Title
Understanding the role of HIV-1 protease in viral maturation

Author
Rose, Jason Robert

Publication Date
1995

Peer reviewed|Thesis/dissertation

Understanding the role of HIV-1 protease in viral maturation: Application of protein engineering to a target for antiviral chemotherapy.

by

Jason Robert Rosé

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmacology

in the

GRADUATE DIVISION

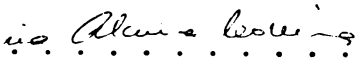
of the

UNIVERSITY OF CALIFORNIA

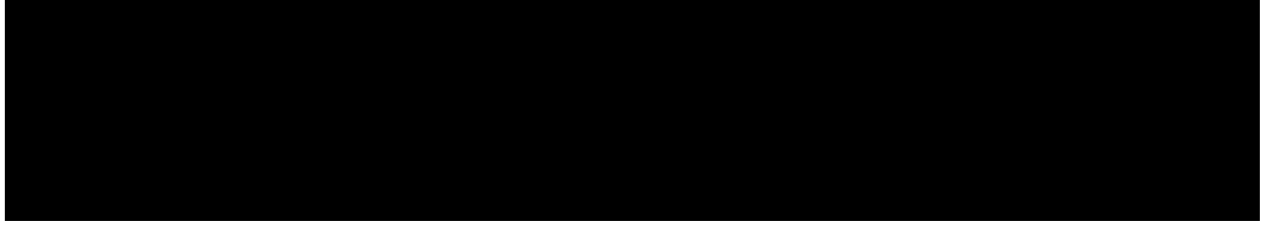
San Francisco

Approved:

Judith M. White



Committee in Charge



To my parents, without whose encouragement
this work would never have been started,

and to Jesus Christ, without whom
this work could never have been finished.

My apologies for the brevity of this preface, but after managing to churn out a tome of this size I tends to find myself with very little left to say. So if you were expecting something erudite and entertaining, I would recommend a trip to the nearest bookstore. I have endeavored to make the work presented here as coherent as possible, although the scientific style presents some limitations that I am sure I have not been able to overcome. This thesis touches on a good number of topics, ranging from bacterial genetics to protein chemistry and, ultimately, virology. I have had the exceptional good fortune to spend the last five years in a laboratory and an institution which has allowed me to draw from a wide variety of disciplines to answer some fundamental questions about the human immunodeficiency virus. I can also thank my undergraduate training at Whitman College for helping me to appreciate the power of a broad-based and multifaceted (dare I say Liberal Arts?) approach to investigating a given problem.

The nature of the problem we are facing with the human immunodeficiency virus is, however, much broader than the issues which are addressed in this thesis. The numerous scientific efforts which are being made to describe the basic character and behavior of HIV are, I believe, essential to our understanding of an infectious agent that will be with us for some time. But I would like to emphasize here that any efforts which we make to understand how this virus works will be meaningless without a corresponding effort to communicate what we know. While I doubt the current epidemic represents some sort of universal Armageddon, it is clear that it will leave us with a lingering harvest of misery and personal disaster that will extend into the next century. Rather than waiting for a cure, we need to emphasize prevention by educating our society and our children so that the spread of this virus is curtailed. Past and present experience suggests that this will certainly be no easy task, but it is ultimately the only road

to true victory. I would encourage those of you reading this work to do everything in your power to support efforts to prevent the spread of this disease.

When I look back on the years I have spent working here, there are more people that come to mind than I have the space to thank. Specifically, I would like to thank Charly for his direction and guidance, as well as for his willingness to allow and encourage my wanderings into the various disciplines I have sampled during my time here. I would also like to thank Robert, Judy, Henry, Don, and the other faculty members here who have challenged me and taught me how to think about science and research.

My thanks to my cohorts and fellow minions in the HIV project - Bethany, Christopher, Dianne, Elaine, Fiona, Gian-Carlo, Nancy, Rafael and Sergio. You have made this project much more interesting and some of the tasks a lot easier with your insights and assistance. My thanks especially to Lilia, who has supported and encouraged my work in virology. You have been a good freind, collaborator, and instructor to me through the last year of this project.

To the other graduate students and post-docs in the Craik lab - Luke, John, Ann, Scott, Chris, David, Qing, Cheng-I and Jeff, for making this enterprise as colorful and entertaining as it has been, despite the tension and stress we've all had to deal with. Thanks for putting up with my occasional hysteria and perpetual hyperactivity. Guess what, it wasn't the coffee...

