides with a more accurate account of the deprotonation equilibria, a somewhat simpler description of the deprotonation can be used, since it can be observed (cf. Figure 3.17) that for the activating peptide each ionisable group deprotonates at distinct pH ranges. Therefore, an alternative simpler description would be as follows:

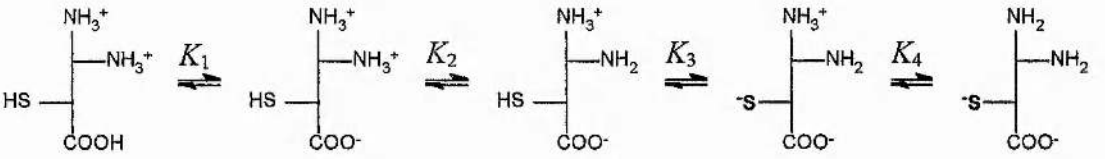


Figure 3.27 - Simplified description of the deprotonation of the peptide GVQSLKRRRCF. The species depicted represent the predominant species only, although the macroscopic constants represent deprotonation steps of the molecule, regardless of where the proton is coming from.

In the above description, the macroscopic constants can be shown to be related to the microscopic constants represented in Figure 3.5 as follows (excluding K_1, for which no determinations were made):

$$K_2 = k_1 + k_2 + k_3 \qquad \text{Eqn. 3.2.17}$$

$$K_3 = \frac{k_1 k_{12} + k_1 k_{13} + k_2 k_{23}}{k_1 + k_2 + k_3} \qquad \text{Eqn. 3.2.18}$$

$$K_4^{-1} = k_{123}^{-1} + k_{132}^{-1} + k_{231}^{-1} \qquad \text{Eqn. 3.2.19}$$

Taking into consideration the values in Table 3.4, it is a straightforward process to show that $K_2 \approx k_2$, $K_3 \approx k_{23}$ and $K_4 \approx k_{231}$, as it would be expected if the simplification stated above was true. This enables the assignation of a macroscopic pK to each of the ionisable groups within the peptide (with the exception of the carboxyl group):

Table 3.8 - Macroscopic p*K* values of the ionisable groups of the peptide GVQSLKRRRCF, with exception of the C-terminal carboxyl group.

Group	p*K*
Lysine-6 amine	5.5
Cysteine-10 thiol	8.85
N-terminal amine	11.4

Therefore, a discussion in terms of "independent" ionisable groups is possible for the effect of mutations of residues within the activating peptide.

Reanalysing the data for the peptide GVQSLKRRRCA, it can be seen that this mutation leads to a decrease of 0.1 pH units in the p*K* of the thiol group in relation to the peptide GVQSLKRRRCF, while keeping the other p*K* values constant (cf. Figure 3.28). This seems to indicate that the phenylalanine does not play any relevant role on the acidity of the cysteine thiol group. The fact that the other p*K* values remain constant is a reflection of the postulate used for the determination of the constants.

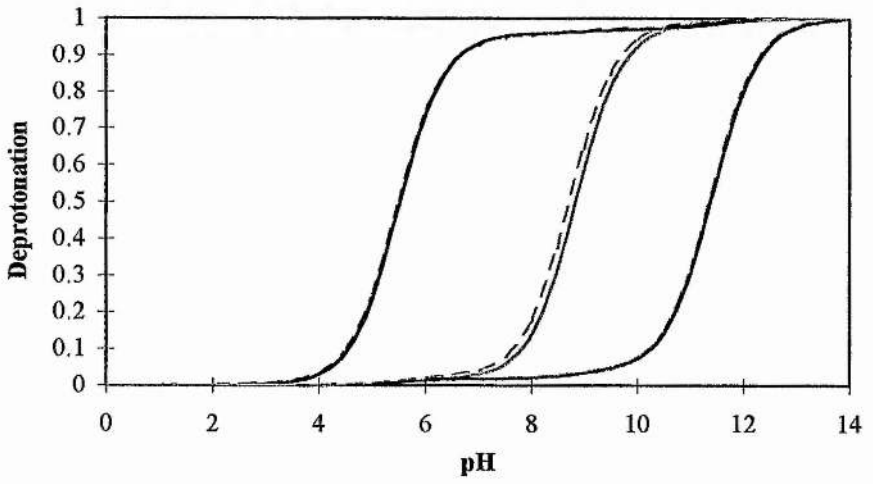


Figure 3.28 - Comparison of the deprotonation curves of the peptides GVQSLKRRRCF (solid lines) and GVQSLKRRRCA (dashed lines). The black lines correspond to the lysine amine, the purple to the cysteine thiol and the dark blue to the N-terminal amine.

When the cysteine is shifted along the arginines, there is a reasonable increase in acidity of the thiol group, with a decrease of 0.35 pH units for the peptide GVQSLKRRCRF, and of 0.55 pH units for GVQSLKRCRRF (cf. Figure 3.29), consistent with the aforementioned observation that charged neighbouring groups affect the acidity of the cysteine group, although this result does indicate that the cysteine in the activating peptide is actually reasonably basic rather than unusually acidic, which is

exactly the opposite effect of what was expected to be observed if the arginines had the reactivity enhancement initially sought for.

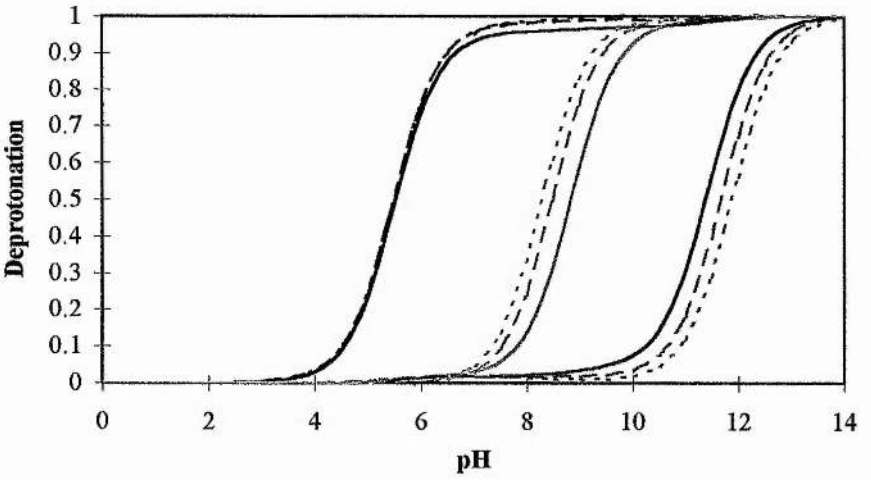


Figure 3.29 - Comparison of the deprotonation curves of the peptides GVQSLKRRRCF (——), GVQSLKRRCRF (– – –) and GVQSLKRCRRF (---). The black lines correspond to the lysine amine, the purple to the cysteine thiol and the dark blue to the N-terminal amine.

Again, the fact that the lysine amine has a constant pK is a result of the constraints imposed for the calculation of the equilibrium constants, although it makes sense that a group that is completely deprotonated before any other group (other than the C-terminal carboxyl) in the molecule will not be affected by small changes of pK of groups that only deprotonate at a reasonably higher pH.

Finally, although the array of arginines close to the cysteine appears not to affect its pK, it is plausible that these could be responsible for the unusually very low pK of the lysine amino group. Also, the very high pK of the N-terminal amino group falls within the range of the usual pK values for arginine groups, which implies that the initial assumption that these would only deprotonate after the complete deprotonation of the amines and thiol cannot be applied, and therefore the equilibria depicted in Figure 3.5 should be much more complex, not enabling the mathematical treatment used herein. Therefore, the only reliable pK values determined are those for the cysteine, as there is experimental evidence for these.

3.3. *Influence of Activating Peptide Length on Activation*

The peptide KRRRCF, which corresponds to the last six residues of the full length activating peptide, was found to be unable to activate the adenovirus protease (Webster *et al.*, 1993). This observation suggested that there had to be some role of the peptide

length in activation. Therefore, several peptides were synthesised in order to evaluate what was the minimum length requirement for adequate activation of the protease.

3.3.1. Importance of GV in Activation

Four peptides were synthesised, LKRRRCF, SLKRRRCF, QSLKRRRCF and VQSLKRRRCF as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2; KRRRCF had been previously synthesised. All peptides were purified by HPLC (Method 2.4.2.1) and as a reference typical percentages of acetonitrile for elution of the peptides (4.6 mm × 25 cm C18 column), as well as migration times on capillary electrophoresis (CE) are shown in Table 3.9. The peptide molecular weights were confirmed by mass spectrometry (Method 2.6) and their sequence was confirmed by Method 2.5.

Table 3.9 - Typical values of purification parameters for several peptides. Migration times for CE are given for a 24 cm × 25 µm capillary.

Peptide	Percentage of acetonitrile for peptide elution	Migration time on CE (min)
VQSLKRRRCF	40-50	3.4
QSLKRRRCF	45-50	3.3
SLKRRRCF	45-50	3.1
LKRRRCF	45-50	3.1
KRRRCF	40-45	2.9

The first approach to the study of these peptides was to perform concentration assays, in the same manner as with the peptides with substitutions in the cysteine. The protease (0.02 mg/ml) was therefore incubated with each of the peptides at 37°C (Method 2.9.1) for 10 min before the addition of substrate. The substrate LSGAGFSW was then allowed to be digested for 30 min at several concentrations (0.15, 0.12, 0.089, 0.059, 0.030 and 0.015 mM), of each of the purified activating peptides. Time samples were taken and analysed using capillary electrophoresis (Method 2.7.1.1). The results are shown in Figure 3.30:

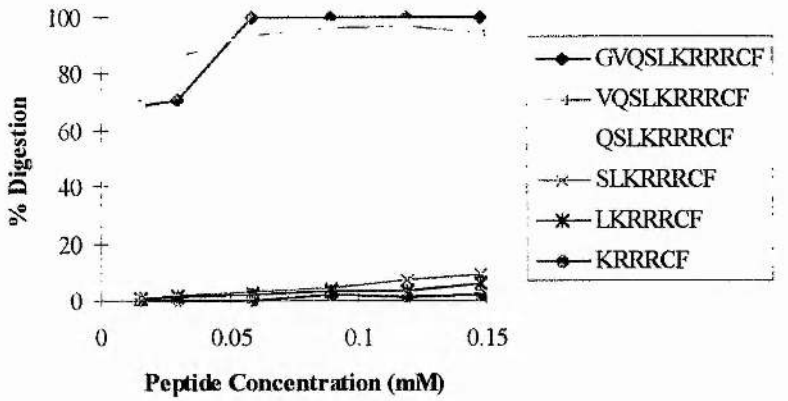


Figure 3.30 - Effect of activating peptide length on activation of the adenovirus protease. Curves correspond to a 30 min digestion of the substrate LSGAGFSW. Purified activating peptides.

These preliminary results indicated that the first two N-terminal residues, glycine and valine, had a definite role in the activation process of the protease. The following step was to quantify the difference between the activation of these peptides. For that purpose, kinetic assays were made in order to obtain values for the initial rates of the digestion reaction.

3.3.2. Determination of the Initial Rates of the N-terminal Truncated Peptides

The kinetic assays consisted on following the digestion curve of LSGAGFSW with time (as described in Method 2.12.1 and discussed in 3.8.3) when the protease was activated with each of the N-terminal truncated peptides and comparing it to the activation with the original activating peptide.

The determination of the initial rates of digestion demanded the use of a curve fitting strategy (Method 2.12) due to the limitation of the method used to determine the time points of digestion (capillary electrophoresis, Method 2.7.1.1), which did not allow to quantify accurately low amounts of digestion. Several initial rate determinations were made for each peptide activation in order to obtain statistically significant values for the final values.

3.3.2.1. The Peptide VQSLKRRRCF

Kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was

activated in duplicate by the peptide VQSLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.10 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide VQSLKRRRCF (IR_{10}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 5 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{10} (%/min)	r^2_{10}	IR_{10}/IR_{11}
0.037	0.01128	0.99984	0.00242	0.97560	0.215
0.037	0.00829	0.93833	0.00200	0.94571	0.242
0.15	0.07947	0.98941	0.02097	0.94202	0.264
0.15	0.11863	0.99577	0.04060	0.98968	0.342
0.037	0.01128	0.99984	0.00387	0.98800	0.343
0.037	0.02599	0.98825	0.00960	0.98294	0.369
0.037	0.01910	0.99029	0.00714	0.96683	0.374
0.037	0.00829	0.93833	0.00318	0.99818	0.384
0.15	0.11863	0.99577	0.04957	0.98323	0.418
0.15	0.07947	0.98941	0.03341	0.97852	0.420
0.037	0.01599	0.96870	0.00684	0.95220	0.428
0.037	0.01599	0.96870	0.00764	0.98279	0.478
0.037	0.01319	0.98735	0.00633	0.99536	0.480
0.15	0.08540	0.99441	0.04109	0.97740	0.481
0.15	0.08540	0.99441	0.04427	0.97429	0.518
0.037	0.01319	0.98735	0.00686	0.98690	0.521
0.037	0.01910	0.99029	0.01037	0.98324	0.543
0.037	0.03418	0.99591	0.02026	0.99812	0.593
0.15	0.07658	0.97816	0.04647	0.95889	0.607
0.037	0.03418	0.99591	0.02211	0.99484	0.647
0.037	0.02599	0.98825	0.01814	0.99250	0.698
0.15	0.09830	0.99466	0.07234	0.97038	0.736
0.15	0.09830	0.99466	0.07595	0.97374	0.773
0.15	0.07658	0.97816	0.06285	0.97380	0.821

The first remark to be made regards the notorious variability found in the initial rate for the activated protease with any of the peptides. This is certainly due to the instability of the protease regarding repeated freeze-thawing whilst being reused (cf. 3.1). Also, the correlation coefficient is always greater than 0.90, indicating a close agreement between

the experimental points and the theoretical curve. However, the final relative initial rate of digestion is reasonably constant, regardless of the concentration of peptide used for activation, allowing for a statistical treatment of the data. An histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated:

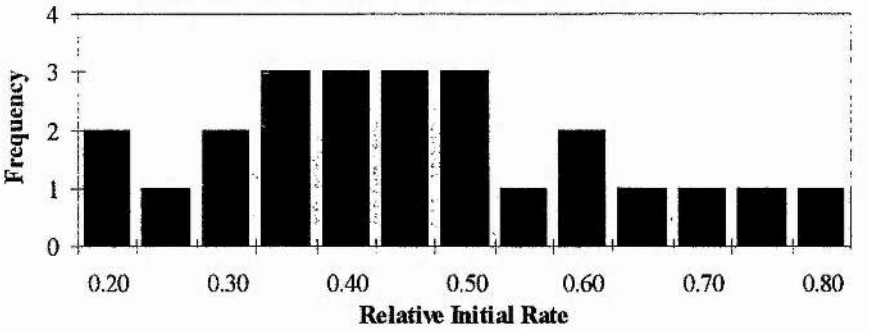


Figure 3.31 - Histogram depicting the distribution of the relative initial rate IR_{10}/IR_{11}.

From the histogram, an average can be estimated to lie between 0.40-0.45. These values led to a final determination of an average value of 0.49 ± 0.17 (standard deviation) for the relative initial rate of protease cleavage when activated with the peptide VQSLKRRRCF as compared to the peptide GVQSLKRRRCF. Therefore, this peptide is still able to activate the protease, albeit only with half of the efficiency of the original activating peptide, which suggests a role of the N-terminal glycine in the activation mechanism.

3.3.2.2. The Peptide QSLKRRRCF

Again, kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 or 0.1 mg/ml) was activated in duplicate by the peptide QSLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.11 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide QSLKRRRCF (IR_9) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_9 (%/min)	r^2_9	IR_9/IR_{11}
0.15	0.10074	0.99192	0.00169	0.96602	0.017
0.15	0.12871	0.97184	0.00263	0.95126	0.020
0.15	0.06186	0.97466	0.00155	0.91595	0.025
0.15	0.06186	0.97466	0.00157	-0.25010	0.025
0.15	0.12871	0.97184	0.00342	0.99171	0.027
0.15	0.10074	0.99192	0.00272	0.94370	0.027
0.15	0.11909	0.98792	0.00355	0.98022	0.030
0.15	0.14347	0.98552	0.00568	0.97872	0.040
0.15	0.14347	0.98552	0.00617	0.99012	0.043
0.15	0.10656	0.99264	0.00474	0.93982	0.045
0.15	0.11408	0.99271	0.00528	0.95972	0.046
0.15	0.11909	0.98792	0.00591	0.95957	0.050
0.15	0.11408	0.99271	0.00661	0.97445	0.058
0.15	0.10656	0.99264	0.00659	0.97167	0.062
0.037	0.01500	0.98818	0.00124	0.99569	0.083
0.037	0.00812	0.96384	0.00073	0.98983	0.090
0.18	0.03496	0.99615	0.00378	0.99042	0.108
0.18	0.01057	0.98228	0.00142	0.99933	0.134*
0.037	0.01500	0.98818	0.00206	0.99134	0.137*
0.18	0.03496	0.99615	0.00477	0.96606	0.137*
0.037	0.01589	0.98845	0.00252	0.98884	0.158*
0.037	0.01069	0.99482	0.00184	0.96240	0.172*
0.18	0.01029	0.99317	0.00184	0.97711	0.179*
0.037	0.00770	0.96415	0.00141	0.97266	0.183*
0.037	0.00812	0.96384	0.00150	0.99824	0.184*
0.18	0.01029	0.99317	0.00214	0.92229	0.208*
0.037	0.01069	0.99482	0.00227	0.99614	0.212*
0.18	0.01057	0.98228	0.00272	0.98001	0.258*
0.037	0.00770	0.96415	0.00215	0.94425	0.280*
0.037	0.01589	0.98845	0.00458	0.97647	0.288*

*- Discarded values in the average calculation.

Once more in these determinations a remarkable variability is found in the initial rate for the activated protease, as well as a value for the correlation coefficient constantly above 0.90, with one exception. However, in this case the final relative initial rate of digestion is not very constant. The causes leading to this occurrence can be attributed to the method used for determination of the values for the percentage of digestion. A first

point to note is that the peptide QSLKRRRCF was not expected to activate very efficiently the protease, as revealed by the preliminary studies (cf. 3.3.1). On the other hand, initial digestion rates of less than 0.005 % digested substrate/min arise from digestion points below 10 % of substrate digestion, which is very near to the limit of detection of the method used (cf. discussion in 3.8.3). Therefore, the results for low initial rates were misleading, and a more concentrated protease was used in order to obtain more reliable results. An histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated, depicting the scattering of higher relative initial rates due to loss of sensitivity of the method (corresponding in the quotient IR_9/IR_{11} to a variable value of IR_{11} with a reasonable constant low value of IR_9):

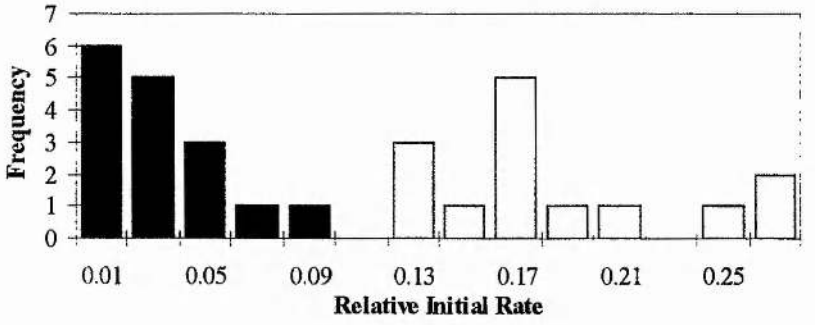


Figure 3.32 - Histogram depicting the distribution of the relative initial rate IR_9/IR_{11}. The void bars correspond to values discarded in the average calculation.

From the histogram, an average can then be estimated to lie between 0.03-0.05, taking into consideration the above arguments for discarding the higher values. This led to a final determination of an average value of 0.047 ± 0.027 for the relative initial rate of protease cleavage when activated with the peptide QSLKRRRCF as compared to the peptide GVQSLKRRRCF. This value should be considered rather as a maximum, consistently with the low initial rates obtained when the protease was activated with the peptide QSLKRRRCF, which might still lie within the detection limits of the method used for these determinations. These results seem to indicate that the presence of the residues glycine and valine in the N-terminal are determinant in the activation process of the protease.

3.3.2.3. The Peptide SLKRRRCF

Once again, kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06/0.1 mg/ml) was activated in duplicate by the peptide SLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.12 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide SLKRRRCF (IR_8) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_8 (%/min)	r^2_8	IR_8/IR_{11}
0.15	0.12815	0.99627	0.00303	0.94653	0.024
0.15	0.07025	0.99804	0.00193	0.97575	0.027
0.15	0.14925	0.97736	0.00427	0.96935	0.029
0.15	0.12815	0.99627	0.00392	0.97551	0.031
0.15	0.12264	0.99838	0.00390	0.82152	0.032
0.15	0.07025	0.99804	0.00224	0.94382	0.032
0.15	0.15668	0.99644	0.00695	0.96895	0.044
0.15	0.06370	0.86419	0.00285	0.98054	0.045
0.15	0.06611	0.55708	0.00307	0.95940	0.046
0.15	0.15668	0.99644	0.00747	0.98019	0.048
0.15	0.12264	0.99838	0.00595	0.98364	0.049
0.15	0.06611	0.55708	0.00325	0.96606	0.049
0.15	0.14925	0.97736	0.00762	0.92186	0.051
0.15	0.06421	0.99929	0.00325	0.95514	0.051
0.15	0.06421	0.99929	0.00372	0.99955	0.058
0.15	0.54545	1	0.04900	0.99865	0.090*
0.15	0.54545	1	0.05252	0.99194	0.096*
0.15	0.06370	0.86419	0.00710	0.98722	0.111*
0.18	0.02936	0.99966	0.00437	0.99047	0.149*
0.18	0.02936	0.99966	0.00633	0.99856	0.216*
0.18	0.02038	0.98904	0.00443	0.97525	0.218*
0.18	0.02164	0.98993	0.00518	0.95017	0.239*
0.18	0.02038	0.98904	0.00695	0.98986	0.341*
0.18	0.02164	0.98993	0.00817	0.98640	0.378*

*- Discarded values in the average calculation.

As with the previous peptides, the initial rate for the activated protease is found to be quite variable, as well as a value for the correlation coefficient lying steadily above 0.90, again with one exception. In one case, this value reaches 1 due to the fact that only one point was used to determine the adjustment curve, as the protease was extremely active, and complete digestion of substrate can not be used in the calculation due to having to compute a logarithm of zero. As with the peptide QSLKRRRCF, the final relative initial rate of digestion is not very constant, the causes leading to this occurrence being similar to the ones pointed out for that peptide. Therefore, again the results for low initial rates are misleading, and a more concentrated protease was used in order to obtain more reliable results. The following histogram depicting the distribution of the relative initial rate illustrates the average value to be calculated, depicting the scattering of higher relative initial rates due to loss of sensitivity of the method (once more corresponding in the quotient IR_8/IR_{11} to a variable value of IR_{11} with a reasonably constant low value of IR_8):

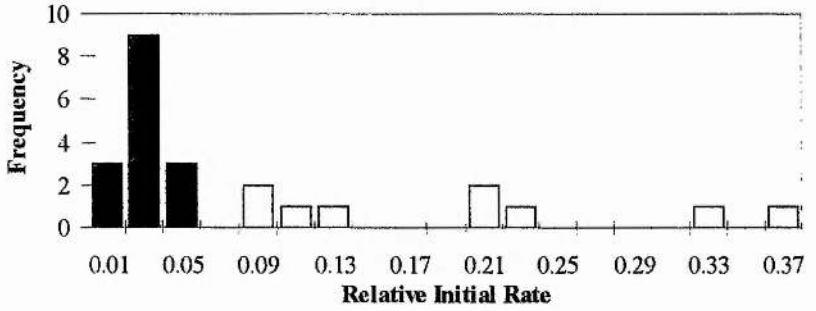


Figure 3.33 - Histogram depicting the distribution of the relative initial rate IR_8/IR_{11}. The void bars correspond to values discarded in the average calculation.

From the histogram, an average can then be estimated to lie between 0.03-0.05, taking into consideration the above arguments for discarding the higher values. This led to a final determination of an average value of 0.041 ± 0.011 for the relative initial rate of protease cleavage when activated with the peptide SLKRRRCF as compared to the peptide GVQSLKRRRCF. As with the peptide QSLKRRRCF, the value for the relative initial rate should be considered rather as a maximum, consistently with the low initial rates obtained when the protease was activated with the peptide SLKRRRCF, which again might still lie within the detection limits of the method used for these determinations. What these results suggest is that any shorter peptide, lacking the N-

terminal glycine and valine, will not be able to activate the protease, which had already been shown to be true in the case of the peptide KRRRCF (Webster *et al.*, 1993).

3.3.3. Non-Complementarity of GVQSL and KRRRCF

From the previous results, it became apparent that the residues glycine-1 and valine-2 were determinant for the activation of the protease with the peptide GVQSLKRRRCF. In order to try and elucidate whether the peptide molecule acted as a whole or in a concerted manner, with more than one molecule being necessary to fully activate the protease, some assays were performed using a mixture of the peptides GVQSL and KRRRCF.

The peptide GVQSL was synthesised as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2; KRRRCF had been previously synthesised. It was then purified by HPLC (Method 2.4.2.1) and typical percentages of acetonitrile for elution of GVQSL were of 40-45% (4.6 mm × 25 cm C18 column), and its migration time on capillary electrophoresis (CE) was of 7.1 min (24 cm × 25 μm capillary).

Activity assays were done essentially as described in Method 2.9.1, and the results are summarised in the following table:

Table 3.13 - Protease activation with GVQSL and KRRRCF (30 min digestion).

Peptide	Concentration (mM)	% digestion
GVQSLKRRRCF	0.074	100
GVQSL	0.074	0
KRRRCF	0.074	0
GVQSL + KRRRCF	0.037	0

These results show that the peptide GVQSLKRRRCF acts as a whole in the activation mechanism, which suggests that the N-terminal region of the peptide has some concomitant role in activation.

3.4. Importance of N-terminal region of Activating Peptide

The results shown in 3.3.2 stress the importance of the N-terminal region of the activating peptide GVQSLKRRRCF in the activation of the protease, namely the role of the two first residues, glycine-1 and valine-2. The next step was to assess the specificity required for these residues in the activation mechanism. Again some peptides were synthesised, with some alterations in those residues: valine was mutated to alanine, a side chain similar in hydrophobicity although not branched; and to threonine, a side chain

similar in structure to valine but hydrophilic. Glycine was also acetylated in the N-terminal in order to eliminate the positive charge. The ability of these peptides in activating the protease was then evaluated.

3.4.1. Determination of the Initial Rates of the N-terminal Mutated Peptides

Three peptides were synthesised, GAQSLKRRRCF, GTQSLKRRRCF and Ac-GVQSLKRRRCF, as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2. All peptides were purified by HPLC (Method 2.4.2.1) and as a reference typical percentages of acetonitrile for elution of the peptides (4.6 mm × 25 cm C18 column), as well as migration times on capillary electrophoresis (CE) are shown in Table 3.14. The peptide molecular weights were confirmed by mass spectrometry (Method 2.6) and their sequence was confirmed by Method 2.5, with the exception of the peptide Ac-GVQSLKRRRCF, which could not be sequenced due to being N-terminal blocked.