My thanks also go out to the people who have been a part of my life outside of UCSF. For the friendship of Carolyn Willis, Paul and Mary Anthony, and my brother Eric, who have been willing to stick with me through all the obsessive behavior, the long silences, the demands that this place have made on me and I, in turn, have made on them. To the Sunnyvale crowd - Paul, Elizabeth, Darlene, Roy, Tim, Kathy, Carolyn, and especially Marie - although I haven't known you

as long, your companionship and kindness has meant a lot to me over the last three years. And to the members of the El Camino Bible Church, especially Rich and Ilene Brown, who have taken me in and given me all the love and support that a family could offer.

Last, but far from least, I want to acknowledge the work of the Lord in guiding and directing my endeavors and my life in the last years of this project. To quote from Isaiah 40,

> 29 He gives strength to the weary and increases the power of the weak.
> 30 Even youths grow tired and weary, and young men stumble and fall;
> 31 but those who hope in the Lord will renew their strength. They will soar on wings like eagles; they will run and not grow weary, they will walk and not be faint.

I offer this work as a tribute to the strength and wisdom you have given me.

Understanding the role of HIV-1 protease in viral maturation: Application of protein engineering to a target for antiviral chemotherapy.

Jason Robert Rosé

The human immunodeficiency virus type 1 (HIV-1) encodes a dimeric aspartyl protease which plays a crucial role in the maturation of viral particles to an infectious state. Our interest in this virus stems from the current drive towards developing inhibitors of this protease as therapeutic agents to treat acquired immune deficiency syndrome (AIDS). Current knowledge about the structure and function of the HIV-1 protease, and proteases from related retroviruses, was used, to engineer mutants with specific properties that could be used to study the maturation process. In order to achieve this goal, a heterologous expression system was developed in *Escherichia coli* which allowed the proteases from HIV-1, HIV-2, and SIV_{mac} to be expressed at levels sufficient for biophysical and biochemical analysis. To facilitate biochemical studies of these proteases, protein engineering was used to modify the enzymes to minimize autoproteolysis, a serious hindrance to both crystallographic and enzymatic analysis. Site-directed mutagenesis was further used to engineer protease variants in which overall catalytic activity was diminished 4-fold and 50-fold without altering substrate specificity. The genes for these variant proteases were cloned into HIV-1 viral vectors to quantitate the effects of diminished protease activity on viral function. Our results demonstrate that a reservoir of protease activity exists in the HIV-1 virion, such that inhibition of as much as 75% of the available protease has only a minor effect on viral maturation and activity. A 50-fold reduction in infectivity blocked infectious particle

formation, suggesting that the protease need not be completely inhibited to prevent spread of the virus.

In an attempt to develop an inhibitor capable of producing this level of inhibition while avoiding resistance mutations, we examined a strategy in which defective protease monomers could inhibit the HIV-1 protease. By expressing inactive monomers of the protease in a target cell, inactive heterodimers are formed between wild type and mutant protease monomers, resulting in an inhibition of proteolytic activity and disruption of the viral maturation process. This approach was successful in inhibiting infectious virus production in co-transfection assays, where polyprotein processing could be inhibited by as much as 90% when equal amounts of wild type and mutant-protease containing genomes were used. The loss of proteolytic activity due to heterodimer formation and the concomitant drop in infectious particle titer correlated with the results from the mutational analysis of the protease. The defective monomer was stably expressed in tissue culture cells and shown to reduce infectious particle formation by 5-7 fold following transfection with wild type viral DNA.

By placing these mutant protease monomers in the appropriate gene-delivery systems to achieve high levels of expression, a gene-therapy approach could be developed which would disrupt the function of the protease in infected individuals and limit the spread of HIV-1.

Table of Contents

List of Tables

List of Figures

Introduction. Retroviruses and Retroviral Proteases

History of Retroviruses

The discovery of the viruses now known as retroviruses began with the identification of agents which transmitted cancers in various organisms. As early as 1910, it was observed that a cell-free transmissible agent derived from chicken solid tumors (sarcomas) was able to induce tumors in other chickens (Rous, 1910). This virus, now known as Rous sarcoma virus (RSV), was the first of a number of strains of related viruses capable of causing tumors in chickens. Two viruses capable of inducing tumors in mice were identified between 1935 and 1952. The mouse mammary tumor virus (MMTV) was identified initially by the observation that the milk of a strain of inbred mice could induce cancers in other mice (Bittner, 1936). This agent could be filtered and heat inactivated, properties consistent with it being a viral agent (Anderwont and Bryan, 1944). A second virus, the murine leukemia virus (MuLV), was identified in another strain of inbred mice with a particularly high incidence of leukemia. Filtrates of these leukemias were capable of inducing similar leukemias in otherwise healthy mice (Gross, 1951). The presence of RNA viruses capable of causing tumors in laboratory animals led to the search for similar viruses in other organisms. Viral particles were found to be present in tissues from a wide variety of species, ranging from cats to baboons and other primates.

The nature of the replicative cycle of RNA tumor viruses became clear in 1970 when the virally encoded enzyme reverse transcriptase (RT) was discovered (Baltimore, 1970; Temin and Mizutani, 1970). This enzyme, isolated from MuLV and RSV, was shown to transcribe viral RNA into a double-stranded DNA form. Temin had proposed that RNA tumor viruses were capable of stably integrating into host cell genomes, thus conferring inherited changes in the host cells (Temin, 1964). The discovery of reverse transcriptase gave credence to this hypothesis

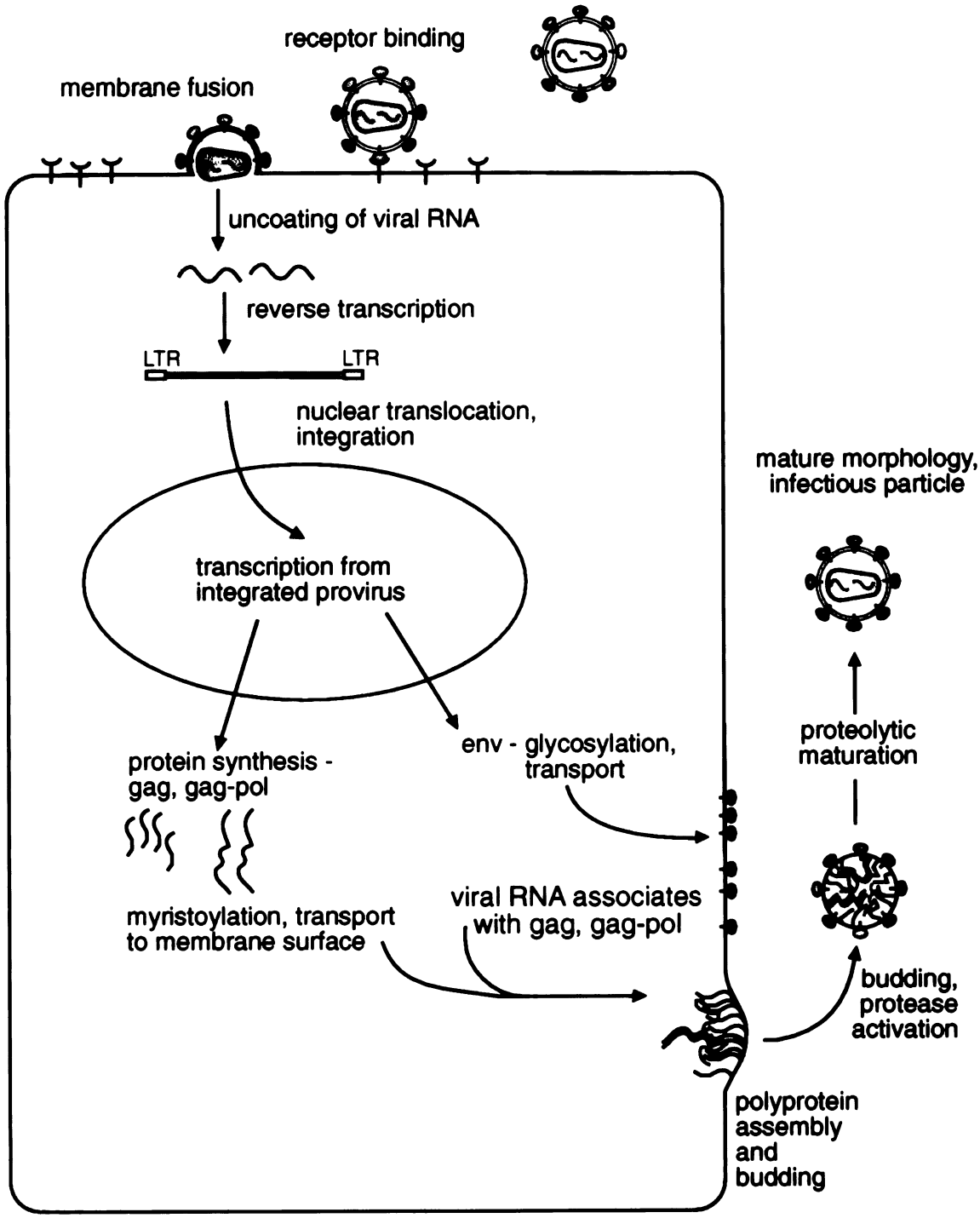
and helped to define some of the early events in the viral life cycle. The RNA tumor viruses, now known as retroviruses, were defined as a family of viruses having a membrane derived from the host cell, two copies of single-stranded genomic RNA, and a virally encoded reverse transcriptase.