Table 3.14 - Typical values of purification parameters for the peptides GAQSLKRRRCF, GTQSLKRRRCF and Ac-GVQSLKRRRCF. Migration times for CE are given for a 24 cm × 25 μm capillary.

Peptide	Percentage of acetonitrile for peptide elution	Migration time on CE (min)
GAQSLKRRRCF	40-45	3.7
GTQSLKRRRCF	40-45	3.6
Ac-GVQSLKRRRCF	40-45	4.3

The kinetic assays consisted on following the digestion curve of LSGAGFSW with time (as described in Method 2.12.1 and discussed in 3.8.3) when the protease was activated with each of the N-terminal truncated peptides and comparing it to the activation with the original activating peptide.

As stated before, the determination of the initial rates of digestion demanded the use of a curve fitting strategy (Method 2.12) due to the limitation of the method used to determine the time points of digestion (capillary electrophoresis, Method 2.7.1.1), which did not allow to quantify accurately little amounts of digestion (cf. 3.8.3). Again, several initial rate determinations were made for each peptide activation in order to obtain statistically significant values for the final values.

3.4.1.1. The Peptide GAQSLKRRRCF

Once more, kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GAQSLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.15 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide GAQSLKRRRCF (IR_{2A}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{2A} (%/min)	r^2_{2A}	IR_{2A}/IR_{11}
0.15	0.01018	0.97545	0.00120	0.97502	0.118
0.18	0.01066	0.95073	0.00181	0.98667	0.170
0.18	0.03440	0.98847	0.00691	0.96337	0.201
0.15	0.05352	0.99646	0.01233	0.98106	0.230
0.18	0.03440	0.98847	0.00857	0.96912	0.249
0.15	0.01043	0.99610	0.00298	0.98372	0.286
0.15	0.02090	0.96148	0.00653	0.94099	0.312
0.15	0.00791	0.98073	0.00259	0.96028	0.328
0.15	0.02014	0.96412	0.00726	0.93341	0.360
0.15	0.03049	0.99430	0.01109	0.94534	0.364
0.15	0.02014	0.96412	0.00770	0.99001	0.382
0.15	0.00791	0.98073	0.00305	0.96836	0.386
0.18	0.02228	0.98468	0.00928	0.99328	0.416
0.15	0.05352	0.99646	0.02329	0.99634	0.435
0.18	0.02228	0.98468	0.00984	0.99372	0.442
0.15	0.00906	0.95198	0.00435	0.97731	0.481
0.15	0.01043	0.99610	0.00505	0.95070	0.485
0.15	0.04747	0.98128	0.02517	0.99562	0.530
0.15	0.03305	0.98649	0.01757	0.99791	0.532
0.18	0.01066	0.95073	0.00586	0.99692	0.550
0.15	0.03305	0.98649	0.01866	0.99820	0.564
0.15	0.00906	0.95198	0.00536	0.96646	0.592
0.15	0.04747	0.98128	0.02873	0.98135	0.605

As with the previous assays, the initial rate for the activated protease with any of the peptides varies to a noticeable extent, again due to the instability of the protease

regarding repeated freeze-thawing whilst being reused. Also, the correlation coefficient is always greater than 0.90, indicating a close agreement between the experimental points and the theoretical curve. However, the relative initial rate of digestion obtained after the calculations is reasonably constant within certain limits, regardless of the concentration of peptide used for activation, allowing for a statistical treatment of the data. An histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated:

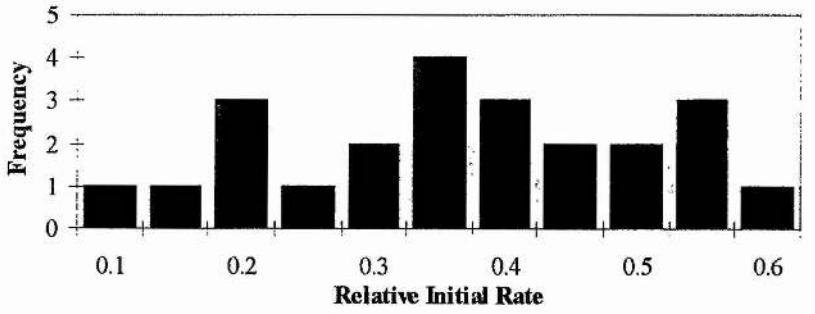


Figure 3.34 - Histogram depicting the distribution of the relative initial rate IR_{2A}/IR_{11}.

From the above histogram, an average can be estimated to lie between 0.30-0.40. For the determination of the average, no values were discarded as all the data seem to lie within the apparent pattern for a normal distribution, albeit quite flat and scattered. This led to a final determination of an average value of 0.39 ± 0.14 for the relative initial rate of protease cleavage when activated with the peptide GAQSLKRRRCF as compared to the peptide GVQSLKRRRCF. This value seems to indicate that a branched side chain in the second residue of the activating peptide is a requirement for a more effective activation of the protease.

3.4.1.2. The Peptide GTQSLKRRRCF

As with the previous peptides, kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GTQSLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.16 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide GTQSLKRRRCF (IR_{2T}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{1I}). The first column indicates the concentration of peptide used in the assay. the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1. and the last column indicates the relative initial rate.

Conc. (mM)	IR_{1I} (%/min)	r^2_{1I}	IR_{2T} (%/min)	r^2_{2T}	IR_{2T}/IR_{1I}
0.15	0.01133	0.98164	0.00291	0.95609	0.257
0.15	0.01011	0.98184	0.00277	0.97950	0.274
0.15	0.00593	0.98647	0.00174	0.99153	0.294
0.15	0.00373	0.97594	0.00118	0.97975	0.318
0.15	0.01133	0.98164	0.00439	0.96493	0.387
0.15	0.01989	0.99706	0.00781	0.95833	0.393
0.15	0.01011	0.98184	0.00433	0.98233	0.428
0.15	0.01989	0.99706	0.00883	0.96019	0.444
0.18	0.02445	0.99339	0.01289	0.99101	0.527
0.18	0.02445	0.99339	0.01411	0.99162	0.577
0.15	0.00806	0.98981	0.00654	0.97038	0.811*

*- Discarded value in the average calculation.

It can again be observed that the initial rate for the activated protease with any of the peptides is varying, for the same reasons already stated. Also, the correlation coefficient is always greater than 0.95, indicating an excellent agreement between the experimental points and the theoretical curve. The relative initial rate of digestion values obtained after the calculations are much less variable, regardless of the concentration of peptide used for activation, allowing for a statistical treatment of the data. Only one point was found to lie reasonably distant from the others, which led to its discarding ($\alpha = 0.1$*, criteria of Dixon (Dixon and Massey, 1957)). An histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated:

* α represents the probability of the discarded point belonging to the average and not being due to some error associated with its experimental determination (Dixon and Massey, 1957).

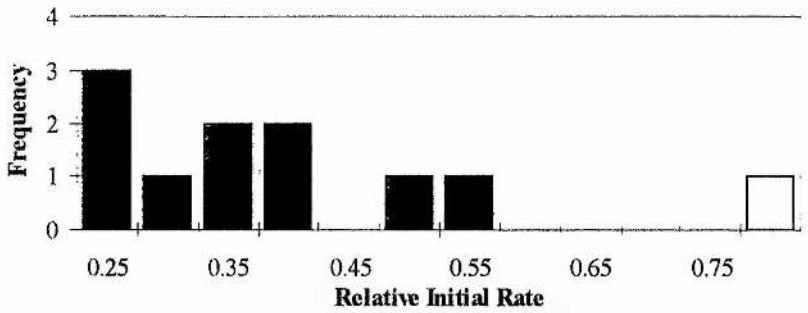


Figure 3.35 - Histogram depicting the distribution of the relative initial rate IR_{2T}/IR_{1T}. The void bar corresponds to the value discarded in the average calculation.

It is not completely clear from the above histogram where the average lies, for the initial rate of digestion with this peptide. For the determination of the average, it was assumed that all the data lay within the pattern for a normal distribution, in spite of the unusual distribution of the data points, which can be due to a reduced number of experiments, as compared to other peptides. Yet, this still allowed a final determination of an average value of 0.39 ± 0.11 for the relative initial rate of protease cleavage when activated with the peptide GTQSLKRRRCF as compared to the peptide GVQSLKRRRCF. The above value suggests that the side chain of the second residue, besides having to be branched, should also be hydrophobic in order to enable an efficient activation of the protease.

3.4.1.3. The Peptide Ac-GVQSLKRRRCF

Also for this peptide kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide Ac-GVQSLKRRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.17 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide Ac-GVQSLKRRRCF (IR_{Ac}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{Ac} (%/min)	r^2_{Ac}	IR_{Ac}/IR_{11}
0.15	0.01090	0.99031	0.00338	0.98527	0.310
0.15	0.01090	0.99031	0.00348	0.95136	0.319
0.15	0.02050	0.98791	0.00828	0.98440	0.404
0.15	0.01880	0.99737	0.00781	0.97149	0.416
0.15	0.01880	0.99737	0.00862	0.95616	0.459
0.15	0.00783	0.95468	0.00370	0.98793	0.472
0.15	0.02050	0.98791	0.01017	0.99226	0.496
0.15	0.00871	0.97358	0.00436	0.98196	0.501
0.15	0.00871	0.97358	0.00492	0.97349	0.565
0.15	0.00880	0.95452	0.00539	0.98532	0.612
0.15	0.00783	0.95468	0.00539	0.99817	0.688
0.15	0.00880	0.95452	0.00654	0.97576	0.743

The same observations previously made regarding protease activity variability and good fitting of the experimental data to the theoretical curve, expressed in correlation coefficient values always greater than 0.95, also apply to these results. The relative initial rate of digestion values obtained after the calculations are exhibiting once more a very moderate variation, regardless of the concentration of peptide used for activation, allowing for a statistical treatment of the data. As previously, an histogram depicting the distribution of the relative initial rate illustrates the average value to be calculated:

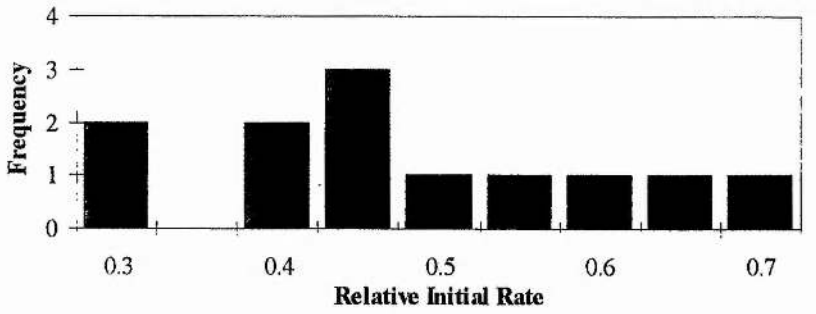


Figure 3.36 - Histogram depicting the distribution of the relative initial rate IR_{Ac}/IR_{11}.

In this histogram, the highest frequency of values is within the interval 0.45-0.50, indicating a first approximation to the value of the average to be calculated. Considering

a normal distribution, this allowed a final determination of an average value of 0.50 ± 0.14 for the relative initial rate of protease cleavage when activated with the peptide Ac-GVQSLKRRRCF as compared to the peptide GVQSLKRRRCF. Again this reveals the importance of the N-terminal residues and their specificity, as a change in the polarity of the amino group decreases activity to half of its original value with the ordinary peptide. This also suggests that peptides with an extended peptide chain along the N-terminal should not be as effective as the cleaved peptide in activating the protease, raising the possibility that pVI is not as effective as the cleaved C-terminal in the activation process. This was studied further below.

3.4.2. N-terminal Prolongation with Consensus Sequence

The previous results, in which the acetylation of the activating peptide in its N-terminal, yielding the peptide Ac-GVQSLKRRRCF, caused a reduction in activation when compared to the native peptide, raised the question of how the activating peptide is released from pVI in the first instance. Certainly, previous studies (Webster *et al.*, 1993) showed that pVI was able to activate the protease, but the effectiveness of this activation had not been assessed. Moreover, as the activating peptide is cleaved from pVI by the protease, it was possible that the mechanism of activation was concomitant with the cleavage of the peptide, with the C-terminal of pVI binding to the protease in such a way that the cleavage site was positioned in the active site and the C-terminal cysteine of pVI close to the cysteine of the protease with which it dimerises (cysteine 104), as illustrated by the following scheme:

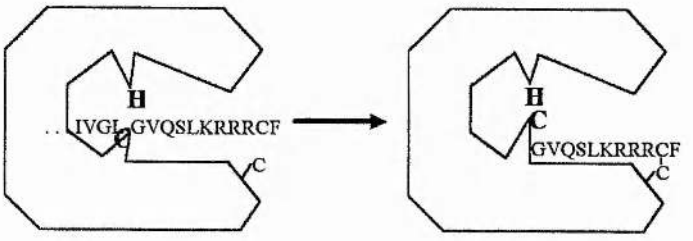


Figure 3.37 - Scheme illustrating a possible mechanism of activation concomitant with the cleavage of the C-terminal of pVI. The binding of the activating peptide would cause a conformational change in the active site by pushing together the residues involved in catalysis, thus converting the protease to its active conformation. The putative active site cysteine (**C**) and histidine (**H**) are also depicted.

In order to test for this hypothesis, several peptides were synthesised, with the following sequences: IVGLGVQSLKRRRCF, IVALGVQSLKRRRCF and VVGLGVQSLKRRRCF. Whilst the first peptide corresponds to the C-terminal 15

residues of pVI, including a cleavage site that matches the consensus sequence for the protease, the latter two comprise point mutations in amino acids that alter the consensus sequence, rendering it uncleavable by the protease, or which will be cleaved in spite of not presenting a consensus sequence, should the above model be verified. Also, if the hypothesis above was true and the peptides were not cleaved, the latter peptides would inhibit the protease, by blocking the active site and preventing other substrates from reaching it.

3.4.2.1. The peptides (I,V)V(G,A)LGVQSLKRRRCF

The above three peptides were synthesised as described in Method 2.4.1, using mini synthesis, Method 2.4.1.4. Several attempts were made to purify all peptides by HPLC (Method 2.4.2.1) (4.6 mm × 25 cm C18 column), but the final samples were still reasonably impure, as assessed by capillary electrophoresis, where several peaks were observed, although the main peak was the intended peptide. This was most probably due to an inefficient synthesis which produced various peptides very similar in composition which were quite difficult to separate. The best purification method was using HPLC with an isocratic gradient of 49% acetonitrile.

The first experiment aimed at verifying whether the peptide IVGLGVQSLKRRRCF activated the protease and was cleaved by it at the same time. Activity assays were performed using 10 µl of protease (0.08 mg/ml), 10 µl of peptide solution (0.75 mM) and 30 µl of assay buffer (cf. Method 2.9.1), and samples were taken and analysed using capillary electrophoresis (Method 2.7.1.1). A possibility that was tested at the same time was to verify whether the other peptides, IVALGVQSLKRRRCF and VVGLGVQSLKRRRCF, were also cleaved, due to being directed to the active site, as described in the scheme of Figure 3.37, which would then give some credence to the model.

In some preliminary experiments, after 5 hours the peptide IVGLGVQSLKRRRCF had been cleaved to IVGL and GVQSLKRRRCF with only ~30% of the original peptide remaining, whereas neither IVALGVQSLKRRRCF nor VVGLGVQSLKRRRCF presented any signs of being digested.

Other experiments tried to assess the ability of the peptides in inhibiting the protease. Activity assays were performed as per Method 2.9.1, using the different peptides (0.75 mM) as activators, and samples were taken and analysed using capillary electrophoresis (Method 2.7.1.1). Also, two possibilities were tested: to check whether

the previous activation of the protease with each peptide would inhibit it as compared to a simultaneous incubation of the protease with the peptides and the substrate (LSGAGFSW).

However, the results can be interpreted in various ways. When the protease is initially incubated with each of the peptides for 45 min prior to the addition of the substrate LSGAGFSW, the protease indeed is activated to a different extent with the different peptides, as the following figure shows:

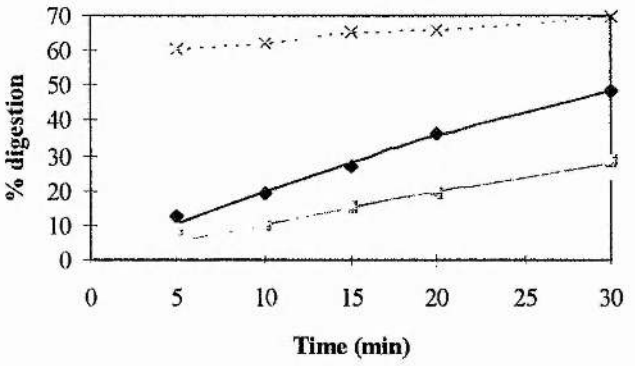


Figure 3.38 - Activation of the protease with: IVGLGVQSLKRRRCF (♦), IVALGVQSLKRRRCF (▪), and VVGLGVQSLKRRRCF (▲). The solid lines show the corresponding percentage of digestion of LSGAGFSW when the protease is activated with each of the above peptides, whereas the dashed line shows the concomitant digestion of the peptide IVGLGVQSLKRRRCF (×) to IVGL and GVQSLKRRRCF. Time zero corresponds to the addition of substrate (LSGAGFSW), after 45 min incubation of the protease with the above peptides.

Although the peptides IVALGVQSLKRRRCF and VVGLGVQSLKRRRCF do seem not to activate the protease as efficiently as IVGLGVQSLKRRRCF, this can be interpreted in much the same way as for the inefficiency of the peptide Ac-GVQSLKRRRCF in activating the protease. Whereas the peptide IVGLGVQSLKRRRCF is itself cleaved, thus producing the peptide GVQSLKRRRCF which activates effectively the protease, the other two peptides are not able to activate the protease as efficiently due to not having the free amino group to the left of the glycine, thus presenting a lower rate of cleavage. This phenomenon could also be explained in terms of protease inhibition, and the following results also allow for this possibility:

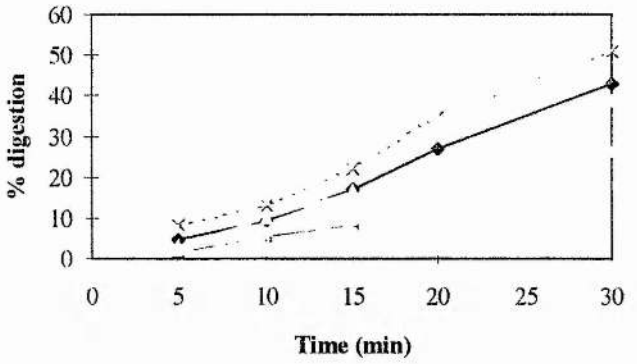


Figure 3.39 - Activation of the protease with: IVGLGVQSLKRRRCF (♦). IVALGVQSLKRRRCF (▪). and VVGLGVQSLKRRRCF (▲). The solid lines show the corresponding percentage of digestion of LSGAGFSW when the protease is activated with each of the above peptides. whereas the dashed line shows the concomitant digestion of the peptide IVGLGVQSLKRRRCF (×) to IVGL and GVQSLKRRRCF. Time zero corresponds to the addition of protease.

In the above experiment, the protease was the last to be added, and if the peptides IVALGVQSLKRRRCF or VVGLGVQSLKRRRCF were inhibiting in the previous experiment, even if at the beginning of this experiment all of them seem to activate to the same extent, then follows a decrease in the rate of digestion which could also be explained in terms of inhibition of the active site of the protease. However, the presence of the extra residues to the left of the GV motif could be removing the properties of a free amino group.

Therefore, the above experiments were inconclusive. The next step would be to try different concentrations of the above peptides, to verify whether the inhibition would increase, a consequence of a greater competition for the active site, or whether it would decrease, in which only the activating properties of the peptides could be accountable for.

Unfortunately, these experiments were not repeated due to the reduced amount of peptides available, thus not allowing a statistical treatment of the data to provide some quantitative answers regarding the possible rate of activation of the protease by the above peptides.

3.5. Importance of C-terminal region of Activating Peptide

The aforementioned experiments tried to elucidate the relevance of the N-terminal region of the activating peptide in the activation mechanism of the protease, both in terms of size of the peptide chain and the specificity of the first two residues in the

efficiency of activation. However, the C-terminal region is known to be involved in the activation (Webster *et al.*, 1993), with the participation of cysteine-10. The nature of the interaction of this cysteine with the protease, as well as the importance of its position within the activating peptide, were the basis of the following experiments.

3.5.1. Cys Shift in Activating Peptide

Some of the questions that remained to be answered regarded the importance of the position of the cysteine in the activating peptide and its environment, namely the striking feature of three consecutive arginines adjacent to the cysteine (cf. Figure 1.9), as well as the distance between the N-terminal of the peptide and the cysteine. The positively charged environment created by the arginines was thought to have a role in the ionisation of the cysteine, thus influencing its oxidation kinetics (cf. 3.2.3). Therefore, by shifting the cysteine along the arginines, a change in the susceptibility of the resulting peptide to ionisation and further oxidation would be expected. However, a change in position altering the activation ability could also signify that the distance between the N-terminal and the cysteine was of importance, again suggesting that the peptide acts as a whole, as discussed in 3.3.3.

3.5.2. Determination of Initial Rates of Peptides with Cys Shift

Two peptides were synthesised, GVQSLKRRCRF and GVQSLKRCRRF, as described in Method 2.4.1, using the semi-automated synthesis, Method 2.4.1.2. Both peptides were purified by HPLC (Method 2.4.2.1) and as a reference typical percentages of acetonitrile for elution of the peptides (4.6 mm × 25 cm C18 column), as well as migration times on capillary electrophoresis (CE) are shown in Table 3.18. The peptide molecular weights were confirmed by mass spectrometry (Method 2.6) and their sequence was confirmed by Method 2.5.

Table 3.18 - Typical values of purification parameters for the peptides GVQSLKRRCRF and GVQSLKRCRRF. Migration times for CE are given for a 24 cm × 25 µm capillary.

Peptide	Percentage of acetonitrile for peptide elution	Migration time on CE (min)
GVQSLKRR**C**R**F**	45-50	3.7
GVQSLKR**C**RR**F**	45-50	4.5

The kinetic assays again consisted in following the digestion curve of LSGAGFSW with time (as described in Method 2.12.1 and discussed in 3.8.3) when the protease was activated with each of the Cys-shifted peptides, and comparing it to the activation with the original activating peptide.

As stated before, the determination of the initial rates of digestion demanded the use of a curve fitting strategy (Method 2.12) due to the limitation of the method used to determine the time points of digestion (capillary electrophoresis, Method 2.7.1.1), which did not allow to quantify accurately reduced amounts of digestion (cf. 3.8.3). Again, several initial rate determinations were made for each peptide activation in order to obtain statistically significant values for the final values.

3.5.2.1. The Peptide GVQSLKRRCRF

Kinetic assays were done with this peptide essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GVQSLKRRCRF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.19 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide GVQSLKRRCRF (IR_{3C}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{3C} (%/min)	r^2_{3C}	IR_{3C}/IR_{11}
0.15	0.06786	0.98798	0.00268	0.99254	0.039
0.15	0.03833	0.99461	0.00315	0.96311	0.082
0.15	0.06507	0.99079	0.00580	0.98137	0.089
0.15	0.06507	0.99079	0.00702	0.97609	0.108
0.15	0.06584	0.98610	0.00724	0.96816	0.110
0.15	0.07528	0.99108	0.00851	0.99771	0.113
0.15	0.03833	0.99461	0.00506	0.99494	0.132
0.15	0.08153	0.98274	0.01105	0.95756	0.136
0.15	0.02614	0.97699	0.00358	0.89796	0.137
0.15	0.06279	0.99034	0.00914	0.96057	0.146
0.15	0.01905	0.99260	0.00281	0.95543	0.147
0.15	0.05471	0.99724	0.00803	0.99303	0.147
0.15	0.05471	0.99724	0.00823	0.95445	0.150
0.15	0.02816	0.84051	0.00437	0.95238	0.155
0.15	0.06279	0.99034	0.01068	0.99446	0.170
0.15	0.02299	0.98671	0.00419	0.96405	0.182
0.15	0.05991	0.99241	0.01089	0.98726	0.182
0.15	0.05991	0.99241	0.01132	0.98575	0.189
0.15	0.05825	0.99410	0.01137	0.95305	0.195
0.15	0.08390	0.98732	0.01648	0.96922	0.196
0.15	0.08390	0.98732	0.01753	0.99549	0.209
0.15	0.01972	0.88564	0.00426	0.82346	0.216
0.15	0.01151	0.98919	0.00275	0.99206	0.239
0.15	0.09990	0.96726	0.02617	0.97369	0.262

As with previous activation assays, protease activity was found to be quite changeable, and a good fitting of the experimental data to the theoretical curve is still observed, expressed in correlation coefficient values almost always greater than 0.95. The calculated values obtained for the relative initial rate of digestion are however within a restricted range, enabling a statistical treatment of the data. Once more, the use of an histogram depicting the distribution of the relative initial rate enables the illustration of the average value to be calculated:

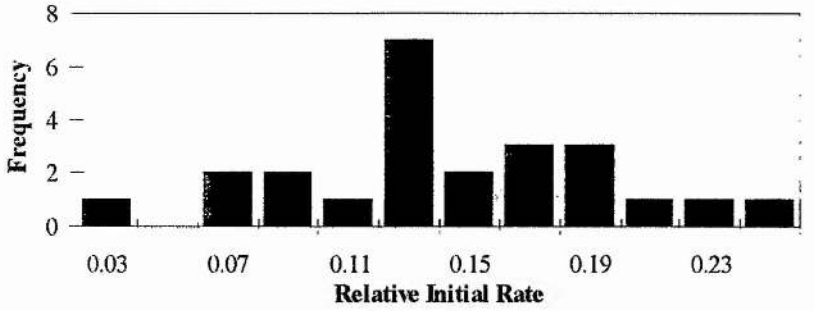


Figure 3.40 - Histogram depicting the distribution of the relative initial rate IR_{30}/IR_{11}.

In this histogram, the highest frequency of values is within the interval 0.13-0.15, indicating a first approximation to the value of the average to be calculated. Assuming a normal distribution, this allowed a final determination of an average value of 0.16 ± 0.06 for the relative initial rate of protease cleavage when activated with the peptide GVQSLKRRCRF as compared to the peptide GVQSLKRRRCF. This result does seem to indicate two possibilities: either the vicinity of an aromatic residue such as phenylalanine is important for the activation step of the protease by the activating peptide, or the distance between the cysteine and the N-terminal of the peptide is of the decisive importance in the activation mechanism, or possibly both these phenomena are acting together in preventing an effective protease activation.