The current model for the major events in the life cycle of a prototypic retrovirus are schematized in Figure i-1. They can be split into two sections - events which take place after infection and lead up to integration of the viral genome (early events), and events which result in the production of new virus from integrated proviral genomes (late events). In the early stages of infection, the virus gains entry into the cell by specific recognition of a cellular receptor by the viral envelope glycoprotein. Upon internalization the viral core is disassembled, freeing the genomic RNA-nucleoprotein complex for reverse transcription. The viral RNA is converted to double-stranded DNA by the action of reverse transcriptase, producing the viral long terminal repeats (LTRs) which contain the transcriptional signals necessary for reactivation of the virus. The proviral genome is integrated into the genome of the target cell through the action of the virally encoded integrase.

When the provirus is reactivated, transcription from the viral LTR yields a primary RNA transcript, used to direct synthesis of the viral structural (gag), enzymatic (pol) and envelope (env) proteins. The *gag* and *pol* reading frames are translated into two precursor forms: gag, in which only the structural proteins matrix (MA), capsid (CA) and nucleocapsid (NC) are present, and gag-pol, in which the enzymatic proteins reverse transcriptase (RT), integrase (IN), and usually the protease (PR) are fused to the structural proteins by means of a frameshift during translation of the viral RNA. The location of the viral protease

Figure i-1: Schematic of the retroviral replication cycle.

The virus enters the host cell by means of a specific receptor. Upon entry the viral core is exposed and the genomic RNA is uncoated, allowing reverse transcription to occur. The reverse transcription reconstructs the viral long terminal repeats (LTR), which contain the viral promoters and integration sites. DNA is integrated into the genome of the host cell to form a stable provirus. Upon reactivation, messages for the viral gag, gag-pol, and env proteins are transcribed from the viral long terminal repeat (LTR). Viral structural and enzymatic proteins are synthesized in the cytoplasm and transported to the membrane surface. Full length viral RNA is packaged by the viral gag proteins during transit to the cell membrane. Virions bud from the surface of the infected cell and acquire the envelope glycoprotein during budding. Activation of the viral protease at this stage leads to maturation of the viral proteins and condensation of the core into the mature morphology associated with infectious particles.


receptor binding
membrane fusion
uncoating of viral RNA
reverse transcription
LTR
LTR
nuclear translocation,
integration
transcription from
integrated provirus
protein synthesis -
gag, gag-pol
myristoylation, transport
to membrane surface
env - glycosylation,
transport
viral RNA associates
with gag, gag-pol
polyprotein
assembly
and
budding
budding,
protease
activation
proteolytic
maturation
mature morphology,
infectious particle

gene varies, sometimes being associated with the *gag* reading frame and sometimes with the *pol* reading frame. The synthesis of gag-pol precursors takes place by a number of mechanisms, involving translational read-through and ribosomal frameshifting. These strategies ensure that the enzymatic proteins are synthesized at levels between five and 10 percent of the gag proteins (Jacks, et al., 1988; Jacks, et al., 1987; Jacks and Varmus, 1985) (Figure i-2).

Viral polyprotein precursors are directed to the plasma membrane by means of an N-terminal myristoylation signal. During this transit process, the nucleocapsid protein associates with full length viral RNAs which are packaged as viral genomic RNAs in the mature virus. The envelope glycoprotein is translated separately and targeted to the plasma membrane through the endoplasmic reticulum, where it is glycosylated. Assembly of the gag and gag-pol polyproteins at the membrane results in budding of an immature viral particle, in which the viral protease is activated by dimerization of the polyproteins. Proteolytic processing of polyproteins results in condensation of the viral core to give a mature and infectious virion, which is schematized in Figure i-3.

Human retroviruses and HIV

The search for retroviral pathogens associated with human cancers was initiated as part of the ongoing search for new retroviral agents in the 1970's. This endeavor was unsuccessful until 1980, when developments in tissue culture techniques permitted the growth of human T-lymphocytes isolated from patients with acute T-cell leukemia. These cells produced viruses with a morphology similar to those of retroviruses from other mammalian species, and which contained reverse transcriptase activity (Poiesz, et al., 1980). In addition, viral

Figure i-2: General genomic organization for retroviruses.

The different organizations of the major viral proteins in the Rous sarcoma virus (RSV), human immunodeficiency virus (HIV), visna lentivirus (VISNA), and bovine leukemia virus (BLV) are shown. Each box represents a separate open reading frame. The gene for the viral protease (PR) is shown in black. In RSV-like viruses the protease is in frame with the *gag* genes and is synthesized as part of gag (Schwartz, et al., 1983). In HIV-like viruses the protease is synthesized as part of a gag-pol fusion generated by a frameshift which occurs with approximately 10% efficiency and places the *pol* genes in the same reading frame as the *gag* genes (Jacks, et al., 1988). In MuLV-like viruses the protease is synthesized in a gag-pol precursor which is generated by suppression of the *gag* stop codon (Yoshinaka, et al., 1985). In BLV-like viruses the protease is synthesized as a gag-PR fusion generated by a frameshift readthrough of the gag termination codon (Yoshinaka, et al., 1986).