3.5.2.2. The Peptide GVQSLKRCRRF

As already mentioned before, also for this peptide kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GVQSLKRCRRF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.20 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide GVQSLKRCRRF (IR_{4C}) together with the initial rates of the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptide used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{4C} (%/min)	r^2_{4C}	IR_{4C}/IR_{11}
0.15	0.09990	0.96726	0.00146	0.98080	0.015
0.15	0.06584	0.98610	0.00185	0.26219	0.028
0.15	0.08153	0.98274	0.00323	0.99201	0.040
0.15	0.01905	0.99260	0.00083	0.90910	0.043
0.15	0.01151	0.98919	0.00051	0.60000	0.045
0.15	0.07528	0.99108	0.00373	0.95661	0.050
0.15	0.01972	0.88564	0.00105	0.83826	0.053
0.15	0.06786	0.98798	0.00380	1.00000	0.056
0.15	0.02299	0.98671	0.00150	0.98129	0.065
0.15	0.05825	0.99410	0.00435	0.98409	0.075
0.15	0.02614	0.97699	0.00217	0.82876	0.083
0.15	0.02816	0.84051	0.00253	0.91030	0.090
0.074	0.01041	0.97634	0.00226	0.71161	0.217*
0.074	0.01041	0.97634	0.00245	0.93312	0.235*
0.074	0.01041	0.97634	0.00288	0.94243	0.277*
0.074	0.00956	0.94157	0.00307	0.95170	0.321*
0.074	0.01639	0.97262	0.00651	0.91120	0.397*
0.074	0.00956	0.94157	0.00401	0.96224	0.420*
0.074	0.01639	0.97262	0.00793	0.96235	0.484*
0.074	0.00991	0.85992	0.00522	0.98128	0.527*
0.074	0.00991	0.85992	0.00522	0.98797	0.527*
0.074	0.01639	0.97262	0.00870	0.96816	0.531*

*- Discarded values in the average calculation.

The same observations previously made regarding protease activity variability and good fitting of the experimental data to the theoretical curve, expressed in correlation coefficient values always greater than 0.95, also apply to these results. The relative initial rate of digestion values obtained after the calculations are exhibiting once more a very moderate variation, regardless of the concentration of peptide used for activation, allowing for a statistical treatment of the data. Several points found to lie reasonably distant from the putative average were discarded, for reasons similar to those stated for the peptide QSLKRRRCF. As previously, an histogram depicting the distribution of the relative initial rate illustrates the average value to be calculated:

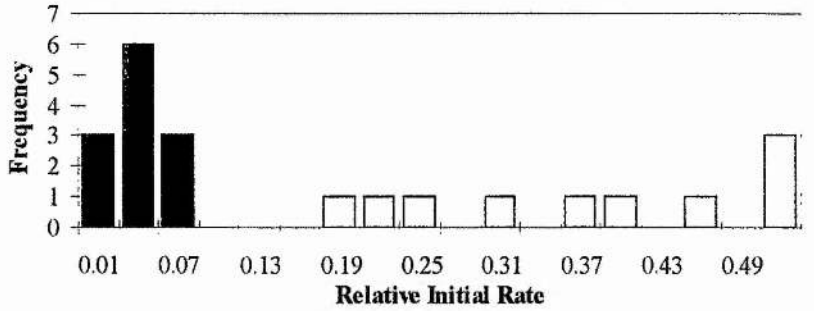


Figure 3.41 - Histogram depicting the distribution of the relative initial rate IR_{4C}/IR_{11}. The void bars correspond to values discarded in the average calculation.

It can be seen that the highest frequency of values lies within the interval 0.04-0.07, with quite a balanced distribution. This histogram resembles the ones in Figure 3.32 and Figure 3.33 in that it presents a maximum towards the lowest values of the relative initial rate, with some scattered values at higher rates corresponding to determinations near the limits of detection, as discussed previously in 3.3.2.2 and 3.3.2.3. Therefore, assuming a normal distribution and discarding the higher values, this allowed a final determination of an average value of 0.054 ± 0.022 for the relative initial rate of protease cleavage when activated with the peptide GVQSLKRCRRF as compared to the peptide GVQSLKRRRCF. The conclusions drawn previously for the activation with the peptide GVQSLKRCRRF can also be applied to this situation. Furthermore, in this case the situation is more extreme, as almost no activation occurs, reinforcing the importance of both the distance between the cysteine and the N-terminal of the peptide, as well as the relevance of the environment surrounding the cysteine.

3.5.3. The Peptide GVQSLARRRCF

In another attempt to try and explain the high conservation of positively charged residues to the left of cysteine-10 in the activating peptide (cf. Figure 1.9), also this peptide was synthesised to try and elucidate any possible role of lysine-6.

The peptide GVQSLARRRCF was synthesised as described in Method 2.4.1, using semi-automated synthesis, Method 2.4.1.2. It was then purified by HPLC (Method 2.4.2.1) and as a reference typical percentages of acetonitrile for elution of the peptide (4.6 mm × 25 cm C18 column), as well as its migration time on capillary electrophoresis (CE) are shown in Table 3.21.

Table 3.21- Typical values of purification parameters for the peptide GVQSLARRRCF. Migration times for CE are given for a 24 cm × 25 μm capillary.

Peptide	Percentage of acetonitrile for peptide elution	Migration time on CE (min)
GVQSLARRRCF	45-50	4.0

As described before, kinetic assays consisted on following the digestion curve of LSGAGFSW with time (as described in Method 2.12.1 and discussed in 3.8.3) when the protease was activated with the above peptide, and comparing it to the activation with the original activating peptide.

Also as already mentioned before, for this peptide too kinetic assays were done essentially as described in Method 2.9.1. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GVQSLARRRCF, and activated by GVQSLKRRRCF in a control assay. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.22 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for activation of the protease with the peptide GVQSLARRRCF (IR_{6A}) together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{6A} (%/min)	r^2_{6A}	IR_{6A}/IR_{11}
0.15	0.00864	0.98511	0.00214	0.96981	0.248*
0.015	0.00823	0.94074	0.00298	0.87185	0.362
0.15	0.02545	0.99555	0.01161	0.99330	0.456
0.15	0.00864	0.98511	0.00412	0.99027	0.477
0.15	0.02086	0.99531	0.01126	0.97575	0.540
0.015	0.00117	0.58264	0.00068	0.26115	0.579
0.015	0.00419	0.98830	0.00255	0.95146	0.608
0.15	0.01343	0.99022	0.00817	0.98724	0.608
0.15	0.00555	0.99635	0.00373	0.99314	0.673

*- Discarded value in the average calculation.

In this set of data, the correlation coefficient is almost always greater than 0.95, an indication of how well the theoretical curve describes the experimental points. In order to

get a value for an average relative initial rate of digestion values were plotted on an histogram. Only one point was found to lie reasonably distant from the others ($\alpha = 0.3$, (Dixon and Massey, 1957)), which led to its discarding:

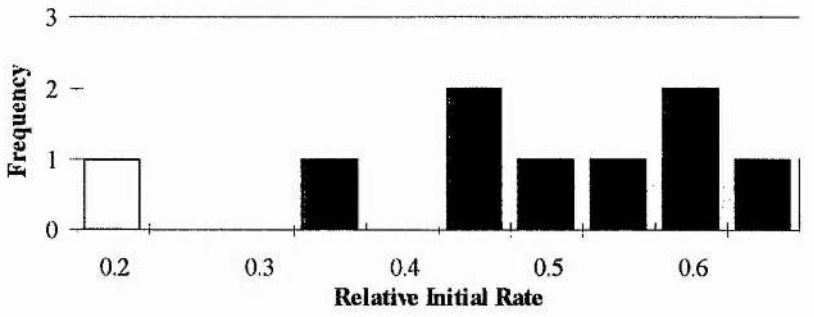


Figure 3.42 - Histogram depicting the distribution of the relative initial rate IR_{6A}/IR_{11}. The void bar corresponds to the value discarded in the average calculation.

From the above histogram, it is reasonably clear where the average is going to lie, although the predictable centre of the distribution does not present the highest frequency of values. Therefore, assuming a normal distribution and discarding the lowest value, which is clearly outside the distribution and can be attributed to experimental error, this allowed a final determination of an average value of 0.54 ± 0.10 for the relative initial rate of protease cleavage when activated with the peptide GVQSLARRRCF as compared to the peptide GVQSLKRRRCF. This reduction of nearly 50% seems to indicate that also the lysine residue seems to be important for the ability of the activating peptide in binding to the protease, possibly due to the removal of a positive charge from the molecule which will affect the pK_a of the N-terminal amino group (cf. 3.2.3.8).

3.5.4. The Effect of the C-terminal Aromatic Residue

Another striking feature of the activating peptide is that an aromatic residue such as phenylalanine or tyrosine is always found at the C-terminus (cf. Figure 1.9). This prompted an investigation into the role of this residue in the activating peptide. By substituting the phenylalanine with an alanine, preliminary results (Kemp, unpublished results) exhibited very little or no activation by the resulting peptide GVQSLKRRRCA, when incubated with the protease. This suggested a role for the aromatic residue in the binding of the peptide to the protease, as this residue is close to the cysteine which seems to be responsible for the covalent binding.

More detailed studies were conducted (Hawkins, 1997), in which the ability of the peptide GVQSLKRRRCA in activating the protease was compared to that of GVQSLKRRRCF both in reduced (monomeric) and oxidised (dimeric) forms. Kinetic assays were done essentially as described in Method 2.9.1, and data were treated using Method 2.12.1. Results showed that the protease (0.005 mg/ml) activated with the reduced form of GVQSLKRRRCA (1 mM) presented a relative initial rate of 0.25 ± 0.03 as compared to that activated with reduced GVQSLKRRRCF, whereas the oxidised form (1 mM) presented a value of 0.34 ± 0.08, not significantly different from the above value ($p = 0.05$). However, these results do stress the importance of the aromatic residue in the activation mechanism. An initial hypothesis that this residue might affect the acidity of the cysteine, rendering it less prone to reaction with the protease cysteine, was eventually discarded as its pK_a was determined (cf. 3.2.3.5) and shown not to be significantly different to that of GVQSLKRRRCF.

Further studies (Hawkins, 1997; Reid, 1997) also showed that even with increasing concentrations these ratios do not seem to vary significantly, although at very high concentrations of each peptide there appeared to be a decrease in the activation abilities of the peptides, more in the case of GVQSLKRRRCF than with GVQSLKRRRCA (Reid, 1997), a phenomenon that could not be explained, and contradicts previous results (Jones *et al.*, 1996).

3.6. *Inhibition Studies*

Of the peptides whose activation ability was studied, the most striking results were those of the N-terminal truncated peptides, in which the removal of GV from the peptide seemed to remove most of the ability of the peptide in activating the adenovirus protease. Also several modifications in that region, such as the acetylation of the N-terminal amino group or the alteration from a valine to either an alanine or a threonine, and their consequent reduction in the ability of the resulting peptides in activating the protease stressed the importance of this region in the activation process. Two possibilities were put forward: (a) the N-terminal region (GV) is necessary for the binding of the peptide to the protease and subsequent activation, or (b) although not necessary for binding, the N-terminal region is responsible for a conformational change which leads to an active form of the protease. Certainly, the interaction of the activating peptide with the protease

was shown to cause a conformational change, measured in a change of tryptophan fluorescence (Jones *et al.*, 1996).

In order to check for the first possibility, mass spectrometry of the protease in the presence of several activating peptides was performed, where an increase of the molecular weight of the protease would indicate the existence of a protease·peptide complex. Another way of verifying this hypothesis was to perform studies of inhibition of activation, whereby adding a peptide to the protease that is unable to activate it would conceivably bind to it and prevent the native peptide from binding and further activate the protease.

3.6.1. Mass Spectrometry of Protease·Peptide Complexes

In order to assert whether the peptides bound to the protease, several approaches have been tried. By radiolabelling the peptide, incubating it with the protease and then performing non-reducing SDS-PAGE, the band corresponding to the protease presented some radioactivity (Kemp, unpublished results). In a similar fashion, by incubating the activating peptide with the protease followed by low acrylamide percentage non-reducing SDS-PAGE, the band corresponding to the putative protease·peptide complex ran at a higher molecular weight than the corresponding band to protease alone (Iqbal, unpublished results). All these results seemed to suggest that indeed the activating peptide bound to the protease.

Using mass spectrometry provided a method of confirming whether the peptide bound to the protease or not, and if it did, whether this binding was covalent or not.

The peptides GVQSLKRRRSF, GVQSL, GVQSLKRRRCF, VQSLKRRRCF, QSLKRRRCF and SLKRRRCF were incubated at a protease to peptide ratio of 1:100 and subjected to mass analysis by time-of-flight mass spectrometry using the matrix-assisted laser desorption technique (Method 2.6), and the results are shown in Figure 3.43.

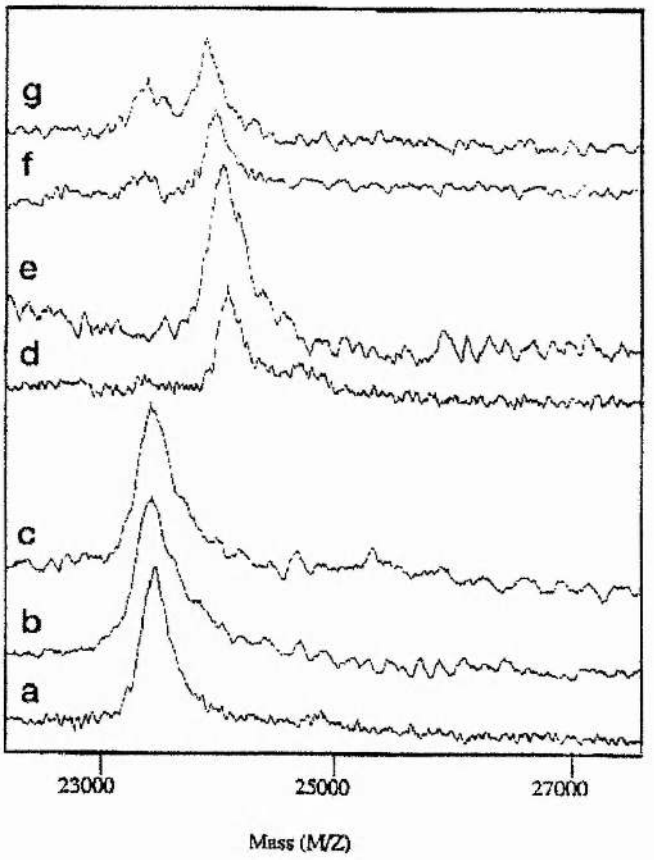


Figure 3.43 - Mass spectra of protease·peptide complexes: protease alone (a), and after the addition of GVQSLKRRRSF (b). GVQSL (c), GVQSLKRRRCF (d), VQSLKRRRCF (e), QSLKRRRCF (f), and SLKRRRCF (g). Taken from (Cabrita *et al.*, 1997).

These results show that whereas the protease does not seem to bind to either GVQSL or GVQSLKRRRSF, it does seem to form a 1:1 complex with GVQSLKRRRCF, VQSLKRRRCF, QSLKRRRCF and SLKRRRCF, which is responsible for the increase of the molecular weight of the protease. The observation that the protease·peptide complexes were only observed when the peptide added contained a cysteine residue provides evidence that these complexes are due to the formation of disulphide-bonded heterodimers. This is further corroborated by the observation (Cabrita *et al.*, 1997) that no complex is observed when the peptide GVQSLKRRRCF is incubated with a protease mutant where cysteine-104 was replaced by alanine.

Furthermore, it was interesting to observe that in the spectra depicting the complex formed with the peptides QSLKRRRCF and SLKRRRCF there seems to be evidence of some uncomplexed protease, even when the peptide is present in a 100-fold molar excess (cf. Figure 3.43), which suggests that these peptides bind less efficiently.

3.6.2. Inhibition by N-terminal Truncated Peptides

Having established that the shorter peptides do bind to the protease (cf. 3.6.1), albeit to a lesser extent, and do not seem to activate it efficiently (cf. 3.3.2.2 and 3.3.2.3), a question was raised as to whether those peptides would remain bound to the protease, thus preventing the native activating peptide from binding to it, thereby having an effect as inhibitors of activation.

To test the above hypothesis, inhibition assays were performed as described in Method 2.9.3. Briefly, these consisted of incubating the protease with the putative inhibiting peptide for 15 min, followed by incubation with the native activating peptide GVQSLKRRRCF for another 15 min. The protease was then tested for digestion by adding the substrate LSGAGFSW.

Also for the inhibition assays initial rates of digestion were determined. These too were done as described previously (Method 2.12). Several initial rate determinations were made for each peptide inhibition in order to obtain statistically significant values for the final values.

3.6.2.1. The Peptide QSLKRRRCF

Assays of inhibition kinetics were done essentially as described in Method 2.9.3. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was inhibited in duplicate by the peptide QSLKRRRCF, and followed by activation with GVQSLKRRRCF. The control assay consisted of adding GVQSLKRRRCF to the protease and adding water when the activating peptide is being added to the test assays. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.23 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for inhibition of the protease with the peptide QSLKRRRCF (IR_9) together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay. the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1. and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_9 (%/min)	r^2_9	IR_9/IR_{11}
0.125	0.07389	0.98995	0.04510	0.99226	0.610
0.125	0.08005	0.96698	0.05168	0.98613	0.646
0.125	0.06464	0.97119	0.04712	0.97846	0.729
0.125	0.06464	0.97119	0.04939	0.97237	0.764
0.125	0.07389	0.98995	0.06057	0.99239	0.820
0.125	0.08005	0.96698	0.07052	0.97853	0.881
0.125	0.04650	0.99958	0.04291	0.98522	0.923

Firstly it should be noticed that, as the values for the initial rates did not vary to a great extent, not so many values were needed to be determined. Also, the correlation coefficient is always greater than 0.95, indicating a close agreement between the experimental points and the theoretical curve. Also, the final relative initial rates of digestion only vary within a reasonably restricted interval, allowing for a straightforward statistical treatment of the data. As in the previous determinations, an histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated:

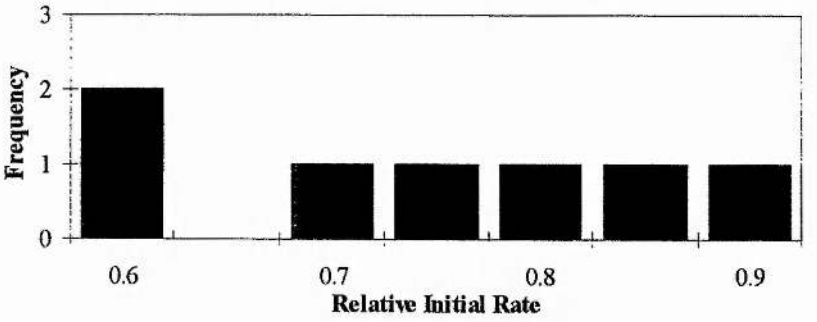


Figure 3.44 - Histogram depicting the distribution of the relative initial rate IR_9/IR_{11} with QSLKRRRCF used as an inhibitor.

Yet, from the above histogram it is not evident which is the final average, and therefore all the values were used in the calculation of the average. This led to a final determination of an average value of 0.77 ± 0.12 for the relative initial rate of protease cleavage when inhibited with the peptide QSLKRRRCF and followed by activation with

GVQSLKRRRCF, as compared to the protease activated initially with the latter peptide. These results seem to indicate that the peptide QSLKRRRCF does bind to the protease, as there is a lower activity of the protease when incubated with it, in agreement with the mass spectrometry results (cf. 3.6.1). However, the fact that there is a recovery in activity with the addition of the peptide GVQSLKRRRCF indicates that the latter peptide is somehow able to displace the inhibiting peptide and bind itself to the protease, thus rendering it active. This greater affinity of the native activating peptide for the protease reiterates the conclusion previously drawn that the presence of the residues glycine and valine in the N-terminal of GVQSLKRRRCF are determinant in the activation process of the protease, probably by helping the peptide in binding to it more efficiently.

3.6.2.2. The Peptide SLKRRRCF

The assays of inhibition kinetics for this peptide were also done essentially as described in Method 2.9.3. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was inhibited in duplicate by the peptide SLKR̄RRCF, and followed by activation with GVQSLKRRRCF. As previously, the control assay consisted of adding GVQSLKRRRCF to the protease and adding water when the activating peptide is being added to the test assays. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.24 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for inhibition of the protease with the peptide SLKRRRCF (IR_8) together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	**IR_{11} (%/min)**	**r^2_{11}**	**IR_8 (%/min)**	**r^2_8**	**IR_8/IR_{11}**
0.125	0.06663	0.98885	0.05045	0.99258	0.757
0.125	0.07016	0.96730	0.05959	0.95881	0.849
0.125	0.06457	0.98259	0.05918	0.95765	0.917
0.125	0.07016	0.96730	0.06497	0.96505	0.926
-	-	-	-	-	1.079*
0.125	0.06457	0.98259	0.07010	0.98819	1.086
0.125	0.04163	0.98458	0.04544	0.97130	1.092
0.125	0.06663	0.98885	0.07578	0.98542	1.137
-	-	-	-	-	1.138*
0.125	0.04163	0.98458	0.05150	0.96682	1.237
-	-	-	-	-	1.251*

*- Values determined by Dr. H. Reddy.

Again for this peptide it can be seen that, as the values for the initial rates did not vary broadly, fewer values were determined than in previous assays. Again the correlation coefficient is always greater than 0.95, indicating a close agreement between the experimental points and the theoretical curve. However, in this case the final relative initial rates of digestion are somewhat scattered, which is possibly a result of low digestion rates and its consequent greater error in their determination. The following histogram depicting the distribution of the relative initial rate illustrates the average value to be calculated:

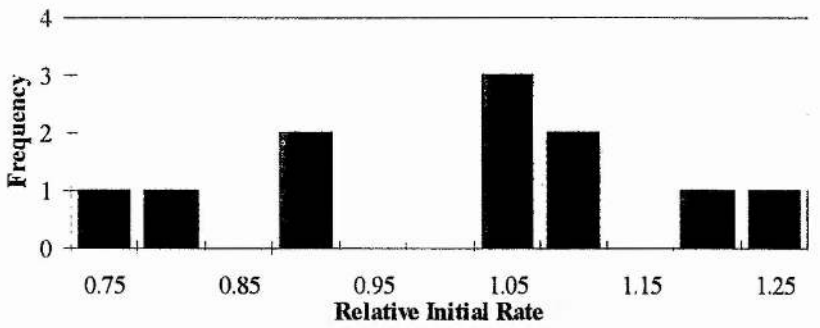


Figure 3.45 - Histogram depicting the distribution of the relative initial rate IR_8/IR_{11} with SLKRRRCF used as an inhibitor.

As with the previous determination, in the above histogram it is not evident which is the final average, and all the values were used in the calculation of the average. This resulted in a final average value of 1.04 ± 0.16 for the relative initial rate of protease cleavage when inhibited with the peptide SLKRRRCF and followed by activation with GVQSLKRRRCF, as compared to the protease activated initially with the latter peptide. These results do not allow for the conclusion that also the peptide SLKRRRCF binds to the protease, as expected from the mass spectrometry results (cf. 3.6.1), as there is no reduction in the activity of the protease, although they do not contradict it either. Assuming that the peptide does bind, the fact that there is a recovery in activity as the peptide GVQSLKRRRCF is added could suggest that the latter peptide is somehow able to displace also this inhibiting peptide and bind itself to the protease, thus rendering it active. This greater affinity of the native activating peptide for the protease would also reinforce the conclusion previously drawn about the importance of the residues glycine and valine at the N-terminal of the peptide in the activation process of the protease.

3.6.2.3. Other Peptides

Some more peptides other than the above mentioned were studied in their ability to inhibit the protease. These determinations were done by Dr. H. Reddy exactly as described previously (cf. 3.6.2) with peptides previously synthesised (with the exception of GV, purchased from Sigma), and the results are summarised in the following table:

Table 3.25 - Inhibition of the adenovirus protease by other peptides. Results are expressed as percent activity (initial rate) relative to that achieved with GVQSLKRRRCF alone and are expressed as the mean ± S.E. of at least 3 determinations (adapted from (Cabrita *et al.*, 1997)).

Peptide	Activity (%)
GVQSLKRRRCF	100
LKRRRCF	100 ± 6
KRRRCF	108 ± 6
GVQSL	109 ± 6
GV	96 ± 4

The results from the above table together with the values obtained in 3.6.2.1 and 3.6.2.2 show that only the peptide QSLKRRRCF is able to inhibit to some extent the protease, which corresponds to the longest peptide that did not activate the protease (cf. 3.3.2). Again, this stresses the importance of the N-terminal region of the activating peptide in the activation of the protease. Also, the fact that the neither the peptide GV nor GVQSL were able to activate the protease confirms that any peptide has to contain a cysteine with which to bind covalently to the protease in order to activate it efficiently. This result does suggest that the activating peptide interacts with the protease via two distinct regions: covalently through cysteine-10, and electrostatically in the N-terminal region of the peptide.

3.6.2.4. Binding Competition between GVQSLKRRRCF and QSLKRRRCF

Having established that the peptide QSLKRRRCF does bind to the protease (cf. 3.6.1), and is not easily displaced by the native peptide GVQSLKRRRCF (cf. 3.6.2.1), the consequent question was to know whether the final binding of each of the peptides was competitive and independent of the order in which each peptide was added, or whether the peptide GVQSLKRRRCF had a greater affinity in binding to the protease. In order to address this question, a modified inhibition assay was performed, in which the native peptide GVQSLKRRRCF was added prior to the addition of the peptide QSLKRRRCF, to check whether the final activation would be the result of a competition between both peptides, or whether the peptide GVQSLKRRRCF remained bound to the protease.

The assays of inhibition kinetics were again done essentially as described in Method 2.9.3. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was activated in duplicate by the peptide GVQSLKRRRCF, and followed by inhibition with QSLKRRRCF. The control assay consisted of adding

GVQSLKRRRCF to the protease and adding water when the activating peptide is being added to the test assays. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table (duplicate assays are displayed where two identical rates of digestion are found for the control peptide):

Table 3.26 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for inhibition of the protease with the peptide QSLKRCRRF (IR_9) after GVQSLKRRRCF was added, together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_9 (%/min)	r^2_9	IR_9/IR_{11}
0.125	0.05050	0.96876	0.03499	0.95923	0.693*
0.125	0.05923	0.98095	0.04108	0.99882	0.694*
0.125	0.04736	0.99817	0.04038	0.97741	0.853
0.125	0.05082	0.98796	0.05048	0.99123	0.993
0.125	0.05923	0.98095	0.05918	0.99041	0.999
0.125	0.05050	0.96876	0.05355	0.99276	1.060
0.125	0.04736	0.99817	0.05060	0.99017	1.068
0.125	0.05082	0.98796	0.05918	0.99385	1.164
0.125	0.04250	0.97105	0.05172	0.99410	1.217
0.125	0.05338	0.98962	0.06602	0.99626	1.237
0.125	0.04250	0.97105	0.05868	0.99618	1.381
0.125	0.03485	0.99994	0.04877	0.95037	1.399
0.125	0.05338	0.98962	0.07556	0.99455	1.415
0.125	0.03485	0.99994	0.06192	0.97909	1.777*

*- Discarded values in the average calculation.