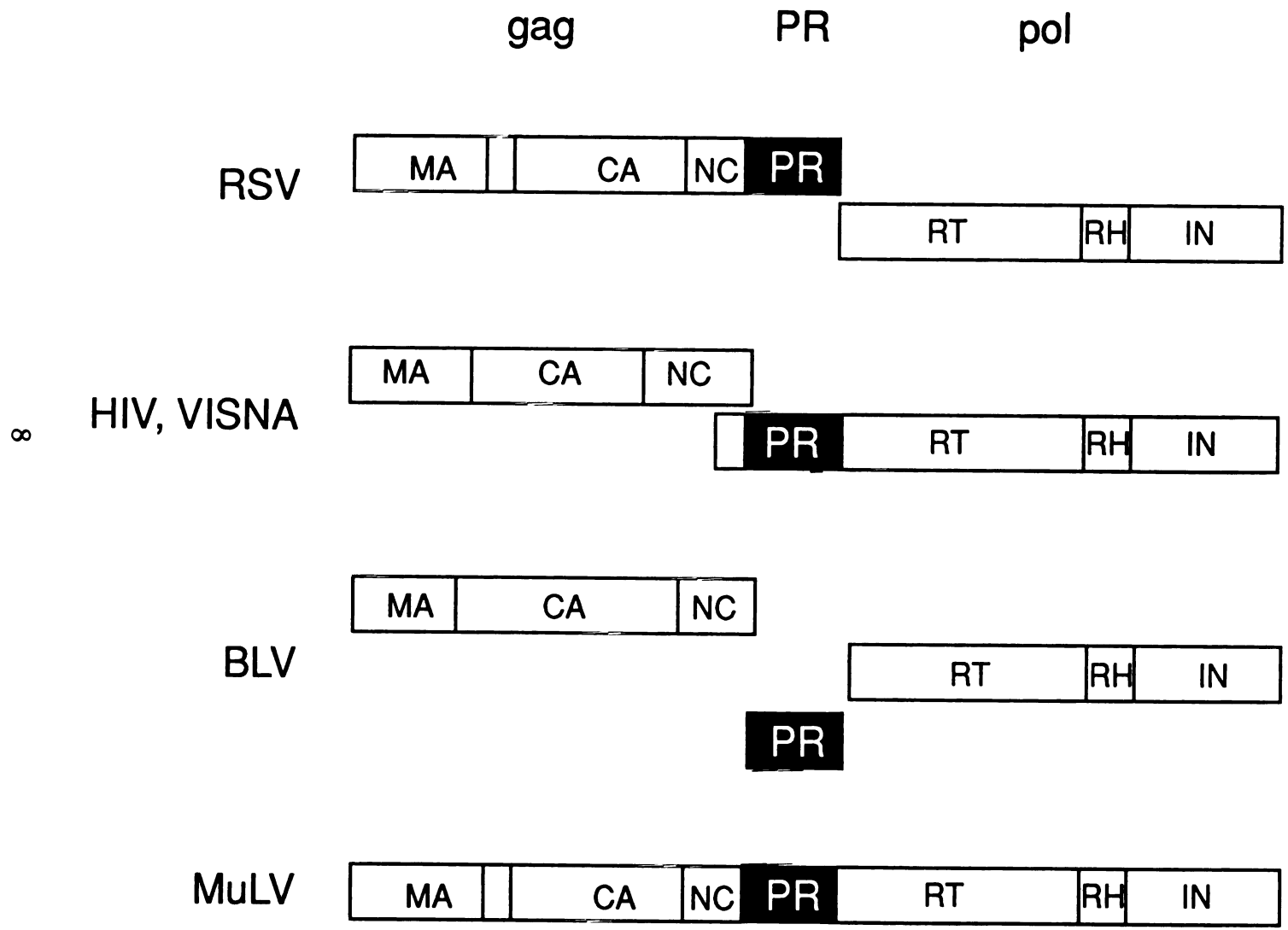
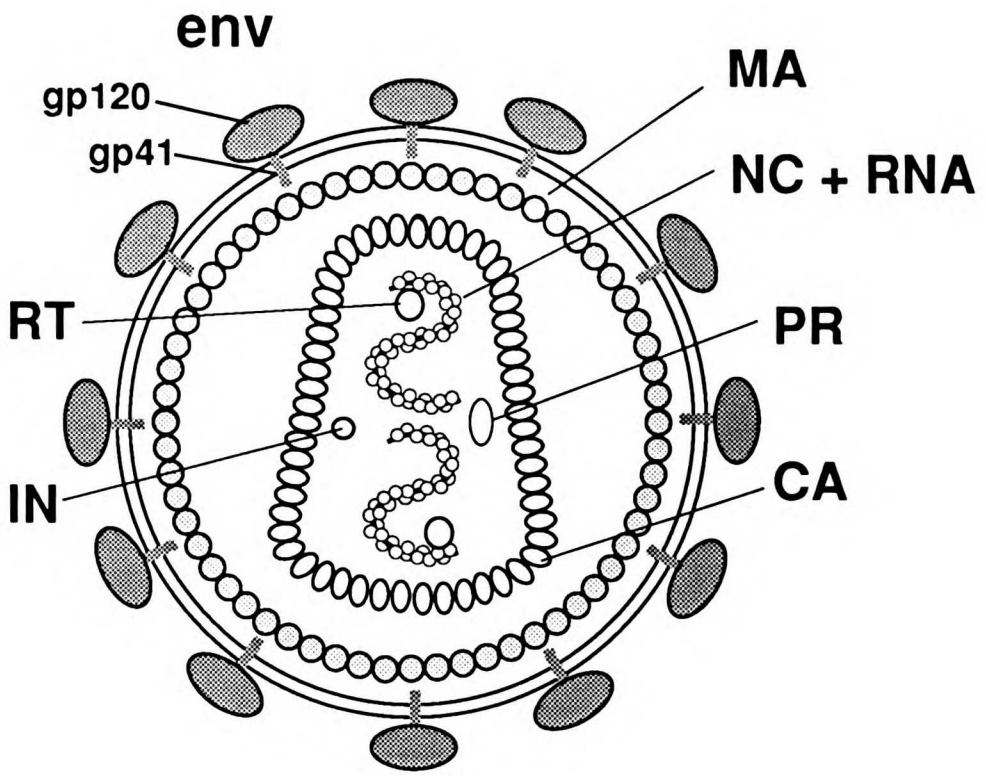

gag
PR
pol
RSV
MA
CA
NC
PR
RT
RH
IN
HIV, VISNA
MA
CA
NC
PR
RT
RH
IN
BLV
MA
CA
NC
RT
RH
IN
PR
MuLV
MA
CA
NC
PR
RT
RH
IN

Figure i-3: Morphology of a retroviral particle.

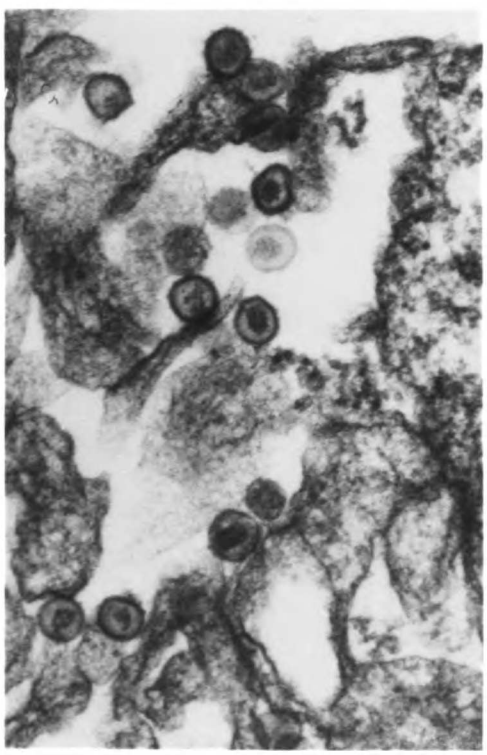
A). Schematic of HIV-1. The core of the viral particle consists of a shell of the capsid protein (CA) which surrounds the nucleocapsid (NC) coated genomic RNA and the enzymatic proteins: protease (PR), reverse transcriptase (RT) and integrase (IN). The core is surrounded by a second shell of matrix (MA) protein. The virus is coated by a lipid bilayer derived from the host cell in which the receptor-specific envelope glycoprotein (env), consisting of a surface component (gp120) and transmembrane component (gp41), is embedded.

B, C) Thin section electron micrographs of mature (B) and protease-inhibitor treated (C) HIV-1 virions released from COS-7 cells. Samples of cells producing viral particles were fixed, thin-sectioned, stained with uranyl acetate, and contrasted with Reynolds lead stain. Virions were examined on a JEOL 1200 EX electron microscope operating at 80 kV. In mature viral particles the condensed cores and outer layer of matrix protein are clearly visible. The thicker ring of electron-dense material in immature virus samples corresponds to a layer of unprocessed gag protein. Magnifications are x20,000 and x25,000 in B) and C), respectively.

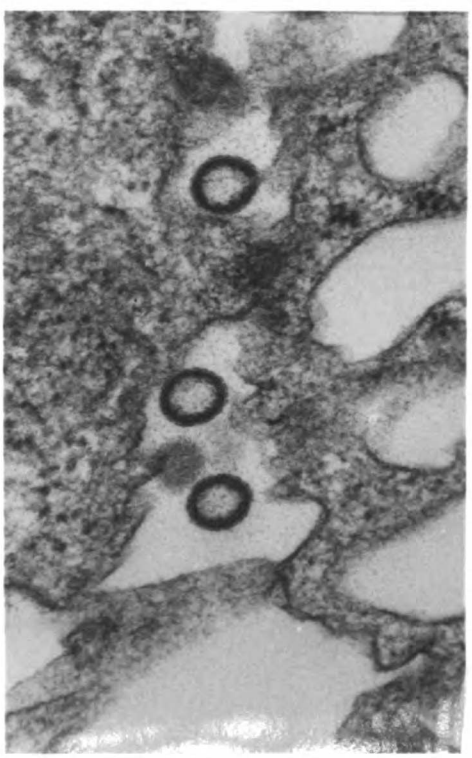
A.



B.



C.



particles could be isolated directly from T-cells in the peripheral blood of patients, as well as in cultured cell lines. This virus, known as human T-cell leukemia virus (HTLV) was joined in 1982 by a second virus which caused a T-cell variant of hairy cell leukemia. This virus was clearly distinct from HTLV, and was designated HTLV-II. The discovery of these viruses indicated that retroviruses could clearly be considered as important human pathogens.