This table shows a reasonable constancy of the values determined for both the initial rates, together with a correlation coefficient always greater than 0.95, indicating an excellent agreement between the experimental points and the theoretical curve. Also, as the final relative initial rates of digestion are constricted to a reasonably narrow interval, this allows for a straightforward statistical treatment of the data. As in the previous determinations, an histogram depicting the distribution of the relative initial rate enables an easier illustration of the average value to be calculated:

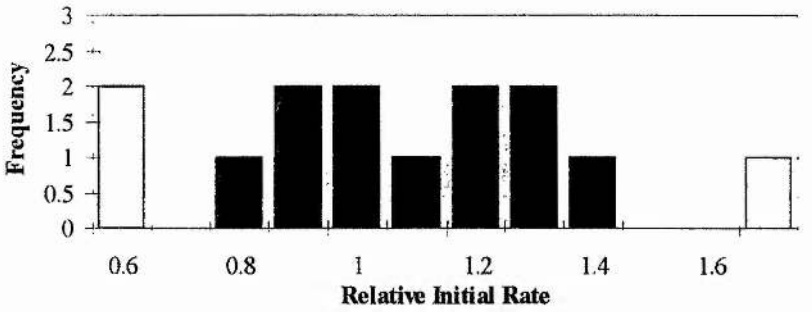


Figure 3.46 - Histogram depicting the distribution of the relative initial rate IR_9/IR_{11} with QSLKRRRCF used as an inhibitor after the addition of GVQSLKRRRCF. The void bars correspond to the values discarded in the average calculation.

From the above histogram the final average seems to lie within the interval 1.1-1.2, although this does not correspond to the highest frequency interval. Some of the extreme values in the histogram were considered to lie outside a normal distribution ($\alpha = 0.3$, (Dixon and Massey, 1957)), and were therefore discarded in the calculation of the average. This led to a final determination of an average value of 1.16 ± 0.19 for the relative initial rate of protease cleavage when activated with the peptide GVQSLKRRRCF and followed by inhibition with QSLKRRRCF, as compared to the protease activated initially with the native peptide. These results seem to indicate that although the peptide QSLKRRRCF binds to the protease when it is added first, the fact that there is no significant change in activity when the peptide GVQSLKRRRCF is added first indicates that the latter peptide has a greater affinity to the protease, thus not being displaced by the other peptide. This greater affinity of the native activating peptide for the protease once more reinforces the conclusion previously drawn that the presence of the residues glycine and valine in the N-terminal of GVQSLKRRRCF are determinant in the binding process of the peptide to the protease.

3.6.3. Inhibition by Cys-Shifted Peptides

It was previously shown (cf. 3.5.1) that the Cys-shifted peptides are poor activators of the protease. However, these peptides (GVQSLKRRCRF and GVQSLKRCRRF) do conform to the previous requisites deemed necessary for an efficient activation of the protease, which are the presence of a GV in the N-terminal region and a cysteine in its sequence. As suggested above (cf. 3.5.2.1), it is also possible that the distance between the N-terminal and the cysteine is of importance in an efficient activation of the protease.

Unfortunately, mass spectrometry of possible protease·peptide complexes with the above peptides, in order to ascertain whether there is binding of the above peptides to the protease or not, was not possible. This would have permitted to assert about the influence of the cysteine environment on its ability to bind the peptide to the protease.

However, inhibition studies were made to try and provide an alternate way of asserting whether these peptides bound to the protease or not. It was expected that, in the case of binding and greater affinity for the protease than the native peptide, some inhibition would be observed.

3.6.3.1. The Peptide GVQSLKRRCRF

Inhibition assays were done essentially as described in Method 2.9.3. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was inhibited by the peptide GVQSLKRRCRF, followed by activation with GVQSLKRRRCF. The control assay consisted of adding GVQSLKRRRCF to the protease and adding water when the activating peptide is being added to the test assays. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table:

Table 3.27 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for inhibition of the protease with the peptide GVQSLKRRCRF (IR_{3C}) together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{3C} (%/min)	r^2_{3C}	IR_{3C}/IR_{11}
0.125	0.01552	0.98049	0.00733	0.99771	0.472
0.125	0.01027	0.90597	0.00607	0.99507	0.591
0.125	0.02248	0.95910	0.01620	0.77583	0.721
0.125	0.05680	0.98087	0.04191	0.95553	0.738
0.125	0.01167	0.74010	0.00874	0.99179	0.749
0.125	0.00778	0.98787	0.00667	0.99605	0.856
0.125	0.01926	0.64588	0.01123	0.82118	0.876
0.125	0.01301	0.98511	0.01161	0.97003	0.893
0.125	0.01389	0.96789	0.01337	0.94385	0.963
0.125	0.01974	0.94394	0.02025	0.99741	1.026
0.125	0.00808	0.93323	0.01042	0.96659	1.290
0.125	0.00949	0.98728	0.01225	0.99076	1.291
0.125	0.00498	0.99215	0.00856	0.78818	1.718
0.125	0.00223	0.80693	0.00526	0.99344	2.358*

*- Discarded value in the average calculation.

The results in the above table show again a reasonable variability in the initial rates determined. Also, the correlation coefficient is not always greater than 0.90, indicating a poorer agreement than usual between the experimental points and the theoretical curve. Even the final relative initial rates of digestion do vary within a considerably broad interval. As in the previous determinations, an histogram depicting the distribution of the relative initial rate allows a better illustration of the average value to be calculated:

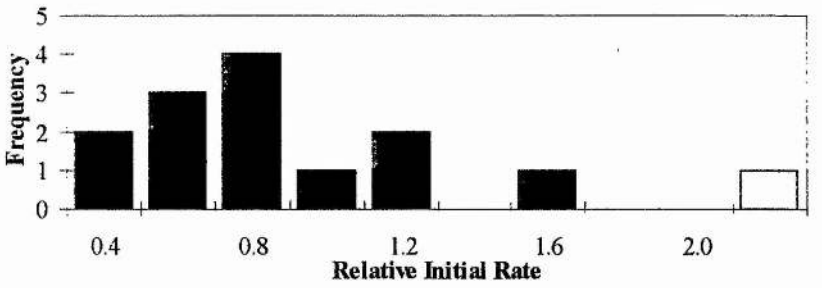


Figure 3.47 - Histogram depicting the distribution of the relative initial rate IR_{3C}/IR_{11} with GVQSLKRRCRF used as an inhibitor. The void bar corresponds to the value discarded in the average calculation.

From the above histogram the final average seems to lie within the interval 0.8-1.0. The extreme value to the right of the histogram was considered to lie outside a normal distribution ($\alpha = 0.01$, (Dixon and Massey, 1957)), and was discarded in the calculation of the average. This resulted in a final average value of 0.94 ± 0.34 for the relative initial rate of protease cleavage when inhibited with the peptide GVQSLKRRCRF and followed by activation with GVQSLKRRRCF, as compared to the protease activated initially with the native peptide. This result means that the Cys-shifted peptide is unable to inhibit the protease. From previous results (cf. 3.5.2.1) the peptide does not seem to be able to activate it to a reasonable extent either, and this can be due to an inefficient binding to the protease or to an incorrect distance between the cysteine and the N-terminal GV. Therefore, these results, although corroborating the previous ones, did not allow for any further conclusions to be withdrawn.

3.6.3.2. The Peptide GVQSLKRCRRF

Inhibition assays for this peptide were also done essentially as described in Method 2.9.3. In a typical assay, three substrate digestions were followed with time, where the protease (0.06 mg/ml) was inhibited by the peptide GVQSLKRCRRF, followed by activation with GVQSLKRRRCF. The control assay consisted of adding GVQSLKRRRCF to the protease and adding water when the activating peptide is being added to the test assays. Time points were plotted against its correspondent percentage of substrate digestion and a curve was fitted using Method 2.12. The results are presented in the following table:

Table 3.28 - Experimentally determined values for initial rates (IR) of digestion of the substrate LSGAGFSW for inhibition of the protease with the peptide GVQSLKRCRRF (IR_{4C}) together with the initial rates of activation with the control peptide GVQSLKRRRCF (IR_{11}). The first column indicates the concentration of peptides used in the assay, the following four columns the initial rates obtained with the respective peptides in percentage of substrate digested per minute (initial conc.: 4.9 mM) together with the correlation coefficient obtained using Method 2.12.1, and the last column indicates the relative initial rate.

Conc. (mM)	IR_{11} (%/min)	r^2_{11}	IR_{4C} (%/min)	r^2_{4C}	IR_{4C}/IR_{11}
0.125	0.05680	0.98087	0.01364	0.97086	0.240
0.125	0.01552	0.98049	0.00604	0.99677	0.389
0.125	0.01167	0.74010	0.00693	0.91872	0.594
0.125	0.02248	0.95910	0.01556	0.89854	0.692
0.125	0.01301	0.98511	0.01026	0.99706	0.789
0.125	0.01027	0.90597	0.00819	0.99895	0.798
0.125	0.01389	0.96789	0.01134	0.97599	0.817
0.125	0.01282	0.83752	0.01570	0.80186	1.224
0.125	0.00808	0.93323	0.01037	0.91543	1.283
0.125	0.01974	0.94394	0.02543	0.89977	1.288
0.125	0.00949	0.98728	0.02187	0.98349	2.305
0.125	0.00498	0.99215	0.01596	0.98400	3.205*
0.125	0.00778	0.98787	0.04530	0.96017	5.819*
0.125	0.00223	0.80693	0.01566	0.92087	7.020*

*- Discarded values in the average calculation.

Again it can be noticed that the values for the initial rates did vary considerably, and the correlation coefficient was not always greater than 0.90, a reflection of a coarser agreement between the experimental points and the theoretical curve. Also the final relative initial rates of digestion vary within a reasonably ample interval, making it a troublesome task to decide how to treat the data. As in the previous determinations, an histogram depicting the distribution of the relative initial rate provided a better illustration of the average value to be calculated:

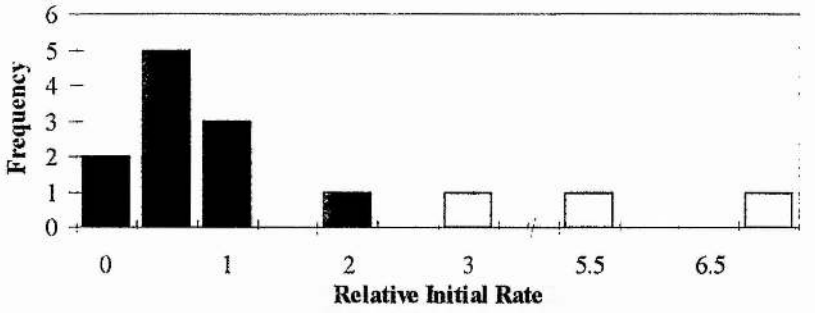


Figure 3.48 - Histogram depicting the distribution of the relative initial rate IR_{4C}/IR_{11} with GVQSLKRCRRF used as an inhibitor. The void bars correspond to values discarded in the average calculation.

From the above histogram it is clearer that the final average seems to lie within the interval 0.5-1.0, quite a broad interval due to the spanning of the data. The values to the right of the histogram were considered to lie outside a normal distribution ($\alpha = 0.005$, (Dixon and Massey, 1957)), and were therefore discarded in the calculation of the average. This allowed a final determination of an average value of 0.95 ± 0.57 for the relative initial rate of protease cleavage when inhibited with the peptide GVQSLKRCRRF followed by activation with GVQSLKRRRCF, as compared to the protease activated initially with the native peptide. The conclusions to take from this result are similar to those for the previous peptide (cf. 3.6.3.1): there is no further evidence that the peptide does bind to the protease, or if it does bind, it is readily displaced by the native activating peptide.

3.7. The Adenovirus CELO Protease

The CELO (Chicken Embryo Lethal Orphan) is the prototype of avian adenoviruses, and it has been shown to be closely related to quail bronchitis virus (Andrewes *et al.*, 1978). It was first identified as an adventitious contaminant of embryonated eggs during attempts to propagate a bovine skin disease agent, and further identified as an infectious agent in 1957 (Yates, cited in (Chiocca *et al.*, 1996)). CELO is classified as a fowl adenovirus type 1 (FAV-1).

The interest in obtaining this protease relied on it being reasonably dissimilar from the human adenovirus type 2 protease and on the sequence of its activating peptide being also known (GVATATRRMCY, cf. Figure 1.9), so that the specificity of the latter peptide to the protease could be studied, and comparisons could be made regarding the interchangeability between activating peptides and proteases. This would possibly allow

a better insight into which parameters govern the specificity of the interaction between the activating peptide and the protease.

3.7.1. Cloning of the Protease Gene into a pET-11c Vector

In order to clone the protease gene into a pET-11c vector (Novagen), Xba I and BamH I restriction sites had to be created (cf. 2.1.1), the former for the 5'-primer and the latter for the 3'-primer. Taking into consideration the DNA genome for the CELO adenovirus (from GenBank, http://www3.ncbi.nlm.nih.gov/Entrez), the following primers were synthesised:

Primer 5': GGCAACGCTG***TCTAGA***AAAAAGACGG (***Xba I***)

Primer 3': CAGTGA***GGATCC***TTGTCGGTCTATCGG (***BamH I***)

The program Amplify (Engels, 1992) was used to test the above primers for primability and stability, and resulted in the following values:

Table 3.29 - Values for efficiency of PCR reaction (Engels, 1992) with the designed primers.

Primer	Primability	Stability
primer-5'	96%	67%
primer-3'	98%	71%

The program also indicated that with these primers only the protease gene should be amplified with PCR, and this was indeed the result of the PCR (Method 2.1.1.2) as verified on an agarose gel (Method 2.1.1.1) presenting only one band at the correct size.

The amplified fragment was then extracted from the gel (Method 2.1.1.3) and its extremities were then cut with the appropriate restriction enzymes Xba I and BamH I (Method 2.1.1.4), and the restricted fragment was then ligated into the pET-11c vector (Method 2.1.1.5) also already cut using Method 2.1.1.4.

XL1-Blue cells (Stratagene) were prepared and transformed with the ligated plasmid according to Method 2.1.1.6 and plated in agar plates with ampicillin and tetracycline (Method 2.1.1.7) in order to select for transformants which contained the plasmid. A mini-prep was done (Method 2.1.1.8) to extract the plasmid, and this was again restricted (Method 2.1.1.4) and run on an agarose gel (Method 2.1.1.1) in order to find which plasmids had the insert.

The appropriate plasmid was selected and retransformed in XL1-Blue cells (Method 2.1.1.6) and a maxi-prep was done (Method 2.1.1.9) in order to prepare a plasmid stock with which to transform the host system for expression of the cloned protein

(BL21(DE3) cells, Novagen). The plasmid was then sequenced (Method 2.1.1.11) and the protease was confirmed to have been cloned into the plasmid.

3.7.2. Transformation of the Host System BL21(DE3) with the Vector and Expression of the Protease

BL21(DE3) cells (Novagen) were prepared and transformed with the plasmid stock according to Method 2.1.1.6 and plated in agar plates with ampicillin (Method 2.1.1.7) in order to select for transformants which contained the plasmid. A single cell colony was then picked and used to inoculate a 5 ml LB medium, which after growth was used to inoculate 2×250 ml of M9 medium, as described in Method 2.1.1.13 for the expression of the cloned protease.

3.7.3. Attempts to Obtain Protease Expression

In the meantime, the peptide GVATATRRMCY, the activating peptide of the avian adenovirus CELO, was synthesised by using Method 2.4.1.2 and Method 2.4.1.3, and purified using Method 2.4.2.1. Its sequence was then verified using Method 2.5.1.

However, when activity assays were performed using Method 2.9.1 using protease directly from the protease extraction, no activity was found after 1 h digestion. A time-course of expression was done (Method 2.1.1.14) and this revealed the system not to be expressing the protease. Growing the cells in LB rather than M9 did not improve in any way the expected expression.

3.7.4. Recloning of the CELO Protease Including a Ribosome Binding Site

An analysis of the vector used (pET-11c) revealed that when the 5' restriction site for Xba I is used it removes a ribosome binding site which is supposed to enhance the translation of the target sequence (Olins and Rangwala, 1989). In order to test whether this was the reason why the protease was not being expressed, a primer was designed (provided by S. Annan) with which to reamplify the gene from the plasmid, and containing the ribosome binding site (rbs); on the other end, a primer designed for priming any sequence cloned into pET-11c (provided by L. Murray) from the 3' end:

Primer 5': TTCCCC***TCTAGA***AAAGGAGCGG (***Xba I***, rbs)

Primer 3': TTATGCTAGTTATTGCTCAGCGGTG

These primers were then used with the plasmid obtained above on a PCR (Method 2.1.1.2), and the band obtained at the right size on an agarose gel (Method 2.1.1.1) was extracted (Method 2.1.1.3).

The new fragment obtained was then cut with the appropriate restriction enzymes Xba I and BamH I (Method 2.1.1.4), and the restricted fragment was then ligated to the pET-11c vector (Method 2.1.1.5) also already cut using Method 2.1.1.4, as before.

The new plasmid was then transformed in XL1-Blue cells (Method 2.1.1.6) and the cells plated in agar plates with ampicillin and tetracycline (Method 2.1.1.7) again in order to select for transformants which contained the plasmid. A mini-prep was done (Method 2.1.1.8) to extract the plasmid, and this was again restricted (Method 2.1.1.4) and run on an agarose gel (Method 2.1.1.1) in order to find which plasmids had the insert.

As with the first plasmid, one that presented the insert was selected and retransformed in XL1-Blue cells (Method 2.1.1.6) and a maxi-prep was done (Method 2.1.1.9) in order to prepare a plasmid stock with which to transform the host system (BL21(DE3) cells, Novagen), and the plasmid was checked for the presence of the insert by cutting it with the appropriate restriction enzymes Xba I and BamH I (Method 2.1.1.4) and running on an agarose gel (Figure 3.49):

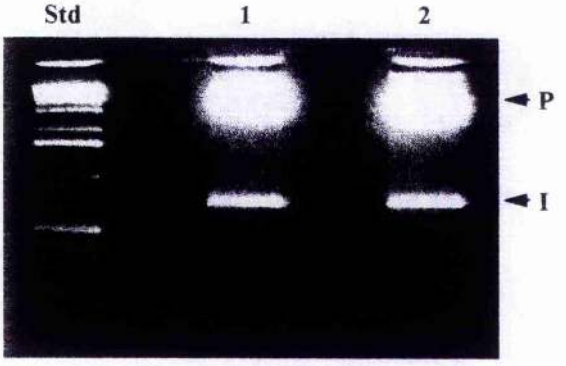


Figure 3.49 - Restriction of plasmid pET-11c cloned with CELO adenovirus protease, with enzymes Xba I and BamH I. The plasmid (**P**) is shown to contain only one insert (**I**), which should correspond to the protease gene. Lanes **1** and **2** are duplicates.

3.7.5. Transformation of the Host System BL21(DE3) with the Corrected Vector and Expression of the Protease

BL21(DE3) cells were again prepared and transformed with the plasmid stock according to Method 2.1.1.6 and plated in agar plates with ampicillin (Method 2.1.1.7) in order to select for transformants which contained the plasmid. A single cell colony was

then picked and used to inoculate a 5 ml LB medium, which after growth was used to inoculate 2×250 ml of M9 medium, as described in Method 2.1.1.13 for the expression of the cloned protease.

3.7.6. New Attempts to Obtain Protease Expression

Again, activity assays were performed using Method 2.9.1 using protease directly from the protease extraction, and again no activity was found after 1 h digestion. An attempt was made at growing the cells in LB rather than M9. A time-course of expression was done (Method 2.1.1.14) and this revealed the system to be expressing a protein that did not present any proteolytic activity as assessed by digestion of the peptide substrate LSGAGFSW:

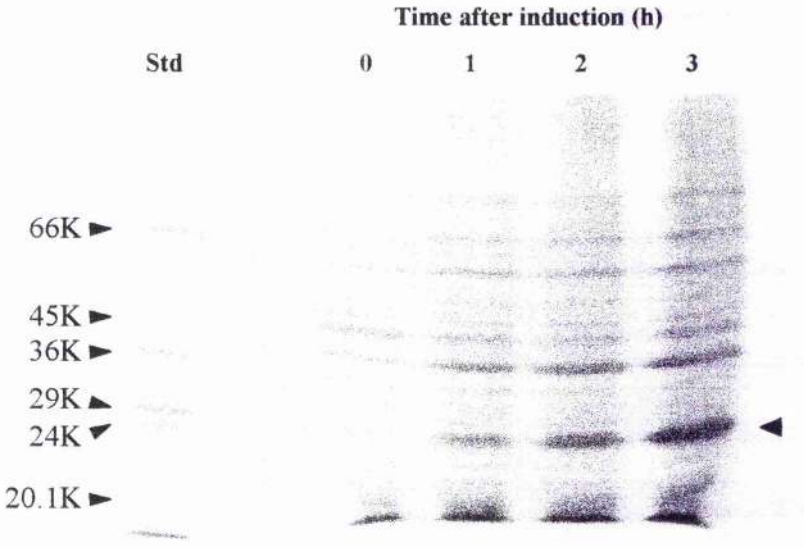


Figure 3.50 - Time course of induction of CELO protease in LB. Above is time after induction (h), on the left the molecular weight of the markers, and on the right the arrow marks the protein being expressed.

This protein seems to have a molecular weight of 24-29K, whereas the protease should have a molecular weight of 23.8K. This is not very different from the above, but every attempt to try and sequence the protein (Method 2.5.2) only revealed host proteins. The reasons for the lack of expression and/or activity were not found.

Possible explanations put forward include (a) the protease is not being expressed or is being expressed in a mutant inactive form; (b) the protease is being expressed in an insoluble form (a very common situation with over-produced proteins in *E. coli* (Grisshammer and Nagai, 1995)); (c) the protease is being expressed but the assay

conditions are not appropriate to enable efficient activation of the protease; or (d) the CELO protease is not able to cleave efficiently the substrate LSGAGFSW, possibly due to a different substrate specificity. Unfortunately, a thorough systematic approach to this problem was not possible.

3.8. Considerations on Some of the Methods Used

Several remarks should be made on some of the procedures used, namely about the reproducibility of the method of production and purification of the recombinant protease, the accuracy of the assessment of peptide concentration and the validity and limitations of the method developed for the determination of initial rates of digestion.

All these points are critically discussed and solutions for the problems encountered are proposed together with the assumptions made for their resolution.

3.8.1. Production of Protease

All the protease used in this work was of recombinant origin. It had been previously cloned in *E. coli* BL21(DE3) strains, using the pET-11 system (cf. 2.1.1) (Anderson, 1990). Several clones were available, which produced protease in different amounts and variable degrees of activity (cf. (Pollard, 1997)). Most of the protease used was kindly provided by Dr. M. Iqbal (in a concentration of 0.06 mg/ml, with a specific activity of 9.1 nmol of GFSW/min/nmol protease), produced by the conical flasks option described in Method 2.1.1.13.

However, as the conical flasks method did not allow for a reproducible source of protease, since experimental conditions such as temperature, aeration and agitation could not be conveniently controlled, the utilisation of a fermenter was also tried in order to try and circumvent these difficulties, as well as providing with a source of greater amounts of recombinant protease. This is also described in Method 2.1.1.13. The extraction procedure also had to be altered, mainly to accommodate for the larger amount of cells to be processed (cf. Method 2.1.2.1).

An attempt to improve the purification process in terms of speed and reproducibility after the extraction of the protease was also done, in order to try and provide with a more reliable source of protease. The main difference from the former method was to elute the protease from the DEAE column directly to the heparin and CM columns, rather than collecting fractions after the first column, pooling them and loading the resulting solution onto the last two columns overnight.

The resulting method consisted of loading the extraction supernatant (~35 ml) onto the DEAE column (40 × 1.6 cm) at a flow rate of 2 ml/min in purification buffer (25 mM Tris-HCl, pH 8.0), for 20 min. Then the heparin (5×1 cm) and CM (11×1 cm) columns were connected in this order to the DEAE column and elution continued at 1 ml/min for 90 min with the same buffer, in order for the protease to attach to these columns. The DEAE column was then disconnected, and the protease eluted at 1 ml/min with purification buffer 10% 1 M NaCl, which usually was collected ~30 min after the salt was applied.

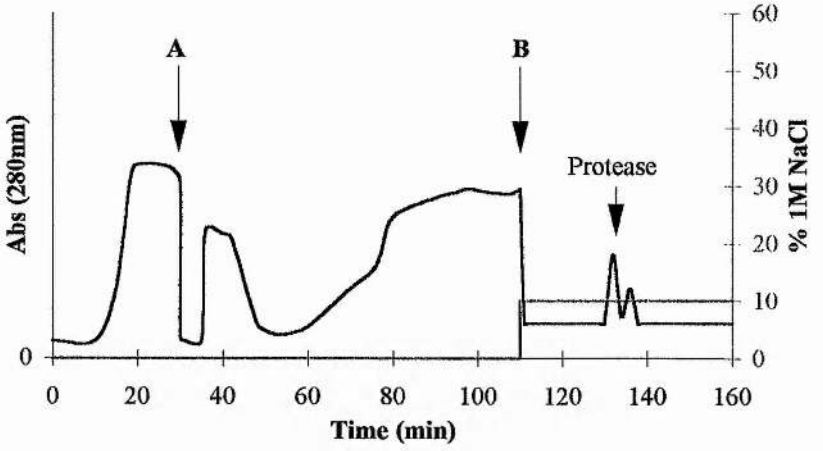


Figure 3.51 - Scheme illustrating a typical recorder plot of a single step protease purification. Point **A** corresponds to the connection of the heparin and CM columns, and point **B** to the disconnection of the DEAE column and increase in salt concentration (cf. text for explanation). From point **B** onwards, the absorbance scale has changed.

The resulting fractions were reasonably pure for assay purposes, with only one band showing on some lanes on SDS-PAGE. This procedure enabled a reduction of purification time from 2 days to one afternoon.

Figure 3.52 - SDS-PAGE of protease fractions purified using the single step purification method. The lane on the left are molecular weight markers, all other lanes are fractions for which a peak was observed at 280 nm.

After this method was established, a more thorough study was possible on the best source of protease to use for purity and activity purposes (Pollard, 1997).

3.8.2. Determination of Peptide Concentration

As a general feature, all the peptides used in this work were synthesised in small quantities, typically in amounts of less than 0.5 g. Therefore, the preparation of solutions and their subsequent concentration determination was not a straightforward process, due to the small amounts involved, which caused the weighing of the solid peptides to be quite a difficult task. A solution to this problem consisted of preparing a reference solution of known concentration of the peptide GVQSLKRRRCF, and preparing other solutions by dissolving a low amount of peptide, with subsequent comparison of concentration by one of two methods, reaction with fluorescamine (Method 2.7.2) or capillary electrophoresis (Method 2.7.1). The solutions were then diluted accordingly, until a uniform concentration was achieved with all the peptide solutions.

Fluorescence as a method for the determination of peptide concentration proved to be quite an accurate way of assessing the concentration of peptide solutions. The following formula was used to calculate the peptide concentration for a 20 µl sample prepared as described in Method 2.7.2.1:

$$C = 0.0028 \times F / S \qquad \text{Eqn. 3.8.1}$$

in which F is the recorder reading (mm), S is the selected sensitivity of the fluorimeter, and C is the concentration of the original sample (mM), prior to the dilution in the assay buffer. As previously described, fluorescamine reacts with free primary

amines in order to produce fluorescent derivatives (cf. 2.7.2). However, the fluorescence of the resulting derivative may vary depending on the primary amine that reacted with fluorescamine (Böhlen *et al.*, 1973; Stein *et al.*, 1973). Therefore, the equation above can only be used for the peptide GVQSLKRRRCF, from which it was derived. Yet, as the other peptides are reasonably similar to this, it was assumed that the final fluorescence would not be very dissimilar. For this reason, a second method was also used to provide with an additional peptide concentration method.

Capillary electrophoresis provided with a rough method of measuring the absorbance at 200 nm of a peptide solution. The resulting electropherogram presented a peak whose area could be related to the amount of peptide present. This relationship was based upon the assumption that the absorbance at 200 nm is roughly proportional to the number of peptide bonds in a peptide (Stoscheck, 1990), provided there is no tryptophan in the sequence, as this amino acid also absorbs at 200 nm due to the indole ring. The rough relationship found for a peptide solution diluted 1:10 and analysed using Method 2.7.1.1 was the following:

$$C = 1.6\times10^{-5} \times A / n \qquad \text{Eqn. 3.8.2}$$

in which A is the peak area as obtained from the integrator application (cf. 2.7.1.2), integrated at a 10 Hz rate, n is the number of peptide bonds in the peptide, and C is the concentration (mM) of the original peptide solution, prior to the 1:10 dilution. It should be stressed once more that this is a very rough method, usually used as a confirmation of concentration, as it is based on a very simplistic model.

3.8.3. Digestion Assays and Determination of Initial Rates

There were three steps involved in the determination of the initial rates of digestion of the adenovirus protease: the experimental method used, the collection and analysis of time point data, and the data treatment in order to calculate the initial rates. Each of these steps comprised certain assumptions as well as some errors that need to be assessed regarding the final validity of the values obtained.

3.8.3.1. The Experimental Method

The main feature of these assays was the small volumes involved, due to the expensive nature of the products utilised. This led to a greater source of experimental

error arising from the handling of small quantities, which was minimised by the repetition of the assays followed by a statistical treatment of the data obtained.

Another origin of variability was the protease preparations, whose activity was not constant, depending on the concentration of each preparation, rendering the measurement of absolute rates impractical. Furthermore, the protease is sensitive to temperature variations, and the thawing of aliquots yielded variable restoration of activity. Therefore, the use of controls and subsequent calculation of relative initial rates was the devised solution to circumvent this variability.

The low solubility of the substrate also posed two problems: (a) low measurable concentrations of the products of reaction, which made it more difficult to quantify the initial substrate digestion, which is discussed further below; and (b) the low substrate solubility imposed a working concentration very close to the K_M of the protease for this substrate (1-10 mM, H. Murray, unpublished results), which should also have contributed to the variability of results mentioned above.

3.8.3.2. The Collection and Analysis of Time Point Data

The collection of point data was made by diluting the sample from the reaction mixture in 1:10 diluted capillary electrophoresis buffer, which is an acidic buffer (pH 2.5) that inhibits the protease (cf. Method 2.9.1).

As mentioned above, the low substrate concentration in the assay made the detection of reaction products at early stages of digestion a troublesome task. To illustrate this problem, Figure 3.53 shows a typical electropherogram where digestion of substrate reaches about 30% of the initial quantity. In this case, it can be seen that all peaks are clearly defined, allowing a straightforward integration for the quantification of digestion. It should be noted that the product peaks exhibit different relative areas, which is due to the presence of tryptophan in the peptide GFSW, making its corresponding peak account for about 75% (average of ~2500 measurements) of the area of products, rather than an otherwise expected equitable proportion of both of the product peak areas.

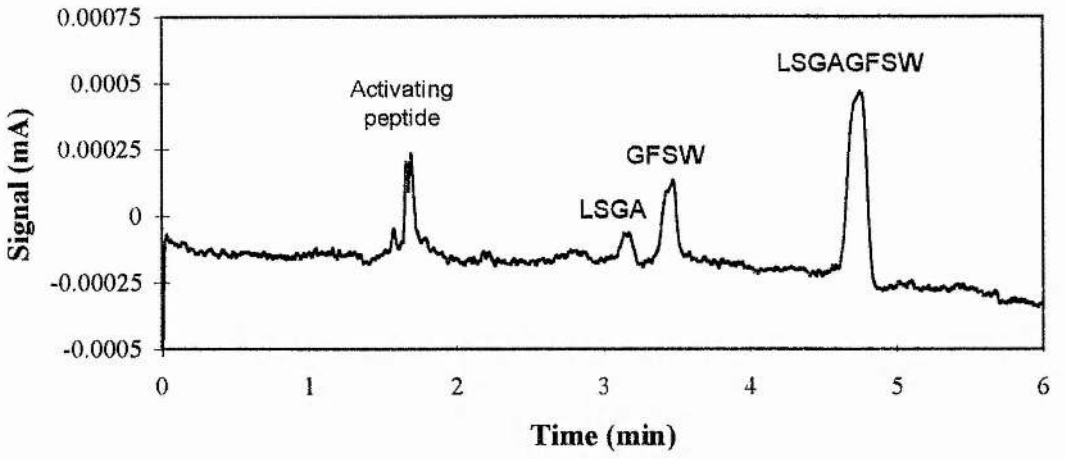


Figure 3.53 - Electropherogram of protease 10 min digestion of the substrate LSGAGFSW. The protease was activated with the peptide GVQSLKRRRCF (peak on the left). The relative areas of the peaks and respective migration times are: LSGA: 7.26%, 3.18 min; GFSW: 22.49%, 3.48 min; LSGAGFSW: 70.24%, 4.75 min.

If the peaks in Figure 3.53 are clearly defined, the situation is not so clear in Figure 3.54. This is a typical electropherogram of a low substrate digestion, where the attribution of peaks becomes more subjective, due to the small product peaks mingling with the background. In this situation, reference peak migration times must be used in order to delineate the limits of the putative peaks. Thus, if the undigested substrate peak is eluting at a time of 4.75 min (as shown in Figure 3.53) and the corresponding product peaks are eluting at 3.18 and 3.48 min (for LSGA and GFSW, respectively, cf. Figure 3.53), then in the electropherogram of Figure 3.54 and Figure 3.55 the peaks for the products are expected to elute at roughly the same corresponding times. This is indeed what happens, and although in the electropherogram in Figure 3.54 the peak for GFSW is still reasonably clear, the peak for LSGA would be difficult to discern, and when magnified (Figure 3.55) the peak can be identified by searching for a maximum near 3.18 min, which occurs at 3.25 min. Obviously, these peak attributions can lead to errors in the calculation of the relative areas, and the solution to this problem is proposed further below.

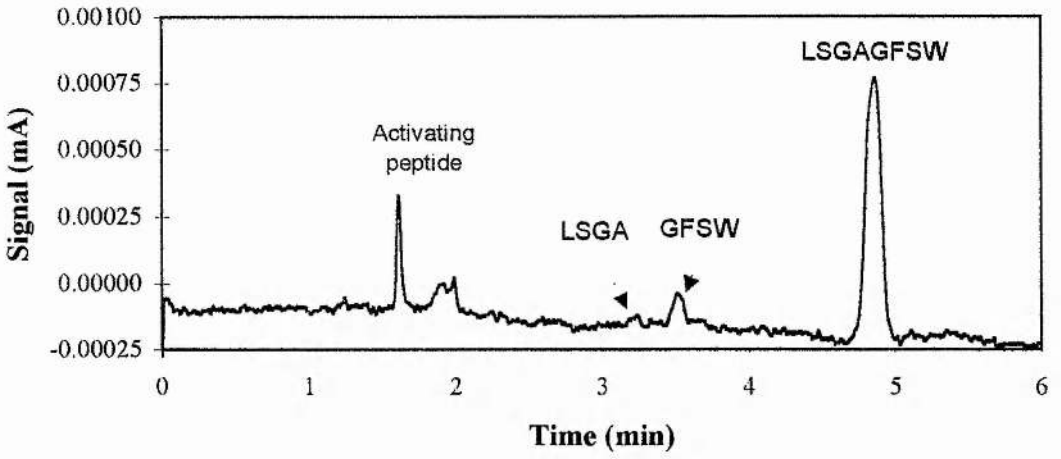


Figure 3.54 - Electropherogram of protease 30 min digestion of the substrate LSGAGFSW. The protease was activated with the peptide GVQSLKRRCRF (peaks on the left). The relative areas of the peaks and respective migration times are: LSGA: 2.34%. 3.25 min; GFSW: 7.30%. 3.52 min; LSGAGFSW: 90.35%, 4.87 min. The dashed rectangle is magnified in Figure 3.55.

Another remark has to be made regarding the presence of a peak (or several, as is the case in Figure 3.54) corresponding to the activating peptide used. This peak (which in Figure 3.54 corresponds to an activating peptide concentration of 0.15 mM in the assay) can also be used as a reference peak, as its position usually remains constant, although its shape may vary. Some analysis of these peaks was made, regarding the oxidation state of the peptide, and the results are discussed in 3.2.2.

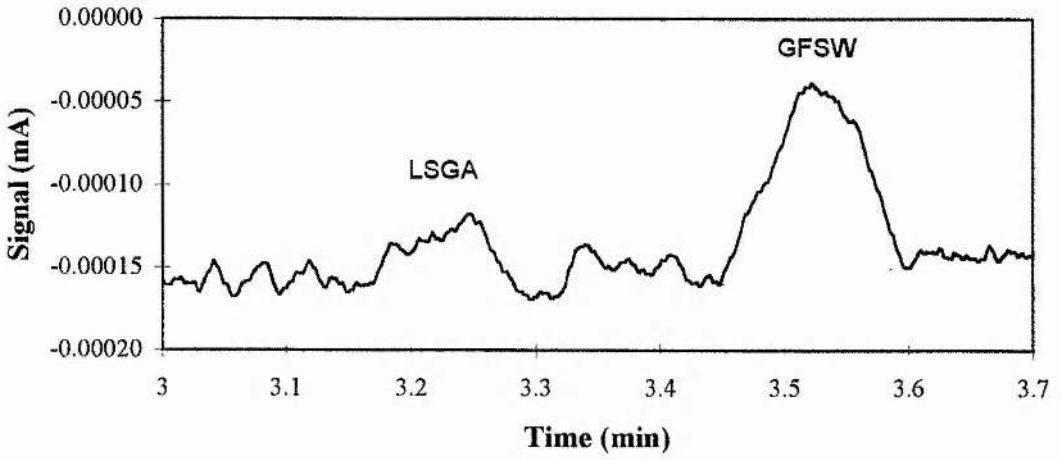


Figure 3.55 - Magnification of the peaks in the dashed rectangle of Figure 3.54. The chosen limits of integration are also represented with red lines.

One advantage of this method is that the determination of the amounts of products and undigested reagent can be directly measured from the same electropherogram, which

leads to a smaller error in the determination of these quantities, as a sampling of the reaction mixture by this method will always yield all the above quantities in one spectrum.

However, one assumption was made that the total area corresponding to the sum of the areas of the peaks LSGA, GFSW and LSGAGFSW remained constant. Therefore, the percentage of the total area corresponding to the peak LSGAGFSW was taken to be equivalent to the percentage of undigested peptide, a value subsequently used in the calculations described in Method 2.12 in order to calculate the initial rates from the electropherogram data, which is discussed below. This was actually observed not to be the case. In fact, the sum of the areas of all three peaks should decrease with increasing percentage of digested products, which can be explained in terms of the loss of a peptide bond that contributes to the absorbance at 200 nm. This decrease in the total area leads to a difference between the percentage of digestion calculated from the peak area of undigested LSGAGFSW and the "real" percentage of digestion.

In order to calculate what is the influence of this difference, the decrease in absorbance had to be calculated as a function of the percentage of digestion obtained from integration. For this purpose, the decrease in total area was plotted as a function of the percentage of the peak area of LSGAGFSW. However, due to experimental error, the amounts of substrate used varied between assays, and also the amount of sample taken varied within each assay, and therefore the total areas varied between as well as within each assay. Consequently, in order to normalise all the data, every total area within an assay for different times was plotted against the corresponding percentage of substrate (LSGAGFSW peak area). A least-squares regression was then performed to fit the decrease in total area with the decrease in percentage of substrate and this was used to calculate the extrapolated value of the total area at 100% substrate area (an example is shown in Figure 3.56). All the values within the assay were then divided by the above extrapolated value to yield a relative total area.

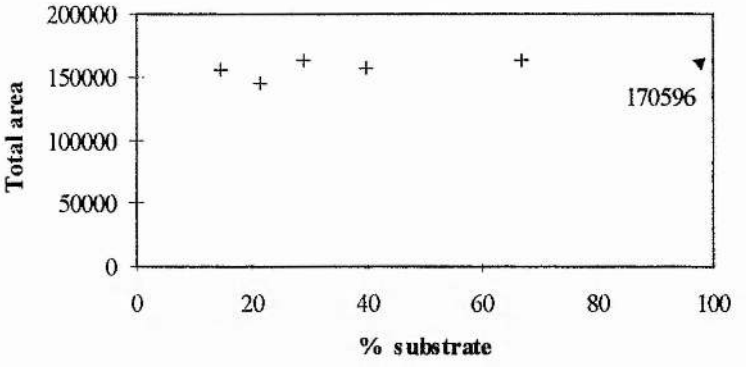


Figure 3.56 - Illustration of the normalisation process of total area values within a digestion assay. A linear regression is calculated in order to obtain an extrapolated value for the total area corresponding to 100% substrate (in this example, 170596).

Once all the values of total area were normalised, with all the extrapolated values for 100% substrate coinciding in the same point, every value from every assay was plotted together in order to calculate the overall trend. The result is shown below:

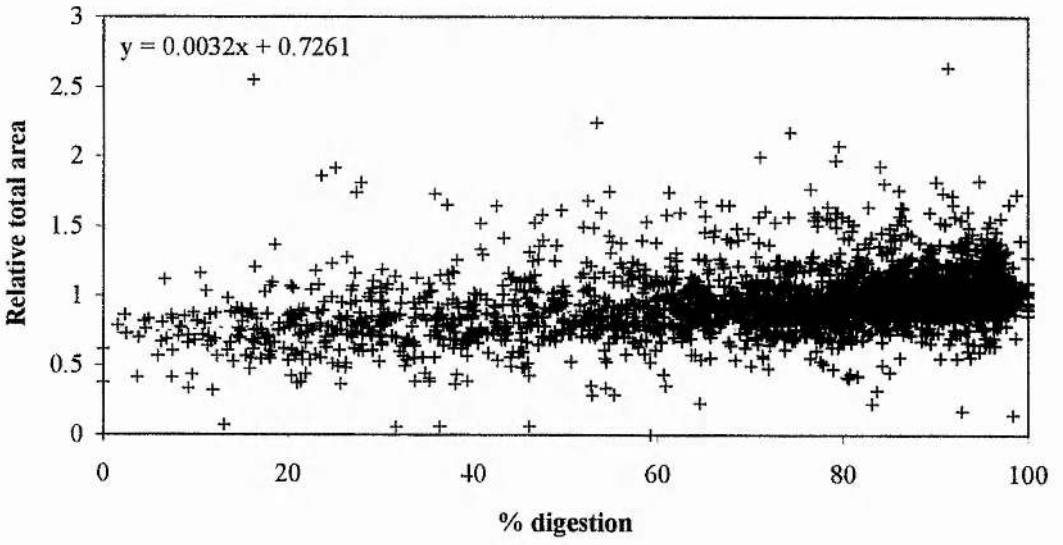


Figure 3.57 - Calculation of the overall trend of relative total area decrease in a digestion assay.

From the graph in Figure 3.57 it can be seen that there is indeed a decrease of total area where there is a greater extent of digestion (lower values for the percentage of substrate left), which agrees with the assumption made before. In order to correct the percentage values, a correction factor must be calculated with which to multiply the areas of the product peaks, so that the total area remains constant, thus allowing for the calculation of the "real" percentage of remaining substrate. The inverse of the value

obtained in Figure 3.57 for 0% substrate (1/0.7261 = 1.377) can then be shown to be the correction factor needed for the calculation of the correct value for the percentage of substrate, whose correction is calculated by the following equation:

$$Corrected\,\%S = \frac{100 \times \%S}{\%S + 1.377(100 - \%S)} \qquad \text{Eqn. 3.8.3}$$

Using Eqn. 3.8.3, the difference between the value for the percentage obtained by integration (area of LSGAGFSW peak) and the "real" percentage can be calculated, and this can be plotted as a function of the area peak percentage:

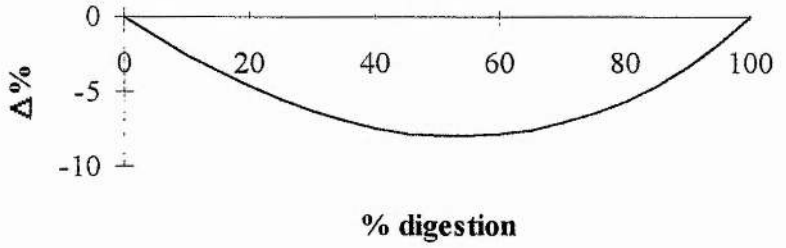


Figure 3.58 - Plot of the difference between percentage of substrate obtained by peak area integration and correction for constant total area.

From the above graph it can be seen that the greatest difference between the values for the percentage of undigested substrate obtained by direct measure of the proportion of the LSGAGFSW peak area to the total area and the "real" percentage is of -8% units, for 50% digestion of substrate. This means that all the values for the percentage of undigested substrate were overestimated to a variable extent, depending on the percentage of the substrate peak area in relation to the total area. The implications of this result are discussed below.

3.8.3.3. Data Treatment

The mathematical treatment done to the experimental data was already described in Methods 2.12 and 2.12.1. As described in these methods, the kinetics assumed were of a first order reaction, and the reasons for this choice were given therein. It should be noted that this assumption was supported by the close fitting of the experimental points, as the correlation coefficient will show further below, thus giving it more credence.

The first topic to be addressed in the treatment of data was the selection of the points to be considered and those to be discarded in the calculation of the slopes for the

determination of the initial rates. As mentioned before, low digestions of the substrate yielded small product peaks and this could cause its smear with the background, making it difficult to delineate the limits of peak areas. The proposed solution around this problem was to plot all the digestion data points for a certain assay against their time of digestion, fit a digestion curve and discard any points that seem to lie outside the general trend. This is illustrated in Figure 3.59. As it can be seen, the point for 5 min digestion can be easily discarded. However, the point corresponding to 20 min is not so easily discerned.

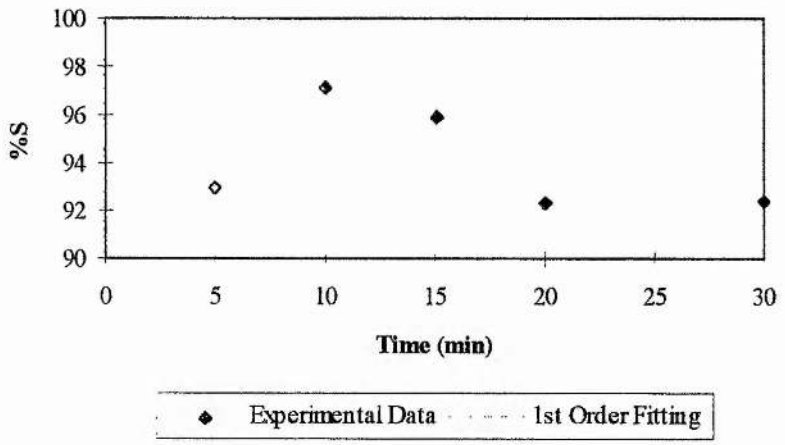


Figure 3.59 - Example of low digestion data. The scattering of the data could be attributed to a poor definition of the product peaks in the electropherograms, as mentioned above (3.8.3.2). The open point was discarded in the calculations. This resulted in a slope of -0.0030 %S/min, with $r^2 = 0.7321$.

Yet, if the 20 min digestion point is also discarded, the fitting is almost exact, as it can be seen from Figure 3.60:

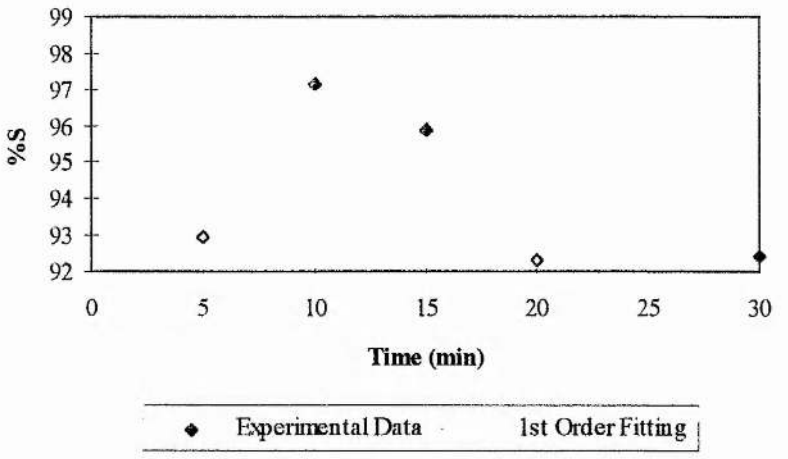


Figure 3.60 - Example of low digestion data. When the point for t = 20 min in Figure 3.59 is also discarded, the resulting slope has a value of -0.0027 %S/min, with $r^2 = 0.9925$.

Therefore, even for low digestions it is possible to obtain reasonably reliable values for initial rates of digestion, provided enough points are collected in order to discern a trend and to select which are the more reliable points to use in the calculations. Hence the collection of at least four time samples in each assay, and usually five.

Ā second point to be addressed was the necessity for a correction factor for the calculation of the "real" percentage of digestion. The following three graphs illustrate the influence of the correction in absolute terms, that is, in each individual calculation of initial rates of digestion. The first graph in Figure 3.61 shows the change in the percentage of undigested substrate (%S) for low amounts of digestion.

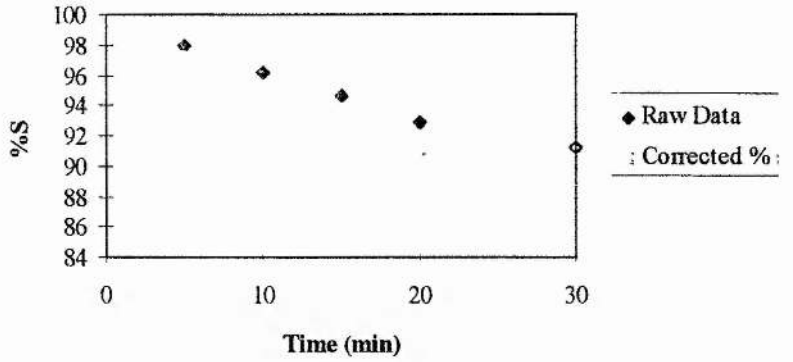


Figure 3.61 - Influence on the correction of data points with the use of Eqn. 3.8.3 for a final digestion of ~10%. The slopes at t = 0 of the dashed curves calculated as described in Method 2.12.1 are -0.00373 %S/min (r^2 = 0.9931) for the original points (Raw Data) and -0.00508 %S/min (r^2 = 0.9920) after the correction, respectively. The open points at 30 min were discarded in the calculations.

The first point to be noted is that at these low rates the digestion curve is quite linear, as it should be expected. The increase in the value of the slope can be calculated to be of about 36%, which is quite considerable. In fact, it can be shown that for (theoretically) extremely low rates of digestion this increase should tend to 37.7%, a value that corresponds to the expected increase arising from the correction factor calculated in 3.8.3.2 (equal to 1.377). Curiously, the quality of fitting of the experimental points to the analytical curve is better in the raw data than in the corrected form, as it can be judged from the values for the correlation coefficient (r^2). This can be attributed to a fortuitous linear behaviour of the raw data that is deformed by the adjustment made with the correction factor.

The second graph, in Figure 3.62, illustrates the variation in the percentage of undigested substrate (%S) for intermediate amounts of digestion.

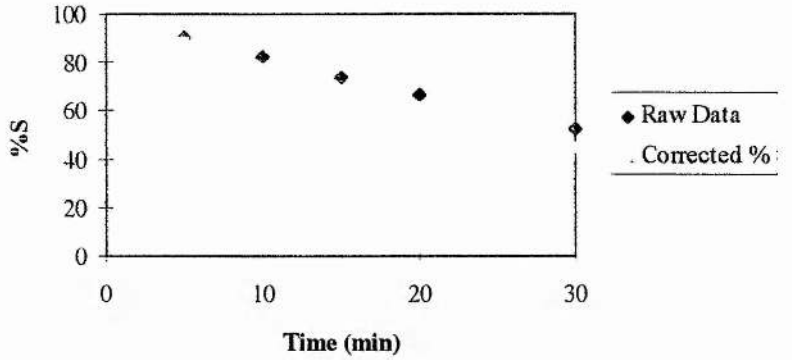


Figure 3.62 - Influence on the correction of data points with the use of Eqn. 3.8.3 for a final digestion of ~50%. The slopes at t = 0 of the dashed curves calculated as described in Method 2.12.1 are -0.02086 %S/min (r^2 = 0.9953) for the original points (Raw Data) and -0.02664 %S/min (r^2 = 0.9992) after the correction, respectively.

In this situation, the increase in the slope has a value of 28%. As the values of the initial rates increase, the difference between the raw data and the corrected values start to decrease. This is observed to a further extent in the next graph, in Figure 3.63.

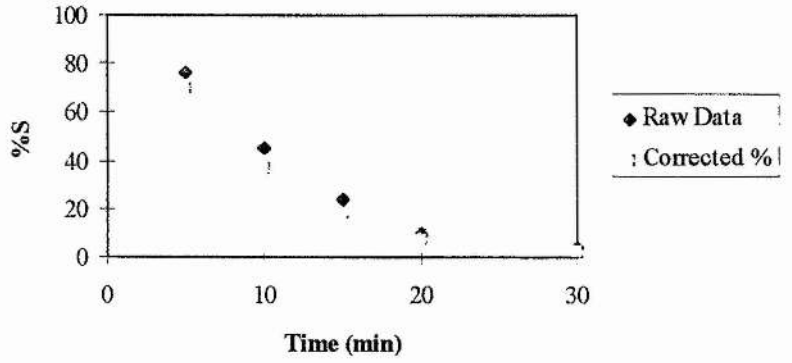


Figure 3.63 - Influence on the correction of data points with the use of Eqn. 3.8.3 for a final digestion of ~95%. The slopes at t = 0 of the dashed curves calculated as described in Method 2.12.1 are -0.10737 %S/min (r^2 = 0.9634) for the original points (Raw Data) and -0.12026 %S/min (r^2 = 0.9752) after the correction, respectively.

For reasonably high rates of digestion, as in the example above, the difference is even more attenuated, reaching in this case an increase of 12%. It can indeed be proved that for higher values of initial rates there is a tendency for the difference to disappear. Therefore, the maximum increase between the raw data (from the percentage of the substrate peak) and the corrected data occurs at low digestion rates. It is then important to know what is the effect of this difference in the calculation of relative initial rates.

The calculation of relative initial rates consisted of obtaining the initial rate of digestion when the protease is activated with the peptide being studied, and dividing it by the initial rate of digestion for the protease being activated with the native activating peptide (GVQSLKRRRCF), thus yielding a quotient which should express the relative capability of activation of the peptide in study.

In order to compare the influence of the correction factor in the above calculations, the relative initial rate obtained between the two most extreme cases, of high and low digestion can be compared, when using the raw data and the corrected results. Therefore, the initial rate using the raw data would be -0.00373/-0.10737 = 0.0347, and using the corrected data it would be -0.00508/-0.12026 = 0.0422. Therefore, in the most extreme situation, there is a relative increase of 26%. However, when comparing two initial rates not too dissimilar, like those obtained for low and intermediate digestion, -0.00373/-0.02086 = 0.179 and -0.00508/-0.02664 = 0.191, a relative increase of 6.7%. The remaining situation, -0.02086/-0.10737 = 0.194 and -0.02664/-0.12026 = 0.222, with a relative increase of 14%, shows that in relative terms the increase is not negligible. But in absolute terms the difference in values will be expressed by the statistical treatment given to these values, and will be subsequently covered by the standard deviation of the sample. In other words, saying that a peptide activates only to an extent of 3.5% or 4.2% (using the most extreme situation) does not have any relevance if the standard deviation covers this difference. Therefore, as the correction factor was a value obtained *a posteriori*, it was deemed unnecessary to use it in the calculations, although it is presented as a more correct approach to the calculations of relative rates.

4. Conclusions

The adenovirus protease, as most viral proteases (Hellen and Wimmer, 1992), is involved in the control of particle assembly and subsequent virus maturation. In order to prevent premature cleavage of protein precursors involved in these processes, it has to present a control mechanism. This control was at first thought to happen through auto-catalytic processing (Chatterjee and Flint, 1987), but this idea was later dismissed (Webster and Kemp, 1993), and the control shown to involve the action of the activating peptide GVQSLKRRRCF (Webster *et al.*, 1993) and possibly also viral DNA (Mangel *et al.*, 1993).

In the present work, the mechanism by which the activating peptide enables the activation of the protease was studied. From a model (cf. Figure 3.3) in which it was supposed that the role of the activating peptide was to promote a disulphide bond rearrangement within the protease (Webster *et al.*, 1993), acting only as a catalyst similar to protein disulphide isomerase (Noiva and Lennarz, 1992) or possibly to thioredoxin (Holmgren, 1985), it became increasingly evident that the peptide was activating the protease by means of covalently binding to it.

However, before discussing the mechanisms by which the protease is rendered active, it is deemed necessary to actually try and define what is understood by an active protease.

4.1. What is an Active Protease?

When protease activity is discussed, the threshold between active and inactive protease always seems to be a choice based on the argument that is trying to be made. Whereas recombinant protease expressed in insect cells or *E. coli* is considered to be fully active when it digests ~50% of ovalbumin or less than 50% of canine adenovirus pVII in 18 h in the absence of activating peptide (Keyvani-Amineh *et al.*, 1995b), the same protease is considered to be effectively inhibited when processing of human adenovirus type 2 pVII is reduced to ~10% digestion in the presence of cystatin, a cysteine protease inhibitor (Sircar *et al.*, 1996).

Also, the recombinant protease is considered to be inactive without activating peptide if no digestion of the peptide MSGGAFSW is observed after 12 h, although it digests >50% pVII in 3 h (presumably due to the presence of pVI in the cell extract), and

active in the presence of activating peptide if it digests <50% of the above peptide in 30 min and also digests >50% pVII in 3 h (Webster *et al.*, 1993).

Clearly, the protease has a different affinity for different substrates and therefore an absolute definition for activity should be referred to a common assay system. In the assay system used throughout this work (Method 2.9.1), the protease (0.15 mg/ml) was found to digest ~10% of the substrate LSGAGFSW in 3 h without activating peptide, whereas in the presence of 0.15 mM of activating peptide it digested ~85% of the substrate in 10 min, indicating an increase of initial rate of more than 300 fold (average results of three assays).

On the other hand, the protease concentration is obviously also of importance, and as it is not mentioned in any of the above results, the comparison of activity between different systems becomes virtually impossible. However, if constant protease concentration is assumed, as well as a similar substrate affinity (a very rough assumption), it can be seen from the graph below that some 'inactive' proteases present higher digestion rates than others considered to be 'active'.

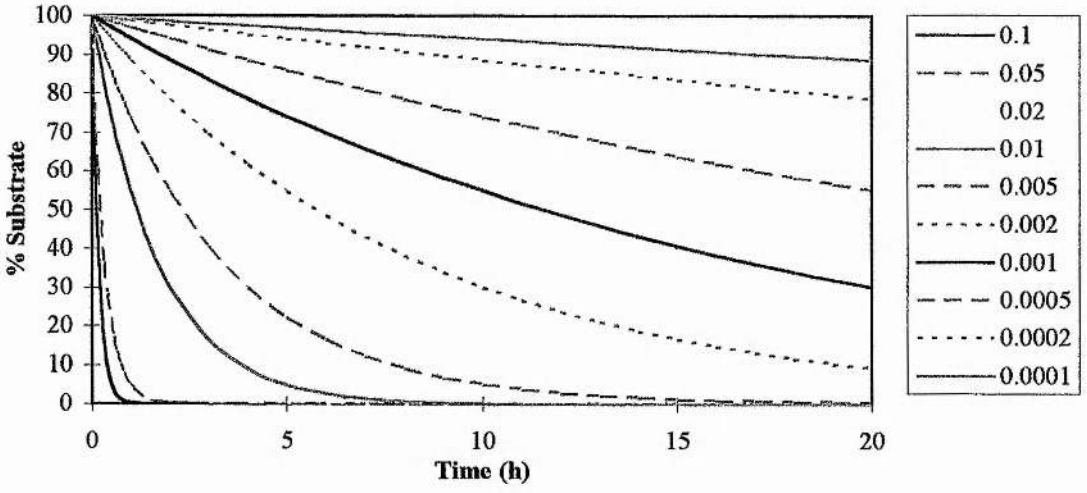


Figure 4.1 - Theoretical plot of substrate digestion curves for different values of initial rates (%S/min), based on first order digestion kinetics (cf. Method 2.12).

Indeed, a ~50% digestion of albumin in 18 h would correspond to an initial rate of ~0.0005 %S/min for an 'active' protease, whereas an 'inactive' protease that digests ~10% of LSGAGFSW in 3 h would have roughly the same initial rate of ~0.0005 %S/min.

Two main conclusions should be drawn from the above observations: (a) the protease does present some proteolytic activity without the presence of activating

peptide, and this activity can arbitrarily be termed 'residual'; (b) it is commonly agreed that the presence of the activating peptide does increase considerably the proteolytic activity of the protease (Diouri *et al.*, 1996; Mangel *et al.*, 1993; Webster *et al.*, 1993).

Therefore, it was deemed convenient in the discussion below to term the protease in the absence of activating peptide as '*inactive*', and when in its presence considered as '*active*', and the process by which the activating peptide activates the protease was termed '*activation*'. Furthermore, the activity of the protease seems to be also either enhanced or even achieved through other means, and this will be discussed below (cf. 4.10).

4.2. Which Cysteines in the Protease are Involved in Activation/Activity?

When the activating peptide was found to activate the adenovirus protease (Webster *et al.*, 1993), in the 11 protease serotypes known by then three cysteines were found to be fully conserved: cysteine-104, 122 and 126 (numbering as referred to the protease from human adenovirus type 2, cf. Figure 1.8). It was then proposed (Webster *et al.*, 1993) that the activation of the protease would possibly involve a rearrangement of an intramolecular disulphide bridge in which the activating peptide would act as a catalyst of a thiol-disulphide interchange (cf. model in Figure 3.3).

However, some subsequent findings started to suggest that this model could not correspond to the activation mechanism: of the eight cysteines in the adenovirus protease, no internal disulphide bonds were found between them (Tihanyi *et al.*, 1993), and the avian adenovirus type 1 protease sequence did not present a conserved cysteine-126 (Cai and Weber, 1993).

Although an initial attempt to classify the protease based on inhibitor studies positioned it among serine proteases (Tremblay *et al.*, 1983), more extensive inhibitor studies (Webster *et al.*, 1989b) suggested that it was rather a member of the cysteine class, albeit with an unusual mechanism, resembling in its inhibition profile the 3C protease of rhinoviruses (Orr *et al.*, 1989).

Furthermore, site-directed mutagenesis studies (Grierson *et al.*, 1994) confirmed the protease to be a cysteine protease, as mutants containing altered serines originally conserved in all of the known serotypes were active, whereas mutants of either cysteine-104 or cysteine-122 presented its activity reduced by over 95%. Other

independent site-directed mutagenesis studies (Rancourt *et al.*, 1994) also showed that histidine-54 or cysteine-104 mutants did not present any activity, although mutation of cysteine-104 to a serine interestingly turned the protease into a serine-like protease in its inhibitory profile. Curiously, these studies revealed that mutation of cysteine-122 into a glycine yielded an active mutant, which would seem to indicate cysteine-104 as the active-site cysteine (Weber and Tihanyi, 1994). However, this was later confirmed not to be the case, as cysteine-122 mutants were actually shown to be cysteine-67 mutants (Weber, personal communication).

Expression of protease mutants in insect cells (Jones *et al.*, 1996) showed however that cysteine-104 did not seem to be essential for proteolytic activity, as this mutant was able to cleave pTP (a viral protein substrate) to iTP, whereas neither of the mutants of histidine-54 or cysteine-122 were able to cleave that substrate. Moreover, when the activating peptide was incubated with recombinant protease, this caused a concentration-dependent increase in tryptophan fluorescence, concomitant with an increase in proteolytic activity, indicating a conformational change of the enzyme when the peptide was added. Also, the incubation of the cysteine-104 protease mutant with the activating peptide did not cause an increase in tryptophan fluorescence, nor did the incubation of wild-type protease with GVQSLKRRRSF or GVQSLKRRRDF, which seemed to conform to cysteine-104 being involved in activation and cysteine-122 with catalysis.

Interestingly enough, labelling experiments of the protease (Iqbal, unpublished results) with radioactively labelled iodoacetamide only revealed labelling of cysteine-104 and cysteine-122, in spite of the existence of another 6 cysteines in the protease, which suggested that either these two cysteines are unusually reactive or the other six are reasonably unreactive, or possibly both. This certainly has been observed before, for cysteine groups have been divided into three groups according to their reactivity (Hellerman *et al.*, cited in (Torchinsky, 1981)): readily reacting, sluggishly reacting, and "masked" or "buried". The causes for this have been ascribed to: activation of thiol groups by neighbouring functional groups such as histidine (Hammond *et al.*, cited in (Torchinsky, 1981)), and "masking" of thiols due to them being sterically inaccessible (Mirsky, cited in (Torchinsky, 1981)) or forming intramolecular bonds, such as thiazolidines or thiazolines (Linderström-Lang *et al.*, cited in (Torchinsky, 1981)), thioesters (Mastin *et al.*, cited in (Torchinsky, 1981)), or simply involvement in hydrophobic interactions (Klotz, cited in (Torchinsky, 1981)).

4.3. What is the Role of the Cysteine in the Activating Peptide?

The first evidence of the importance of the cysteine within the activating peptide came from an experiment (Webster *et al.*, 1993) where an incubation of the protease with the peptide GVQSLKRRRAF was not able to render the protease active. Furthermore, it was also shown that the peptide GVQSLKRRRCF activated the protease more efficiently (10 fold increase) in the oxidised (dimeric) form than in the reduced (monomeric) form.

Results presented here (cf. 3.2) tried to assess how did the replacement of the cysteine in the activating peptide affect the ability of the activating peptide in rendering the protease active. It was initially thought that the cysteine had some particular reactivity which would enable the activation of the protease, and thus some substitutions of that residue aimed at determining what was that peculiar property. Substitution for a serine, quite similar to cysteine in both polarity and structure, although not as reactive and not allowing the formation of disulphide bridges (thus acting as an analogue of the reduced form of the activating peptide, albeit less reactive), aimed at assessing the importance of nucleophilicity; and aspartate, a residue analogous in structure and function to a sulphenic acid (RSOH), which has been observed to contribute to redox control involving cysteine residues (Abate *et al.*, 1990), intended to mimic a possible redox control mechanism, although aspartate, unlike sulphenic acid, can only be reduced. The peptide GVQSLKRRRAF was used as a negative control.

The results shown in Figure 3.2 show that not only is cysteine the only residue at that position that enables an efficient activation of the protease, but also the peptide GVQSLKRRRAF, which was supposed to act as the negative control, is more effective than any of the others in activating the protease, thus enabling the conclusion that cysteine has a particular property that enables the activation, possibly by being able to either participate in a thiol-disulphide exchange or by binding itself to some cysteine in the protease. This, together with the results presented previously (cf. 4.2), seemed to point at an interaction between cysteine-10 of the activating peptide with possibly cysteine-104 in the protease.

If cysteine-10 of the activating peptide was responsible for activation, was it particularly reactive? To address this question, determination of the pK_a values of the peptide was done to assess whether the cysteine was unusually acidic (cf. 3.2.3), and this revealed (cf. 3.2.3.8) the latter hypothesis not to be sustainable. The arginines beside the

cysteine did not alter significantly its acidity either, nor the mutation of the phenylalanine to an alanine. Peptides with the cysteine shifted along the arginines also did not activate the protease efficiently (cf. 3.5.1), confirming that there had to be another factor involved in the activation mechanism, which will be dealt with further below (cf. 4.5).

4.4. Does the Activating Peptide Bind to the Protease?

All the results presented above were not able to elucidate what was the nature of the interaction between the activating peptide and the adenovirus protease. In order to address this question, several approaches have been tried. As described before (cf. 3.6.1), radiolabelling of the peptide followed by incubation with the protease and analysis with non-reducing SDS-PAGE revealed some radioactivity on the band corresponding to the protease (Kemp, unpublished results). Similarly, incubation of the activating peptide with the protease followed by low acrylamide percentage non-reducing SDS-PAGE, showed a band corresponding to the putative protease·peptide complex running at an higher molecular weight than the corresponding band to protease alone (Iqbal, unpublished results). Also, reduced protease was observed to migrate unexpectedly faster than its oxidised counterpart (Greber *et al.*, 1996).

However, the definite proof came from mass spectrometry (cf. Figure 3.43, (Cabrita *et al.*, 1997)), where activating peptide incubated with the protease showed an increase on the molecular weight of the protease consistent with a 1:1 protease·peptide complex. The fact that the peptide GVQSLKRRRSF when incubated with the protease did not cause an increase in the molecular weight of the protease agrees with the assumption that the binding of the peptide to the protease is made via a disulphide bond.

Yet, although the peptide KRRRCF has a cysteine, it did not activate the protease (Webster *et al.*, 1993), which indicated that something else was missing in this peptide in order to produce an effective activation.

4.5. How is the N-terminal Region of the Activating Peptide Important?

The fact that KRRRCF was not able to activate the protease prompted an investigation into what was the minimum length required for an effective activation of the protease by the activating peptide.

The results determined here (cf. 3.3) showed that the removal of the N-terminal glycine of the activating peptide causes a reduction in protease activity of about 50%, whereas by removing both the glycine and the valine reduced the protease activity to 5%

of the value achieved with the native activating peptide. Further removal of more residues had a similar effect to that of removing glycine and valine. Also, as the mass spectra of the N-terminal truncated peptides (cf. Figure 3.43) showed that these peptides do bind to the protease, this leads to the conclusion that the N-terminal GV amino acid sequence in the activating peptide is essential for efficient activation of the protease.

These results were further confirmed when mutations of the valine in position 2 to either threonine or alanine, or an acetylation of the N-terminal, in the activating peptide (cf. 3.4) induced a reduction of the protease activity to 50% or less than the activity obtained with the native peptide, which shows that these two residues are quite specific for activation requirements to be met. This agrees with sequence data for the activating peptide (cf. Figure 1.9), in which glycine is conserved in all of the known serotypes and valine is only conservatively substituted by leucine.

This is in agreement with the tridimensional structure of the protease·peptide complex recently published (Ding *et al.*, 1996):

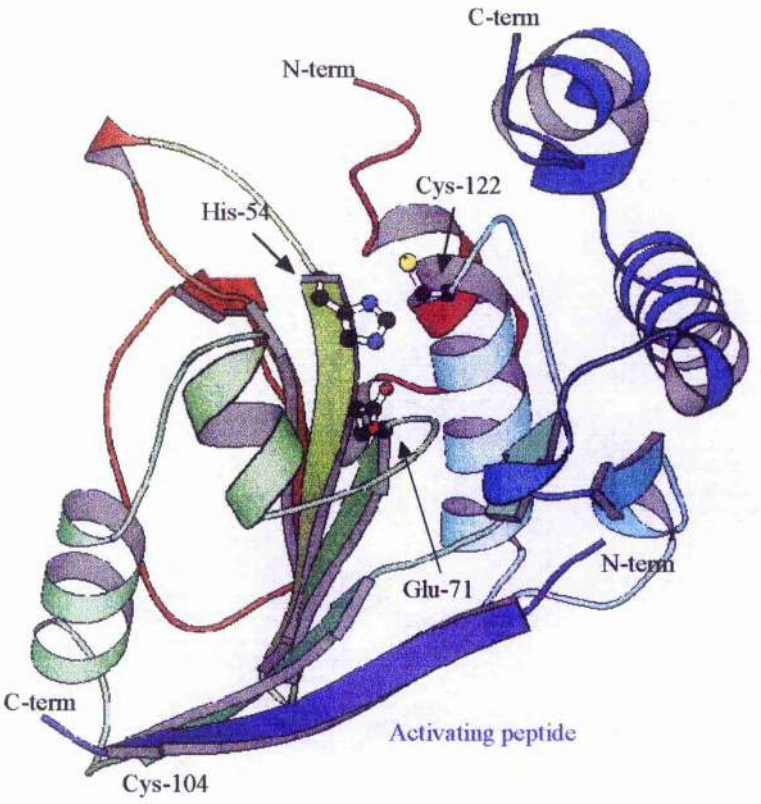


Figure 4.2 - Ribbon scheme of the tridimensional structure of the adenovirus type 2 protease complexed with the activating peptide (Ding *et al.*, 1996).

From the above structure, it can be seen that indeed the activating peptide binds to cysteine-104 via its cysteine-10, and the N-terminal of the activating peptide is enfolded in a loop region of the protease in the region of residues 135-140, with contacts between the N-terminal glycine of the activating peptide and an aspartate at position 142. This loop region immediately follows a region where cysteine-122 lies at the far end of a helix. It is then quite possible that the mechanism through which the activating peptide activates the protease is by "tying together" two minidomains of the protease that in its absence are able to move independently. This would have the effect of stabilising the catalytic residues in their optimum orientation, which would require interaction from both ends of the activating peptide. Therefore, the presence of a GV amino acid sequence in the N-terminal at a fixed distance from the cysteine residue seems to be critical for an efficient activation of the protease by the activating peptide. This would explain why the cysteine-shifted peptides (cf. 3.5.1) were not able to activate the protease to a reasonable extent, as the aforementioned distance was shortened.

This model also allows for an explanation as to why the protease presents a 'residual' activity in the absence of activating peptide, for the dynamic nature of the protease would enable a small proportion of protease molecules to be in the active conformation at any one instant.

These results together with the resolved tridimensional structure of the protease also explain why the peptides IVALGVQSLKRRRCF and VVGLGVQSLKRRRCF do not activate the protease to the same extent as the native activating peptide (cf. 3.4.2.1), as the extra sequence at the N-terminal would have a similar effect to that of the acetylation.

4.6. *Why is there Always an Aromatic Residue at the C-Terminal of the Activating Peptide?*

An unexpected result was the observation that the peptide GVQSLKRRRCA was capable of activating the protease only to 35% or less of the activity achieved by the native activating peptide (cf. 3.5.4). Indeed, this peptide seems to present all the previously mentioned requisites for an efficient activation of the protease: (a) a cysteine in position 10, (b) a GV amino acid sequence at the N-terminal, and (c) eight residues between GV and the cysteine. Therefore, the phenylalanine to the right of the cysteine must have an effect on the activation, presumably in the binding process, due to it being so close to the cysteine. However, the pK_a of the cysteine did not seem to be significantly altered (cf. 3.2.3.7), which ruled out the possibility of a change in acidity of the cysteine.

It can indeed be noticed (cf. Figure 1.9) that an aromatic residue is always present at the C-terminal of the activating peptide, which gives further credence to the assumption that this residue must play a particular role in the binding process. Interestingly, it has been reported (Bodner *et al.*, 1980; Morgan and McAdon, 1980; Morgan *et al.*, 1978) that the proximity of aromatic residues to cysteines enables the formation of sulphur-π-bonded chains, which confer extra stability to the folding of proteins, as disulphide bonds in proteins usually present a dihedral angle of 90° which takes on partial double character (Boyd, cited in (Morgan *et al.*, 1978)), and thus π-π bonding could explain the tendency of cystine disulphide pairs to interact with aromatic rings. On the other hand, a reaction of thiol-disulphide interchange (which presumably will happen between the oxidised activating peptide and cysteine-104 of the protease) is reversible (Eldjarn and Pihl, 1957; Keire *et al.*, 1992; Kolthoff *et al.*, 1955; Shaked *et al.*, 1980). Therefore, the vicinity of

an aromatic residue to the cysteine would presumably stabilise the formation of the disulphide, thus maintaining it bound to the protease, keeping it active. If this aromatic residue is missing, the binding of the peptide to the protease will be less stable and therefore more prone to detachment of the activating peptide. This could also be the reason why the peptides GVQSLKRRCRF and GVQSLKRCRRF were not able to activate the protease (cf. 3.5.1) nor inhibit it (cf. 3.6.3).

An examination of the tridimensional structure of the protease would allow to locate where the aromatic ring of the phenylalanine is binding to in the protease molecule, thus permitting the confirmation of the above mentioned stabilising effect. Unfortunately, although the structure has been solved for more than a year, its co-ordinates still have not been made publicly available.

Mass spectrometry of possible protease·peptide complexes between the adenovirus protease and the peptide GVQSLKRRRCA (or GVQSLKRRCRF, and GVQSLKRCRRF), as well as tryptophan fluorescence change measurements, would also benefit the understanding of this mechanism.

4.7. *Proposal of a Model of Activation of the Adenovirus Protease*

Based on all the results presented above, a model was devised which tries to encompass all the observations made about the protease:

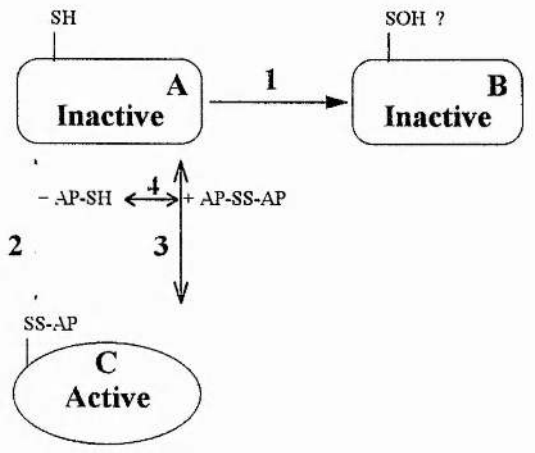


Figure 4.3 - Proposal for a mechanism of activation of the adenovirus protease. AP-SH denotes the reduced activating peptide. AP-SS-AP the oxidised activating peptide. the round rectangles the inactive protease and the ellipse represents the active protease. See text for explanation.

When the protease is newly synthesised, it is produced in the inactive form **A**. This molecule presents a thiol group (cysteine-104) in the reduced form. This cysteine should

be reasonably reactive, due to being present at the surface of the molecule. Therefore, when purified the protease is quite prone to irreversible oxidation via pathway **1** to an inactive (and inactivatable) form **B**, arguably by forming a sulphenic acid. Possibly, cysteine-122 in the active site should also be more prone to oxidation if the catalytic site is not in the right conformation. This oxidation can be prevented by adding a reducing agent such as β-MeOH or DTT to the buffer. This is a process that should only happen *in vitro*, as *in vivo* the protease should be protected from oxidation, due to the reducing environment in the cytosol (Greber *et al.*, 1996).

However, the conversion of the protease to the active form **C** requires the binding of activating peptide (AP). This can occur more slowly via pathway **2** by reacting with reduced activating peptide (as it will require the presence of an oxidant, which is being removed if a reducing agent is added), or via pathway **3** by thiol-disulphide interchange with the oxidised (dimeric) activating peptide. Either of these two reactions should be equilibria demanding a large excess of the activating peptide in order for it to bind to the protease. Yet, once the activating peptide binds to the protease, the aromatic residue should stabilise the disulphide bond and thus protect reasonably against subsequent detachment, as well as the N-terminal. On the other hand, the activating peptide itself should exist in an equilibrium between its reduced and oxidised forms (pathway **4**).

If a reducing agent is added before the activating peptide, the observed inhibition (Webster *et al.*, 1993) could be explained in terms of the activating peptide being reduced in pathway **4** and its oxidation leading to activation being prevented. However, a reducing agent added after activation should help maintain the active site cysteine in a reduced form.

A poorer activation with reduced activating peptide as compared to the oxidised form could be due to the fact that a slower rate of activation via slow pathway **2** would allow more protease to become inactivatable (inactive **B**) through pathway **1**, whereas the oxidised activating peptide would react reasonably quickly via thiol-disulphide interchange through pathway **3**. The same explanation could account for the lack of increase in activation once more activating peptide is added after an initial amount was allowed to incubate. A possible verification of this would be to use degassed buffers throughout the purification and assay procedures, in which no difference other than the time taken to activate should be observed in the activation.

It was also observed (Iqbal, unpublished observations) that a removal of the excess unbound activating peptide did reduce the activity, which can be explained in terms of the equilibria **2** and **3** being moved towards the release of activating peptide from the protease. It would be interesting to find out whether a subsequent increase of the activating peptide concentration would restore activity in oxygen-free conditions.

In terms of the actual binding of the activating peptide to the protease, the data presented here do suggest that the binding is reversible (cf. 4.9). Therefore, the mechanism by which the peptide can remain reasonably attached to the protease is by a combined effort of the N-terminal (shorter peptides do not bind as efficiently as the full length activating peptide, cf. 4.5), the cysteine, which provides with the covalent bond (cf. 4.3) and the aromatic residue close to it which should protect the disulphide bond from reduction (cf. 4.6).

The specificity in binding of the activating peptide to the protease and consequent conformational change is consistent with the behaviour observed for the protease of the adenovirus type 2 temperature-sensitive mutant (ts1, (Weber, 1976)). This mutant was shown to present a protease with a mutation from proline to leucine at position 137 (Weber and Houde, 1987) that prevents proper encapsidation of the protease within the newly formed viral particles (Rancourt *et al.*, 1995). It is argued that the mutation from proline to leucine causes a small but crucial structural change in the protease that prevents its activation with pVI at the non-permissive temperature (39°C), but still allowing for activation with the activating peptide GVQSLKRRRCF (Rancourt *et al.*, 1995), albeit somewhat to a lesser extent than in the case of the wild type (Keyvani-Amineh *et al.*, 1995a). As stated above (cf. 4.5), the N-terminal region of the activating peptide interacts with a loop region comprising residues 135-140, precisely where the proline is located. A mutation from a proline to a leucine should affect the conformation of this region, presumably altering the affinity of the activating peptide in binding and producing the conformational change that leads to activation, thus conferring a lower activity. However, the ability of the activating peptide in rescuing the infectivity of the ts1 mutant together with the observation that the recombinant ts1 is not temperature-sensitive *in vitro* (Rancourt *et al.*, 1995), showing almost the same activity as the wild type recombinant protease (Keyvani-Amineh *et al.*, 1995a), are inconsistent with the assumption that the conformational change caused by the mutation from proline to leucine prevents the activation with pVI (Rancourt *et al.*, 1995). This rescue with

GVQSLKRRRCF rather seems to indicate that at the non-permissive temperature the protease is not interacting with pVI due to the protease not being directed to the presence of pVI, thus confirming that the ts1 mutant is likely a protein trafficking phenotype. The binding of the activating peptide (rescuing) should correct the conformation of the ts1 protease, thus changing its recognition pattern and enabling it to be targeted properly (Vaughan, personal communication).

The reversible transition between the inactive form **A** of the protease and the active form **C** by reversible binding of the activating peptide should be a reflection of an *in vivo* regulation mechanism of the protease, which might also involve targeting of the protease, as the behaviour of the ts1 mutant mentioned above seems to suggest. This is supported by the observation (Rancourt *et al.*, 1995) that the activating peptide inhibits wild type virus production when added to adenovirus infected cells, presumably by subverting the control mechanism.

4.8. How and Where Does the Activating Peptide Bind to the Protease?

Certainly, there is enough evidence that a covalent bond is formed between cysteine-104 of the protease and cysteine-10 of the activating peptide. But where does the N-terminal bind to?

The results above also do seem to suggest that proline-137 might be a residue involved in the N-terminal interaction between the activating peptide and the protease. Certainly, it is a very well conserved residue among different serotypes (cf. Figure 1.8), and its mutation to leucine in the ts1 mutant did present a lower activity as compared to the wild type, and this could be explained in terms of the replacement of a proline, which confers a certain rigidity to the loop, by another residue that results in a greater flexibility of the protease, which could compromise the correct positioning of the catalytic site residues. Furthermore, in the serotypes where this residue is mutated (bovine type 7, ovine strain 287, avian type 1 and avian strain 127, cf. Figure 1.8) the corresponding peptide is also different. Whereas in ovine strain 287 and avian strain 127 the cysteine in the activating peptide is flanked by two tyrosines (cf. Figure 1.9), in avian type 1 there is a methionine to the left of the cysteine, but cysteine-104 in its corresponding protease presents a phenylalanine to its right. Thus, it is plausible that a weaker interaction between the N-terminal of the activating peptide and the protease in the serotypes where

proline-137 is absent might be compensated by a more stable interaction at the disulphide bond, provided by aromatic residues flanking on both sides rather than just on one.

This hypothesis would also give more credence to the importance of the existence of a binding equilibrium between the protease and the activating peptide, as a peptide that would bind too strongly would activate the protease constutively, and a peptide not binding tightly enough would not confer enough activity to the protease.

4.9. Is Inhibition of Activation Possible?

If the protease does require the binding of the activating peptide in order to be fully active, a mechanism which prevents the binding of the activating peptide to the protease should operate as an inhibitor of activation. This could be achieved by means of finding a molecule that would bind to cysteine-104 in the protease, not activating it and not being replaced by the native activating peptide. Alternatively, a molecule that would modify chemically cysteine-104 to a derivative not allowing disulphide bonding could also serve the same purpose, but this would present the disadvantage of possibly reacting with other cysteine residues, both in the protease and in other viral and host proteins.

The result that the shorter peptides QSLKRRRCF and SLKRRRCF did not activate the protease to a reasonable extent (cf. 3.3.2.2 and 3.3.2.3), but still bound to it (cf. 3.6.1) suggested that they could act as inhibitors of activation. However, when inhibition assays were performed (cf. 3.6.2), only the peptide QSLKRRRCF was able to inhibit activation to some extent (reducing the activity to 77% of that achieved with the native activating peptide), which indicated that the native activating peptide was able to dislodge the shorter peptides, suggesting a reversible binding of the peptides. The reverse (substitution of GVQSLKRRRCF by QSLKRRRCF, cf. 3.6.2.4) did not happen, indicating a greater affinity of the native peptide for the protease.

A good inhibitor of activation should be able to bind reasonably tightly to cysteine-104, and an analogue of the activating peptide should present some characteristics to ensure a good binding (cf. 4.7). But the very same characteristics that confer good binding properties also activate the protease.

Yet, the mechanism by which the protease is presumably activated (cf. 4.5) involves the amino acid sequence GV in the activating peptide to be at a certain distance from the cysteine-10, which would stabilise two different regions in the protease at an optimal distance for the catalytic apparatus to be efficiently positioned. Therefore, it is possible

that peptides such as GVQSLLKRRRCF or GVQSKRRRCF, which present all the requirements for efficient binding and are reasonably similar to the native activating peptide, would indeed bind to the protease, although causing a conformational change that would put the catalytic residues either too close together or too far apart, thus affecting its catalytic properties.

Another possibility is attempting to block cysteine-104 with a peptide such as YCY. This peptide would presumably form a stable and protected disulphide bond with the cysteine due to the stabilising effect of aromatic residues on disulphide bonds already mentioned above (cf. 4.6 and 4.8). Curiously enough, two pVI C-terminal adenovirus serotypes already present this sequence (cf. Figure 1.9), which could suggest that this is indeed a mechanism already used by some adenovirus serotypes to enhance the binding of their activating peptides to their respective proteases.

4.10. What else Activates the Protease?

It has been reported that adenovirus DNA also enhances protease activity (Mangel *et al.*, 1993; Mangel *et al.*, 1996), although the requirement seems to be any negatively charged polymer (Mangel *et al.*, 1993). However, it also has been argued (Webster *et al.*, 1994) that DNA acts only as a stabiliser against reducing agents. The crystal structure (Ding *et al.*, 1996) does present four clusters of positive charge ranging in area from 45 to 65 $Å^2$, and the shortest distance between two of these clusters being ~24 Å is commensurate with the rise of a single turn of double-stranded DNA. Recombinant protease was observed to increase its K_m by 2-fold and its k_{cat} by 3-fold when incubated with adenovirus DNA in comparison to protease alone, whereas an increase of K_m by 2-fold and k_{cat} by 355-fold was observed when activating peptide was incubated to the protease in comparison to the protease alone (Mangel *et al.*, 1996). Yet, when both DNA and activating peptide are incubated together, there is an increase of K_m by 2-fold and its k_{cat} by 6072-fold as compared to the protease alone (Mangel *et al.*, 1996), suggesting that the two molecules act together cooperatively. Curiously, increasing concentrations of adenovirus DNA seem to increase activity up to a concentration of 2-4 μg/ml, followed by a decrease of activity.

Recently, it was reported (Diouri *et al.*, 1996) that a new stimulation ligand, the peptide VEGGS, also stimulates the recombinant protease to a similar extent of that achieved with the activating peptide. It was also reported that this ligand increases

fluorescence emission of tryptophan residues in the protease, suggesting a conformational change, and inhibited virus infection, all in a similar fashion to the activating peptide. Once the protease has been activated by one of VEGGS or GVQSLKRRRCF, the other peptide does not further enhance activity, which suggests that the two peptides cause the same conformational change in the protease. However, attempts to find if and where does this ligand bind to have proven unsuccessful.

4.11. Future Perspectives

Some characteristics of the mechanism by which the activating peptide converts the adenovirus protease into its most active form have been elucidated in this work. However, this is a never ending subject, and many questions remain unanswered.

Namely, the understanding of the nature of the binding of the N-terminal region of the activating peptide to the protease would enable a more rational approach to inhibitor design, as well as the understanding of possible targeting mechanisms during virus infections.

Also, the results presented above seem to indicate that cysteine-104 in the protease exhibits an unusual reactivity, and a better idea of the reasons for this could provide with a better insight on the nature of the reversibility of the covalent binding to the activating peptide.

The divergence of results concerning the effect of DNA on the activity of the protease is also puzzling, and could possibly be related to the assay systems used to determine its influence, or could be an indication that DNA can be used as a guidewire for the protease to cleave proteins within the viral capsid (Ding *et al.*, 1996).

The ligand VEGGS is also intriguing. The results presented in this work seem to indicate that an efficient binding of the activating peptide to the protease requires a joint effort of several regions of the peptide in keeping it bound to the protease. This ligand does not conform to any of those requirements, and yet it is claimed that it has a similar effect to that of the activating peptide (Diouri *et al.*, 1996). The understanding of its mechanism of binding and subsequent activation should be able to reveal alternative paths for activation, and consequently inhibition.

The understanding of the mechanism by which the adenovirus protease is activated should enable a rational design of inhibitors of activation which, together with inhibitors directed to the active site (Brown *et al.*, 1996; Cornish *et al.*, 1995; Sircar *et al.*, 1996),

would allow for the conception of specific inhibitor cocktails which should prove useful in anti-viral therapy. Since the protease presents quite a high specificity for both the cleavage consensus sequence (Anderson, 1990; Webster *et al.*, 1989a) and for the activation mechanism, the resulting inhibitors should also be highly selective and not interfere with the host proteases, which would present these inhibitors as ideal anti-viral agents.

5. References

Abate, C., Patel, L., III, F. J. R., and Curran, T. (1990). *Science* **249,** 1157-1161.

Adam, S. A., and Dreyfuss, G. (1987). *J. Virol.* **61,** 3276-3283.

Akusjärvi, G., Zabielski, J., Perricaudet, M., and Pettersson, U. (1981). *Nucleic Acids Res.* **9,** 1-17.

Anderson, C. W. (1990). *Virology* **177,** 259-272.

Anderson, C. W., Baum, P. R., and Gesteland, R. F. (1973). *J. Virol.* **12,** 241-252.

Anderson, C. W., Young, M. E., and Flint, S. J. (1989). *Virology* **172,** 506-512.

Andrewes, C., Pereira, H. G., and Wildy, P. (1978). Adenoviridae. *In* "Viruses of Vertebrates" (B. Tindall, ed.), pp. 293-311, London.

Applied Biosystems. (1990). Introduction To Cleavage Techniques, pp. 71. Applied Biosystems, Foster City.

Arshady, R., Atherton, E., Clive, D. L. J., and Sheppard, R. C. (1981). *J. Chem. Soc. Perkin Trans. I*, 529-537.

Atherton, E., Cameron, L. R., and Sheppard, R. C. (1988). *Tetrahedron* **44,** 843-857.

Atherton, E., Logan, C. J., and Sheppard, R. C. (1981). *J. Chem. Soc. Perkin Trans. I*, 538-546.

Atherton, E., and Sheppard, R. C. (1989). Solid phase peptide synthesis: A practical approach. *In* "Practical Approach Series" (D. Rickwood and B. D. Hames, eds.), pp. 203. IRL Press, Oxford.

Babiss, L. E., and Ginsberg, H. S. (1984). *J. Virol.* **50,** 202-212.

Babiss, L. E., Ginsberg, H. S., and Darnell, J. E., Jr. (1985). *Mol. Cell. Biol.* **5,** 2552-2558.

Bachenheimer, S., and Darnell, J. E. (1975). *Proc. Natl. Acad. Sci. USA* **72,** 4445-4449.

Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1993). *J. Virol.* **67,** 6056-6063.

Barrett, A. J. (1977). Proteinases in Mammalian Cells and Tissues. North Holland Publishing Company, Amsterdam.

Baum, E. Z., Ding, W.-D., Siegel, M. M., Hulmes, J., Bebernitz, G. A., Sridharan, L., Tabei, K., Krishnamurthy, G., Carofiglio, T., Groves, J. T., Bloom, J. D., DiGrandi, M., Bradley, M., Ellestad, G., Seddon, A. P., and Gluzman, Y. (1996a). *Biochemistry* **35,** 5847-5855.

Baum, E. Z., Siegel, M. M., Bebernitz, G. A., Hulmes, J. D., Sridharan, L., Sun, L., Tabei, K., Johnston, S. H., Wildey, M. J., Nygaard, J., Jones, T. R., and Gluzman, Y. (1996b). *Biochemistry* **35,** 5838-5846.

Belin, M., and Boulanger, P. (1987). *J. Virol.* **61,** 2559-2566.

Beltz, G. A., and Flint, S. J. (1979). *J. Mol. Biol.* **131,** 353-373.

Benesch, R. E., and Benesch, R. (1955). *J. Am. Chem. Soc.* **77,** 5877-5881.

Berget, S. M., Moore, C., and Sharp, P. A. (1977). *Proc. Natl. Acad. Sci. USA* **74,** 3171-3175.

Bhatti, A. R., and Weber, J. (1978). *Biochem. Biophys. Res. Comm.* **81,** 973-979.

Bhatti, A. R., and Weber, J. (1979). *Virology* **96,** 478-485.

Bio-Rad. Piperazine di-Acrylamide (PDA). Bio-Rad.

Bio-Rad. (1996). BioFocus Capillary Electrophoresis Systems, pp. 270. Bio-Rad.

Bodnar, J. W., Hanson, P. I., Polvino-Bodnar, M., Zempsky, W., and Ward, D. C. (1989). *J. Virol.* **63,** 4344-4353.

Bodner, B. L., Jackman, L. M., and Morgan, R. S. (1980). *Biochem. Biophys. Res. Comm.* **94,** 807-813.

Böhlen, P., Stein, S., Dairman, W., and Udenfriend, S. (1973). *Arch. Biochem. Biophys.* **155,** 213-220.

Bond, J. S., and Butler, P. E. (1987). *Ann. Rev. Biochem.* **56,** 333-364.

Borelli, E., Hen, R., and Chambon, P. (1984). *Nature* **312,** 608-612.

Bosher, J., Robinson, E. C., and Hay, R. T. (1990). *New Biol.* **2,** 1083-1090.

Bradford, M. M. (1976). *Anal. Biochem.* **72,** 248-254.

Bridge, E., and Ketner, G. (1990). *Virology* **174,** 345-353.

Brown, M. T., McGrath, W. J., Toledo, D. L., and Mangel, W. F. (1996). *FEBS Lett.* **388,** 233-237.

Burnette, W. N. (1981). *Anal. Biochem.* **112,** 195-203.

Butkiewicz, N. J., Wendel, M., Zhang, R., Jubin, R., Pichardo, J., Smith, E. B., Hart, A. M., Ingram, R., Durkin, J., Mui, P. W., Murray, M. G., Ramanathan, L., and Dasmahapatra, B. (1996). *Virology* **225,** 328-338.

Cabrita, G., Iqbal, M., Reddy, H., and Kemp, G. (1997). *J. Biol. Chem.* **272,** 5635-5639.

Cabrita, G. J. M. (1992). Adenovirus Protease - A Fluorescent Protease Assay. Research Project for Practicals in Biochemical Engineering, University of St. Andrews.

Cai, F., and Weber, J. M. (1993). *Virology* **196,** 358-362.

Carpino, L. A. (1993). *J. Am. Chem. Soc.* **115,** 4397-4398.

Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985). *Cell* **43,** 439-448.

Challberg, M. D., Desiderio, S. V., and Kelly, T. J., Jr. (1980). *Proc. Natl. Acad. Sci. USA* **77,** 5105-5109.

Challberg, M. D., and Kelly, T. D. (1989). *Ann. Rev. Biochem.* **58,** 671-717.

Challberg, M. D., and Kelly, T. J., Jr. (1981). *J. Virol.* **38**.

Chambers, T. J., Grakoui, A., and Rice, C. M. (1991). *J. Virol.* **65,** 6042-6050.

Chang, L.-S., and Shenk, T. (1990). *J. Virol.* **64,** 2103-2109.

Chardonnet, Y., and Dales, S. (1970a). *Virology* **40,** 462-477.

Chardonnet, Y., and Dales, S. (1970b). *Virology* **40,** 478-485.

Chatterjee, P. K., Bruner, M., Flint, S. J., and Harter, M. L. (1988). *EMBO J.* **7,** 835-841.

Chatterjee, P. K., and Flint, S. J. (1987). *Proc. Natl. Acad. Sci. USA* **84,** 714-718.

Chatterjee, P. K., Vayda, M. E., and Flint, S. J. (*1986*). *J. Mol. Biol.* **188,** 23-37.

Chen, M., Mermod, N., and Horwitz, M. S. (1990). *J. Biol. Chem.* **265,** 18634-18642.

Chen, P. H., Ornelles, D. A., and Shenk, T. (1993). *J. Virol.* **67,** 3507-3514.

Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V., and Cotten, M. (1996). *J. Virol.* **70,** 2939-2949.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). *Cell* **12,** 1-8.

Chrambach, A., Reisfeld, R. A., Wyckoff, M., and Zaccari, J. (1967). *Anal. Biochem.* **20,** 150-154.

Clarke, A. R., Purdie, C. A., Harrison, D. J., Morris, R. G., Bird, C. C., Hooper, M. L., and Wyllie, A. H. (1993). *Nature* **362,** 849-852.

Cleat, P. H., and Hay, R. T. (1989). *EMBO J.* **8,** 1841-1848.

Coates, E., Marsden, C. G., and Rigg, B. (1969). *Trans. Faraday Soc.* **65,** 3032-3036.

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

Cornish-Bowden, A. (1995). Fundamentals of Enzyme Kinetics, pp. 343. Portland Press, Cambridge.

Cotten, M., and Weber, J. M. (1995). *Virology* **213,** 494-502.

Dales, S., and Chardonnet, Y. (1973). *Virology* **56,** 465-483.

Darbyshire, J. H. (1966). *Nature* **211,** 102.

Davidson, A. J., Telford, E. A., Watson, M. S., McBride, K., and Mautner, V. (1993). *J. Mol. Biol.* **234,** 1308-1316.

de Bernardo, S., Weigele, M., Toome, V., Manhart, K., Leimgruber, W., Böhlen, P., Stein, S., and Udenfriend, S. (1974). *Arch. Biochem. Biophys.* **163,** 390-399.

de Groot, R. J., Hardy, W. R., Shirako, Y., and Strauss, J. H. (1990). *EMBO J.* **9,** 2631-2638.

de St. Groth, S. F., Webster, R. G., and Datyner, A. (1963). *Biochem. Biophys. Acta* **71,** 377-391.

Debbas, M., and White, E. (1993). *Genes Dev.* **7,** 546-554.

Defer, C., Belin, M. T., Caillet-Boudin, M. L., and Boulanger, P. (1990). *J. Virol.* **64,** 3661-3673.

Devaux, C., Caillet-Boudin, M. L., Jacrot, B., and Boulanger, P. (1987). *Virology* **161,** 121-128.

D'Halluin, J.-C., Milleville, M., Martin, G. R., and Boulanger, P. (1980). *J. Virol.* **33,** 88-99.

Diller, L., Kassel, J., Nelson, C. E., Gryka, M. A., Litwak, G., Gebbhardt, M., Bressac, B., Ozturk, M., Baker, S. J., Vogelstein, B., and Friend, S. H. (1990). *Mol. Cell. Biol.* **10,** 5772-5781.

Ding, J., McGrath, W. J., Sweet, R. M., and Mangel, W. F. (1996). *EMBO J.* **15,** 1778-1783.

Dion, A. S., and Pomenti, A. A. (1983). *Anal. Biochem.* **129,** 490-496.

Diouri, M., Girouard, G. S., Allen, C. M., Sircar, S., Lier, J. E. V., and Weber, J. M. (1996). *Virology* **224,** 510-516.

Dixon, W. J., and Massey, F. J. (1957). Introduction to Statistical Analysis, pp. 450. McGraw-Hill, New York.

Doerfler, W. (1994). Adenoviruses: Molecular Virology. *In* "Encyclopedia of Virology" (R. G. Webster and A. Granoff, eds.), pp. 8-14. Academic Press, London.

Dolph, P. J., Huang, J., and Schneider, R. J. (1990). *J. Virol.* **64,** 2669-2677.

Dyson, N., Guida, P., Munger, K., and Harlow, E. (1992). *J. Virol.* **66,** 6893-6902.

Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., and Craik, C. S. (1992). *J. Biol. Chem.* **267,** 7411-7420.

Edsall, J. T., Martin, R. B., and Hollingworth, B. R. (1958). *Proc. Natl. Acad. Sci. USA* **44,** 505-518.

Eisenthal, R., and Danson, M. J. (1993). Enzyme Assays: A Practical Approach. *In* "Practical Approach Series" (D. Rickwood and B. D. Hames, eds.), pp. 351. IRL Press, Oxford.

Eldjarn, L., and Pihl, A. (1957). *J. Am. Chem. Soc.* **79,** 4589-4593.

Ellis, K. J., and Morrison, J. F. (1982). *Methods Enzymol.* **87,** 405-426.

Ellman, G. L. (1959). *Arch. Biochem. Biophys.* **82,** 70-77.

Enders, J. F., Bell, J. A., Dingle, J. H., Francis, T., Hilleman, M. R., Huebner, R. J., and Payne, A. M.-M. (1956). *Science* **124,** 119-120.

Engels, B. (1992). Amplify, Madison.

Enomoto, T., Lichy, J. H., Ikeda, J., and Hurwitz, J. (1981). *Proc. Natl. Acad. Sci. USA* **78,** 6779-6783.

Eron, L., Wesphal, H., and Callahan, R. (1974). *J. Virol.* **14,** 375-383.

Evans, C. H., and Ridella, J. D. (1984). *Anal. Biochem.* **142,** 411-420.

Everitt, E., Lutter, L., and Philipson, L. (1975). *Virology* **67,** 197-208.

Everitt, E., Sundquist, B., Petersson, U., and Philipson, L. (1973). *Virology* **62,** 130-147.

Falgout, B., Miller, R. H., and Lai, C.-J. (1993). *J. Virol.* **67,** 2034-2042.

Ferguson, B., Krippl, B., Andrisani, O., Jones, N., and Westphal, H. (1985). *Mol. Cell. Biol.* **5,** 2653-2661.

Field, J., Gronostajski, R. M., and Hurwitz, J. (1984). *J. Biol. Chem.* **259,** 9487-9495.

FitzGerald, D. J. P., Padmanabhan, R., Pastan, I., and Willingham, M. C. (1983). *Cell* **32,** 607-617.

Flegel, M., and Sheppard, R. C. (1990). *J. Chem. Soc. Chem. Commun.* **7,** 536-538.

Fredman, J. N., and Engler, J. A. (1993). *J. Virol.* **67,** 3384-3395.

Freeman, A., Black, P. H., Wolford, R., and Huebner, R. J. (1967). *J. Virol.* **1,** 362-367.

Ginsberg, H. S., Pereira, H. G., Valentine, R. C., and Wilcox, W. C. (1966). *Virology* **28,** 782-783.

Girardi, A. J., Hilleman, M. R., and Zwickey, R. E. (1964). *Proc. Soc. Exp. Biol. Med.* **115,** 1141-1150.

Glover, D. M., and Hames, B. D. (1995). DNA Cloning 2: Expression Systems - A practical approach. *In* "Practical Approach Series" (D. Rickwood and B. D. Hames, eds.), pp. 255. IRL Press, Oxford.

Goodrich, D. W., Wang, N. P., Qian, Y.-W., Lee, E. Y.-H. P., and Lee, W. H. (1991). *Cell* **67,** 293-302.

Gorin, G., and Clary, C. W. (1960). *Arch. Biochem. Biophys.* **90,** 40-45.

Grable, M., and Hearing, P. (1992). *J. Virol.* **66,** 723-731.

Greber, U. F., Webster, P., Weber, J., and Helenius, A. (1996). *EMBO J.* **15,** 1766-1777.

Greber, U. F., Willetts, M., Webster, P., and Helenius, A. (1993). *Cell* **75,** 477-486.

Green, M., and Daesch, G. E. (1961). *Virology* **13,** 169-176.

Green, M., and Piña, M. (1963). *Virology* **20,** 199-207.

Green, M., Piña, M., Kimes, R., Wensink, P., MacHattie, L., and Thomas, C. A., Jr. (1967). *Proc. Natl. Acad. Sci. USA* **57,** 1302-1309.

Green, M., Wold, W. S. M., and Brackmann, K. H. (1980). Human adenovirus transforming genes: group relationships, integration, expression in transformed cells and analysis of human cancers and tonsils. *In* "Seventh Cold Spring Harbor Conference on Cell Proliferation Viruses in Naturally Occurring Tumors" (M. Essex, G. Todaro and H. zurHausen, eds.), pp. 373-397. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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

Grisshammer, R., and Nagai, K. (1995). Purification of over-produced proteins from *E. coli* cells. 2nd ed. *In* "DNA Cloning 2: Expression Systems - A practical approach" (D. M. Glover, and B. D. Hames, Eds.), Vol. 2, pp. 255. 4 vols. IRL Press, Oxford.

Halbert, D. N., Cutt, J. R., and Shenk, T. (1985). *J. Virol.* **56,** 250-257.

Hall, D. L., and Darke, P. L. (1995). *J. Biol. Chem.* **270,** 22697-22700.

Hammarskjold, M. L., and Winberg, G. (1980). *Cell* **20,** 787-795.

Han, K.-K., Belaiche, D., Moreau, O., and Briand, G. (1985). *Int. J. Biochem.* **17,** 429-445.

Handa, H., Kingston, R. E., and Sharp, P. A. (1983). *Nature* **302,** 545-547.

Hardy, S., Engel, D., and Shenk, T. (1989). *Genes Dev.* **3,** 1062-1074.

Harlow, E., Franza, B. R., and Schley, C. (1985). *J. Virol.* **55,** 533-546.

Harris, E. L. V., and Angal, S. (1989). Protein purification methods: a practical approach. *In* "Practical Approach" (D. Rickwood and B. D. Hames, eds.), pp. 317. IRL Press, Oxford.

Hartwell, L. (1992). *Cell* **71,** 543-546.

Hassel, J. A., and Weber, J. (1978). *J. Virol.* **28,** 671-678.

Hasson, T. B., Ornelles, D. A., and Shenk, T. (1992). *J. Virol.* **66,** 6133-6142.

Hasson, T. B., Soloway, P. D., Ornelles, D. A., Doerfler, W., and Shenk, T. (1989). *J. Virol.* **63,** 3612-2621.

Hawkins, C. (1997). Investigation into the Activation Mechanism of the Adenovirus Serotype 2 Protease. Senior Honours Research Project, University of St. Andrews.

Hay, R. T., and Russell, W. C. (1989). *Biochem. J.* **258,** 3-16.

Hayes, B. W., Telling, G. C., Myat, M. M., Williams, J. F., and Flint, S. J. (1990). *J. Virol.* **64,** 2732-2742.

Hearing, P., Samulski, R., Wishart, W., and Shenk, T. (1987). *J. Virol.* **61,** 2555-2558.

Hellen, C. U. T., and Wimmer, E. (1992). *Experientia* **48,** 201-215.

Hennache, B., and Boulanger, P. (1977). *Biochem. J.* **166,** 237-247.

Hierholzer, J. C., Stone, Y. O., and Broderson, J. R. (1991). *Arch. Virol.* **121,** 179-197.

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

Hinds, P. W., Mittnacht, S., Dulic, V., Arnold, A., Reed, S. I., and Weinberg, R. A. (1992). *Cell* **70,** 993-1006.

Hoeffler, W. K., Kovelman, R., and Roeder, R. G. (1988). *Cell* , 907-920.

Hoeffler, W. K., and Roeder, R. G. (1985). *Cell* **41,** 955-963.

Holmgren, A. (1985). *Ann. Rev. Biochem.* **54,** 237-271.

Horne, R. W., Brenner, S., Waterson, A. P., and Wildy, P. (1959). *J. Mol. Biol.* **1,** 84-86.

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

Horwitz, M. S., Scharff, M. D., and Maizel, J. V. (1969). *Virology* **39,** 682-694.

Hosakawa, K., and Sung, M. T. (1976). *J. Virol.* **17,** 924-934.

Houde, A., and Weber, J. M. (1990). *Gene* **88,** 269-273.

Huang, J., and Schneider, R. J. (1991). *Cell* **65,** 271-280.

Huang, M.-M., and Hearing, P. (1989). *Genes Dev.* **3,** 1699-1710.

Huebner, R. J., Casey, M. J., Chanock, R. M., and Scheel, K. (1965). *Proc. Natl. Acad. Sci. USA* **54,** 381-388.

Huebner, R. J., Chanock, R. M., Rubin, B. A., and Casey, M. J. (1964). *Proc. Natl. Acad. Sci. USA* **52,** 1333-1340.

Huebner, R. J., Rowe, W. P., and Lane, W. T. (1962). *Proc. Natl. Acad. Sci. USA* **48,** 2051-2058.

Huebner, R. J., Rowe, W. P., Turner, H. C., and Lane, W. T. (1963). *Proc. Natl. Acad. Sci. USA* **50,** 379-389.

Huebner, R. J., Rowe, W. P., Ward, T. G., Parrott, R. H., and Bell, J. A. (1954). *N. Engl. J. Med.* **251,** 1077-1086.

Hull, R. N., Johnson, I. S., Culbertson, C. G., Reimer, C. B., and Wright, H. F. (1965). *Science* **150,** 1044-1046.

Ibelgaufts, H., Jones, K. W., Maitland, N., and Shaw, N. F. (1982). *Acta Neuropathol.* **56,** 113-117.

Ishibashi, M., and Maizel, J. V., Jr. (1974a). *Virology* **57,** 409-424.

Ishibashi, M., and Maizel, J. V., Jr. (1974b). *Virology* **58,** 345-361.

Jocelyn, P. C. (1987). *Methods Enzymol.* **143,** 44-67.

Jones, S. J., Iqbal, M., Grierson, A. W., and Kemp, G. (1996). *J. Gen. Virol.* **77,** 1821-1824.

Keire, D. A., Strauss, E., Guo, W., Noszál, B., and Rabenstein, D. L. (1992). *J. Org. Chem.* **57,** 123-127.

Keyvani-Amineh, H., Diouri, M., Guillemette, J. G., and Weber, J. M. (1995a). *J. Biol. Chem.* **270,** 23250-23253.

Keyvani-Amineh, H., Diouri, M., Tihanyi, K., and Weber, J. M. (1996). *J. Gen. Virol.* **77,** 2201-2207.

Keyvani-Amineh, H., Labrecque, P., Cai, F., Carstens, E. B., and Weber, J. M. (1995b). *Virus Res.* **37,** 87-97.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996). *Cell* **87,** 343-355.

Kitchingman, G. R. (1994). Adenoviruses: Malignant Transformation and Oncology. *In* "Encyclopedia of Virology" (R. G. Webster and A. Granoff, eds.), pp. 17-23. Academic Press, London.

Kohnken, R. E., Ladror, U. S., Wang, G. T., Holzman, T. F., Miller, B. E., and Krafft, G. A. (1995). *Exp. Neurol.* **133,** 105-112.

Kolthoff, I. M., Stricks, W., and Kapoor, R. C. (1955). *J. Am. Chem. Soc.* **77,** 4733-4739.

Kruijer, W., van Schaik, F. M. A., and Sussenbach, J. S. (1980). *Nucleic Acids Res.* **8.**

Kurosaki, T., Tsutsui, K., Tsutsui, K., Aoyama, K., and Oda, T. (1984). *Biochem. Biophys. Res. Comm.* **123,** 729-734.

Laemmli, U. K. (1970). *Nature* **227,** 680-685.

Larsson, V. M., Girardi, A. J., Hilleman, M. R., and Zwickey, R. E. (1965). *Proc. Soc. Exp. Biol. Med.* **118,** 15-24.

Lechner, R. L., and Kelly, T. J., Jr. (1977). *Cell* **12,** 1007-1020.

Leong, M. M. L., and Fox, G. R. (1990). *Methods Enzymol.* **184,** 442-451.

Lichy, J. H., Horwitz, M. S., and Hurwitz, J. (1981). *Proc. Natl. Acad. Sci. USA* **79,** 2678-2682.

Lin, D., Shields, M. T., Ullrich, S. J., Appella, E., and Mercer, W. E. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 9210-9214.

Lindenbaum, J. O., Field, J., and Hurwitz, J. (1986). *J. Biol. Chem.* **261,** 10218-10227.

Londberg-Holm, K., and Philipson, L. (1969). *J. Virol.* **4,** 323-338.

Lonsdale-Eccles, J. D., Mpimbaza, G. W. N., Nkhungulu, Z. R. M., Olobo, J., Smith, L., Tosomba, O. M., and Grab, D. J. (1995). *Biochem. J.* **305,** 549-556.

Lowe, S. W., and Ruley, H. E. (1993). *Genes Dev.* **7,** 535-545.

Lowe, S. W., Schmitt, E., Smith, S., Osborne, B., and Jacks, T. (1993). *Nature* **362,** 847-849.

Luftig, R. B., and Weihing, R. R. (1975). *J. Virol.* **16,** 696-706.

Mackey, J. K., Rigden, P. M., and Green, M. (1976). *Proc. Natl. Acad. Sci. USA* **73,** 4675-4681.

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

Mangel, W. F., Toledo, D. L., Brown, M. T., Martin, J. H., and McGrath, W. J. (1996). *J. Biol. Chem.* **271,** 536-543.

Maran, A., and Mathews, M. B. (1988). *Virology* **164,** 106-113.

Margosiak, S. A., Vanderpool, D. L., Sisson, W., Pinko, C., and Kan, C.-C. (1996). *Biochemistry* **35,** 5300-5307.

Martinez-Palomo, A., and Granboulan, N. (1967). *J. Virol.* **1,** 1010-1018.

Mathews, M. B. (1980). *Nature* **285,** 575-577.

Matthews, D. A., and Russell, W. C. (1995). *J. Gen. Virol.* **76,** 1959-1969.

Maxam, A. M., and Gilbert, W. (1977). *Proc. Natl. Acad. Sci. USA* **74,** 560-564.

McBride, W. D., and Wiener, A. (1964). *Proc. Soc. Exp. Biol. Med.* **115,** 870-874.

McGrath, W. J., Abola, A. P., Toledo, D. L., Brown, M. T., and Mangel, W. F. (1996). *Virology* **217,** 131-138.

Merrifield, R. B. (1963). *J. Am. Chem. Soc.* **85,** 2149-2154.

Merrifield, R. B. (1964a). *J. Am. Chem. Soc.* **86,** 304-305.

Merrifield, R. B. (1964b). *Biochemistry* **3,** 1385-1390.

Merril, C. R. (1990). *Methods in Enzimology* **182,** 477-488.

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

Miyamoto, N. G., Moncolin, V., Egly, J. M., and Cambon, P. (1985). *EMBO J.* **4,** 3563-3570.

Moran, E., and Mathews, M. B. (1987). *Cell* **48,** 177-178.

Morgan, R. S., and McAdon, J. M. (1980). *Int. J. Pept. Prot. Res.* **15,** 177-180.

Morgan, R. S., Tatsch, C. E., Gushard, R. H., McAdon, J. M., and Warme, P. K. (1978). *Int. J. Pept. Prot. Res.* **11,** 209-217.

Morin, N., Delsert, C., and Klessig, D. F. (1989). *J. Virol.* **63,** 5228-5237.

Moyne, G., Pichard, E., and Bernhard, W. (1978). *J. Gen. Virol.* **40,** 77-92.

Mul, Y. M., and van der Vliet, P. C. (1992). *EMBO J.* **11,** 751-760.

Mul, Y. M., Verrijzer, C. P., and van der Vliet, P. C. (1990). *J. Virol.* **64,** 5510-5518.

Muller, U., Kleinberger, T., and Shenk, T. (1992). *J. Virol.* **66,** 5867-5878.

Muller, U., Roberts, M. P., Engel, D. A., Doerfler, W., and Shenk, T. (1989). *Genes Dev.* **3,** 1991-2002.

Nakai, N., Lai, C. Y., and Horecker, B. L. (1974). *Anal. Biochem.* **58,** 563-570.

Nakamura, H., Tsuzuki, S., and Tamura, Z. (1980). *J. Chromatogr.* **200,** 324-329.

Neill, S. D., Hemstrom, C., and Virtanen, A. (1990). *Proc. Natl. Acad. Sci. USA* **87,** 2008-2012.

Nene, V., Gobright, E., Musoke, A. J., and Lonsdale-Eccles, J. D. (1990). *J. Biol. Chem.* **265,** 18047-18050.

Nevins, J. R. (1981). *Cell* **26,** 213-220.

Nevins, J. R. (1992). *Science* **258,** 424-429.

Nevins, J. R., and Winkler, J. J. (1980). *Proc. Natl. Acad. Sci. USA* **77,** 1893-1897.

Nielsen, B. L., and Brown, L. R. (1984). *Anal. Biochem.* **141,** 311-315.

Noiva, R., and Lennarz, W. J. (1992). *J. Biol. Chem.* **267,** 3553-3556.

Norrby, E. (1966). *Virology* **28,** 236-248.

Norrby, E. (1969a). *Virology* **37,** 565-576.

Norrby, E. (1969b). *J. Gen. Virol.* **5,** 221-236.

Oakley, B. R., Kirsch, D. R., and Morris, N. R. (1980). *Anal. Biochem.* **105,** 361-363.

Olins, P. O., and Rangwala, S. H. (1989). *J. Biol. Chem.* **264,** 16973-16976.

O'Malley, R. P., Mariano, T. M., Siekierka, J., Merrick, W. C., Reichel, P. A., and Mathews, M. B. (1986). *Cancer Cells* **4,** 291-301.

Ornelles, D., and Shenk, T. (1991). *J. Virol.* **65,** 424-439.

Orr, D. C., Long, A. C., Kay, J., Dunn, B. M., and Cameron, J. M. (1989). *J. Gen. Virol.* **70,** 2931-2942.

Parke, T. V., and Davis, W. W. (1954). *Anal. Chem.* **26,** 642-645.

Pastan, I., Seth, P., FitzGerald, D., and Willingham, M. (1986). Adenovirus entry into cells; some new observations on an old problem. *In* "Concepts in viral pathogenesis II" (A. Notkins and M. Oldstone, eds.), pp. 141-146. Springer-Verlag, New York.

Pejler, G., and Maccarana, M. (1994). *J. Biol. Chem.* **269,** 14451-14456.

Pereira, H. G., Huebner, R. J., Ginsberg, H. S., and Van der Veen, J. (1963). *Virology* **20,** 613-620.

Pereira, M. S., Pereira, H. G., and Clarke, S. K. (1965). *Lancet* **1,** 21-23.

Perry, M. E., and Levine, A. J. (1993). *Curr. Opin. Gen. Dev.* **3,** 50-54.

Perry, R. H., and Green, D. W. (1984). Perry's Chemical Engineer's Handbook. *In* "McGraw-Hill Chemical Engineering Series". McGraw-Hill, Singapore.

Pettersson, U., and Roberts, R. J. (1986). *Cancer Cells* **4,** 37-57.

Philipson, L. (1984). Adenovirus assembly. *In* "The adenoviruses" (H. Ginsberg, ed.). Plenum Press, New york.

Pilder, S., Moore, M., Logan, J., and Shenk, T. (1986). *Mol. Cell. Biol.* **6,** 470-476.

Pollard, B. (1997). Purification of Recombinant Adenovirus Type 2 Protease. Senior Honours Research Project, University of St. Andrews.

Pombeiro, A. J. L. O. (1983). Técnicas e Operações Unitárias em Química Laboratorial, pp. 1070. Fundação Calouste Gulbenkian, Lisboa.

Pursiainen, M., Jauhiainen, M., and Ehnholm, C. (1994). *Biochem. Biophys. Acta* **1215,** 170-175.

QIAGEN. (1993). QIAGEN Plasmid Handbook, pp. 32. QIAGEN, Chatsworth.

QIAGEN. (1995a). QIAGEN Plasmid Handbook - New Edition, pp. 34. QIAGEN, Dorking.

QIAGEN. (1995b). QIAquick Handbook, pp. 22. QIAGEN, Dorking.

Qiu, X., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W., and Abdel-Meguid, S. S. (1996). *Nature* **383,** 275-279.

Rabenstein, D. L. (1973). *J. Am. Chem. Soc.* **95,** 2797-2803.

Rabson, A. S., Kirschstein, R. L., and Paul, F. J. (1964). *J. Natl. Cancer Inst.* **32,** 77-87.

Rancourt, C., Keyvani-Amineh, H., Diouri, M., and Weber, J. M. (1996). *Virology* **224,** 561-563.

Rancourt, C., Keyvani-Amineh, H., Sircar, S., Labrecque, P., and Weber, J. M. (1995). *Virology* **209,** 167-173.

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

Rao, L., Debbas, M., Sabbatini, P., Hockenbery, D., Korsmeyer, S., and White, E. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 7742-7746.

Raychaudhuri, P., Bagchi, S., Neill, S. D., and Nevins, J. R. (1990). *J. Virol.* **64,** 2702-2710.

Reid, V. M. (1997). The Influence of Oxidised and Reduced Peptides on the Activation of Recombinant Protease. Senior Honours Research Project, University of St. Andrews.

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

Roberts, R. J., O'Neill, K. E., and Yen, C. T. (1984). *J. Biol. Chem.* **259,** 13968-13975.

Rose, M. E., and Johnstone, R. A. W. (1982). Mass spectrometry for chemists and biochemists. *In* "Cambridge Texts in Chemistry and Biochemistry", pp. 308. Cambridge University Press, Cambridge.

Rosenberg, A. H., Lade, B. N., Chui, D.-S., Lin, S.-W., Dunn, J. J., and Studier, F. W. (1987). *Gene* **56,** 125-135.

Rosenfeld, P. J., O'Neill, E., Wides, R. J., and Kelly, T. J. (1987). *Mol. Cell. Biol.* **7,** 875-886.

Roswell, D. F., and White, E. H. (1978). *Methods Enzymol.* **57,** 409-423.

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

Russell, W. C. (1994). Adenoviruses: Animal Adenoviruses. *In* "Encyclopedia of Virology" (R. G. Webster and A. Granoff, eds.), pp. 14-17. Academic Press, London.

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

Ryklan, L. R., and Schmidt, C. L. A. (1944). *Arch. Biochem.* **5,** 89-98.

Salisbury, S. A., Tremeer, E. J., Davies, J. W., and Owen, D. E. I. A. (1990). *J. Chem. Soc. Chem. Commun.* **7,** 538-540.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: a laboratory manual (C. Nolan, ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanger. (1945). *Biochem. J.* **39,** 507-.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). *Proc. Natl. Acad. Sci. USA* **74,** 5463-5467.

Sarma, P. S., Huebner, R. J., and Lane, W. T. (1965). *Science* **149,** 1108.

Sarnow, P., Hearing, P., Anderson, C. W., Halbert, D. N., Shenk, T., and Levine, A. J. (1984). *J. Virol.* **49,** 692-700.

Sarnow, P., Ho, Y.-S., Williams, J., and Levine, A. J. (1982). *Cell* **28,** 387-394.

Sawadago, M., and Roeder, R. G. (1985). *Cell* **43,** 165-175.

Schaack, J., Ho, W. Y.-W., Freimuth, P., and Shenk, T. (1990). *Genes Dev.* **4,** 1197-1208.

Schellenberger, V., and Jakubke, H.-D. (1991). *Angew. Chem. Int. Ed. Engl.* **30,** 1437-1449.

Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Laureto, P. P. d., DasGupta, B. R., and Montecucco, C. (1992). *Nature* **359,** 832-835.

Schwabe, C. (1973). *Anal. Biochem.* **53,** 484-490.

Seth, P., FitzGerald, D. J., Willingham, M. C., and Pastan, I. (1984). *J. Virol.* **51,** 650-655.

Shaked, Z. e., Szajewski, R. P., and Whitesides, G. M. (1980). *Biochemistry* **19,** 4156-4166.

Shapiro, A. L., Viñuela, E., and Maizel, J. V. (1967). *Biochem. Biophys. Res. Comm.* **28,** 815-820.

Shaw, A. R., and Ziff, E. B. (1980). *Cell* **22,** 905-916.

Shaw, P., Bovey, R., Tardy, S., Sahli, R., Sordat, B., and Costa, J. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 4495-4499.

Shen, Y., and Shenk, T. (1994). *Proc. Natl. Acad. Sci. USA* **91,** 8940-8944.

Shenk, T. (1996). *Adenoviridae:* The Viruses and Their Replication. *In* "Fundamental Virology" (B. N. Fields, D. M. Knipe and P. M. Howley, eds.), pp. 979-1016. Lippincot-Raven, Philadelphia.

Shenk, T., and Flint, S. J. (1991). *Adv. Cancer Res.* **57,** 47-85.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). *Anal. Chem.* **68,** 850-858.

Shieh, H.-S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C., and Stallings, W. C. (1996). *Nature* **383,** 279-282.

Shimizu, Y., Yamaji, K., Masuho, Y., Yokota, T., Inoue, H., Sudo, K., Satoh, S., and Shimotohno, K. (1996). *J. Virol.* **70,** 127-132.

Shrager, R. I., Cohen, J. S., Heller, S. R., Sachs, D. H., and Schechter, A. N. (1972). *Biochemistry* **11,** 541-547.

Silver, L., and Anderson, C. W. (1988). *Virology* **165,** 377-387.

Simms, H. S. (1926a). *J. Am. Chem. Soc.* **48,** 1239-1251.

Simms, H. S. (1926b). *J. Am. Chem. Soc.* **48,** 1251-1261.

Singh, A. K., and Sonar, S. M. (1988). *Biochem. Biophys. Acta* **955,** 261-268.

Sircar, S., Keyvani-Amineh, H., and Weber, J. M. (1996). *Antiviral Res.* **30,** 147-153.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986). *Nature* **321,** 674-679.

Snijder, E. J., Wassenaar, A. L. M., and Spaan, W. J. M. (1992). *J. Virol.* **66,** 7040-7048.

Snyder, G. H., Cennerazzo, M. J., Karalis, A. J., and Field, D. (1981). *Biochemistry* **20,** 6509-6519.

Snyder, G. H., Reddy, M. K., Cennerazzo, M. J., and Field, D. (1983). *Biochem. Biophys. Acta* **749,** 219-226.

Solé, N. A., and Barany, G. (1992). *J. Org. Chem.* **57,** 5399-5403.

Stein, S., Böhlen, P., Stone, J., Dairman, W., and Udenfriend, S. (1973). *Arch. Biochem. Biophys.* **155,** 202-212.

Stein, S., Böhlen, P., and Udenfriend, S. (1974). *Arch. Biochem. Biophys.* **163,** 400-403.

Stein, S., and Udenfriend, S. (1984). *Anal. Biochem.* **136,** 7-23.

Stephens, C., and Harlow, E. (1987). *EMBO J.* **6,** 2027-2035.

Stewart, P. L., Burnett, R. M., Cyrklaff, M., and Fuller, S. D. (1991). *Cell* **67,** 145-154.

Stillman, B. (1989). *Ann. Rev. Cell. Biol.* **5,** 197-245.

Stillman, B. W., Lewis, J. B., Chow, L. T., Mathews, M. B., and Smart, E. (1981). *Cell* **23,** 497-508.

Stoscheck, C. M. (1990). *Methods Enzymol.* **182,** 50-68.

Stryer, L. (1988). Biochemistry, pp. 1090. Freeman, New York.

Stuiver, M. H., and van der Vliet, P. C. (1990). *J. Virol.* **64,** 379-386.

Sundquist, B., Everitt, E., Philipson, L., and Höglund, S. (1973). *J. Virol.* **11,** 449-459.

Sung, M. T., Cao, T. M., Lischwe, M. A., and Coleman, R. T. (1983). *J. Biol. Chem.* **258,** 8266-8272.

Svensson, U. (1985). *J. Virol.* **55,** 442-449.

Switzer, R. C., III, Merril, C. R., and Shifrin, S. (1979). *Anal. Biochem.* **98,** 231-237.

Tamanoi, F., and Stillman, B. W. (1982). *Proc. Natl. Acad. Sci. USA* **79,** 2221-2225.

Temperley, S. M., and Hay, R. T. (1992). *EMBO J.* **11,** 761-768.

Tibbetts, C. (1977). *Cell* **12,** 243-249.

Tihanyi, K., Bourbonnière, M., Houde, A., Rancourt, C., and Weber, J. M. (1993). *J. Biol. Chem.* **268,** 1780-1785.

Tong, L., Qian, C., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G., and Lagacé, L. (1996). *Nature* **383,** 272-275.

Toogood, C. I., Crompton, J., and Hay, R. T. (1992). *J. Gen. Virol.* **73,** 1429-1425.

Tooze, J. (1981). DNA tumor viruses, pp. 943-1054. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Torchinsky, Y. M. (1981). Sulfur in proteins, pp. 294. Pergamon Press, Oxford.

Toth, M., Doerfler, W., and Shenk, T. (1992). *Nucleic Acids Res.* **20,** 5143-5148.

Towbin, H., Staehelin, T., and Gordon, J. (1979). *Proc. Natl. Acad. Sci. USA* **76,** 4350-4354.

Tremblay, M. L., Déry, C. V., Talbot, B. G., and Weber, J. (1983). *Biochem. Biophys. Acta* **743,** 239-245.

Trentin, J. J., Yabe, Y., and Taylor, G. (1962). *Science* **137,** 835-849.

Tribouley, C., Lutz, P., Staub, A., and Kedinger, C. (1994). *J. Virol.* **68,** 4450-4457.

Udenfriend, S., Stein, S., Böhlen, P., Dairman, W., Leimgruber, W., and Weigele, M. (1972). *Science* **178,** 871-872.

Ulfendahl, P. J., Linder, S., Kreivi, J.-P., Nordquist, K., Sevensson, C., Hultberg, H., and Akusjärvi, G. (1987). *EMBO J.* **6,** 2037-2044.

Van der Eb, A. J., Kestern, L. W., and Van Bruggen, E. F. J. (1969). *Biochem. Biophys. Acta* **182,** 530-541.

van Oostrum, J., and Burnett, R. M. (1985). *J. Virol.* **56,** 439-448.

van Ormondt, H., Maat, J., and Dijkema, R. (1980). *Gene* **12,** 63-76.

Van Woerkom, W. J., and Van Nispen, J. W. (1991). *Int. J. Pept. Prot. Res.* **38,** 101-113.

Varga, M. J., Weibull, C., and Everitt, E. (1991). *J. Virol.* **65,** 6061-6070.

Velcich, A., and Ziff, E. (1985). *Cell* **40,** 705-716.

Velicer, L. F., and Ginsberg, H. S. (1970). *J. Virol.* **5,** 338-347.

Verrijzer, C. P., van Oosterhout, A. W. M., van Weperen, W. W., and van der Vliet, P. C. (1991). *EMBO J.* **10,** 3007-3014.

Vrati, S., Brookes, D. E., Strike, P., Khatri, A., Boyle, D. B., and Both, G. W. (1996). *Virology* **220,** 186-199.

Wadell, G. (1994). Adenoviruses: General Features. *In* "Encyclopedia of Virology" (R. G. Webster and A. Granoff, eds.), pp. 1-7. Academic Press, London.

Weber, J. (1976). *J. Virol.* **17,** 462-471.

Weber, J., Jelinek, W., and Darnell, J. E. J. (1977). *Cell* **10,** 611-616.

Weber, J. M., Dery, C. V., Mirza, M. A., and Horvath, J. (1985). *Virology* **140,** 351-359.

Weber, J. M., and Houde, A. (1987). *Virology* **156,** 427-428.

Weber, J. M., and Tihanyi, K. (1994). *Methods Enzymol.* **244,** 595-604.

Webster, A., Hay, R. T., and Kemp, G. (1993). *Cell* **72,** 97-104.

Webster, A., and Kemp, G. (1993). *J. Gen. Virol.* **74,** 1415-1420.

Webster, A., Leith, I. R., and Hay, R. T. (1994). *J. Virol.* **68,** 7292-7300.

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

Webster, A., Russell, W. C., and Kemp, G. D. (1989b). *J. Gen. Virol.* **70,** 3215-3223.

Weigele, M., DeBernardo, S. L., Tengi, J. P., and Leimgruber, W. (1972). *J. Am. Chem. Soc.* **94,** 5927-5928.

Weinberg, D. H., and Ketner, G. (1986). *J. Virol.* **57,** 833-838.

Weinberger, R. (1993). Practical Capillary Electrophoresis, pp. 312. Academic Press, London.

Wesley, J. (1969). Enzymic Catalysis. *In* "Modern Perspectives In Biology" (H. O. Halvorson, H. L. Roman and E. Bell, eds.), pp. 206. Harper & Row, New York.

White, E. (1993). *Genes Dev.* **7,** 2277-2284.

White, E., and Cipriani, R. (1989). *Proc. Natl. Acad. Sci. USA* **86,** 9886-9890.

White, E., and Cipriani, R. (1990). *Mol. Cell. Biol.* **10,** 120-130.

Wickham, T. J., Mathias, P., Cheresh, D. A., and Nemerow, G. R. (1993). *Cell* **73,** 309-319.

Williams, G. T., and Smith, C. A. (1993). *Cell* **74,** 777-779.

Wiskerchen, M., and Collett, M. S. (1991). *Virology* **184,** 341-350.

Wolf, D. H. (1992). *Experientia* **48,** 117-118.

Wray, W., Boulikas, T., Wray, V. P., and Hancock, R. (1981). *Anal. Biochem.* **118,** 197-203.

Wyckoff, E. E., Hershey, J. W. B., and Ehrenfeld, E. (1990). *Proc. Natl. Acad. Sci. USA* **87,** 9529-9533.

Yabe, Y., Samper, L., Bryan, E., Taylor, G., and Trentin, J. J. (1964). *Science* **143,** 46-47.

Yamamoto, Y., and Kise, H. (1993). *Chem. Lett.* **11,** 1821-1824.

Yamamoto, Y., and Kise, H. (1994). *Bull. Chem. Soc. Jpn.* **67,** 1367-1370.

Yeh-Kai, L., Akusjärvi, G., Aleström, P., Pettersson, U., Tremblay, M., and Weber, J. (1983). *J. Mol. Biol.* **167,** 217-222.

Yih, L.-H., and Lee, T.-C. (1994). *Biochem. Biophys. Res. Comm.* **202,** 1015-1022.

Yonish-Rouach, E., Resnitzky, D., Lotem, J., Sachs, L., Kimchi, A., and Oren, M. (1991). *Nature* **352,** 345-347.

Yoshinaga, S., Dean, N., Han, M., and Berk, A. J. (1986). *EMBO J.* **5,** 343-354.

Young, S. C., White, P. D., Davies, J. W., Owen, D. E. I. A., Salisbury, S. A., and Tremeer, E. J. (1990). *Biochem. Soc. Trans.* **18,** 1311-1312.

Zhai, Z., Wang, X., and Qian, X. (1988). *Cell Biol. Int. Rep.* **12,** 99-108.

Zhang, Y., and Schneider, R. (1993). *Seminars in Virology* **4,** 229-236.

Zhang, Y., and Schneider, R. J. (1994). *J. Virol.* **68,** 2544-2555.