In the early 1980's a transmissible immunodeficiency syndrome was observed in human patients. This syndrome resembled an immunodeficiency disease present in cats infected with the feline leukemia virus (FeLV) (Jarrett, et al., 1964). The then-recent isolation of human retroviruses associated with T-cell leukemias led researchers to suspect that a retrovirus might be responsible. Using the tissue culture methodologies developed during the discovery of HTLV-I and HTLV-II, a retrovirus was identified in peripheral blood lymphocytes from patients with acquired immunodeficiency syndrome (AIDS) (Barre-Sinoussi, et al., 1983; Gallo, et al., 1983). This virus contained RNA and reverse-transcriptase activity, suggesting it was a member of the HTLV family, although morphological and serological analysis showed it was clearly distinguishable from these viruses (Schüpbach, et al., 1984). Upon cloning and sequencing of the virus, tentatively designated HTLV-III, it was apparent that the virus was clearly distinct from the HTLVs in terms of the arrangement of its open reading frames and the presence of a number of previously unidentified open reading frames at the 3' end of the viral genome (Ratner, et al., 1985). In addition, the nature of the disease induced by this virus was more similar to the slow immunodeficiencies associated with the lentivirus subgroup of retroviruses than the cancers induced by the oncoviruses. The lentivirus subgroup includes VISNA, equine infectious anemia

virus (EIAV) and several simian immunodeficiency viruses (SIV) The virus was renamed human immunodeficiency virus, or HIV (Coffin, et al., 1986).

Following the discovery of HIV another human retrovirus was identified and isolated from patients in West Africa (Clavel, et al., 1986). This virus, now known as HIV-2, had similar morphological and biological properties to the original virus, now named HIV-1, but was serologically distinct and contained a different organization of the open reading frames downstream of the *pol* genes. HIV-1 and HIV-2 are approximately 50% identical at the nucleic acid level, and share approximately 60% identity at the amino acid sequence level in the *gag* and *pol* regions (Guyader, et al., 1987). It was also noted that HIV-2 was very closely related to a simian retrovirus STLV-III_{mac} (SIVmac) (Daniel, et al., 1985; Kanki, et al., 1985), which was shown to cause an AIDS-like disease in captured rhesus macaques (Letvin, et al., 1985). It is clear that HIV is one of a large family of primate retroviruses which share similar cell tropism, pathogenesis, and genetic organization (Desrosiers, et al., 1989; Doolittle, et al., 1990).

Early studies on retroviral proteases

Original work on the retroviral proteolytic maturation process was carried out in the avian oncogenic retroviruses RSV and avian myeloblastosis virus (AMV), and the murine viruses MuLV and MMTV. As early as 1977 a major component of the RSV structural proteins, p15, was shown to be associated with the ability to process the $p76^{gag}$ precursor into the mature p27, p19, and p12/15 proteins (von der Helm, 1977). Virion lysates were also shown to contain this processing activity, suggesting that the responsible protease was packaged in virions. Similar results were observed in MuLV, in which a soluble, low molecular weight processing factor could be purified from virion lysates

(Yoshinaka and Luftig, 1977). Unlike RSV, however, the protease was not a major component of the virus.

Studies of the avian viral processing enzyme were facilitated by the large amounts of p15 protein in isolated virions. The processing activity purified from these virions was found to be highly specific for viral polyprotein substrates, and the processing reaction appeared to proceed optimally at low pH (4-7) and high ionic strength. When the *in vitro* processing reaction was probed with protease inhibitors it was found that it could be blocked by sulfhydryl modifying reagents and metal ions. Initial conclusions from these studies were that the viral proteases were similar to the papain family of cysteine proteases (Dittmar and Moelling, 1978; Sen, et al., 1980).

Genetic analysis was used to demonstrate that the p15 protein was the viral protease. In polyproteins from which p15 had been deleted no processing was observed, but mature forms of the polyprotein could be generated by addition of wild type viral lysates (Vogt, et al., 1979). By 1983, methodologies had been developed which permitted the expression of the RSV *gag* genes in *E. coli*. It was shown that bacteria producing the $p76^{gag}$ gene product contained only mature forms of the viral proteins. Disruption of the p15 coding region resulted in the accumulation of the precursor protein (Mermer, et al., 1983). This confirmed that p15 could process the viral proteins in a heterologous system, and established the use of bacterial expression as a means of probing viral protease function.

Concurrent genetic studies in the MuLV system led to the discovery that in this virus the protein containing protease activity was found not as part of the *gag* genes, but rather as part of the *pol* coding region (Figure i-2) (Levin, et al., 1984). Although deletions in this region did not affect particle formation or assembly, they prevented particle maturation and blocked infectivity at an early

stage in infection (Crawford and Goff, 1985; Katoh, et al., 1985). Examination of the amino acid sequences of the protease region showed that a high degree of homology existed with these sequences and the protease sequences from retroviruses of feline and simian origin (Yoshinaka, et al., 1985). The identity of the active site residues, however, remained unclear.

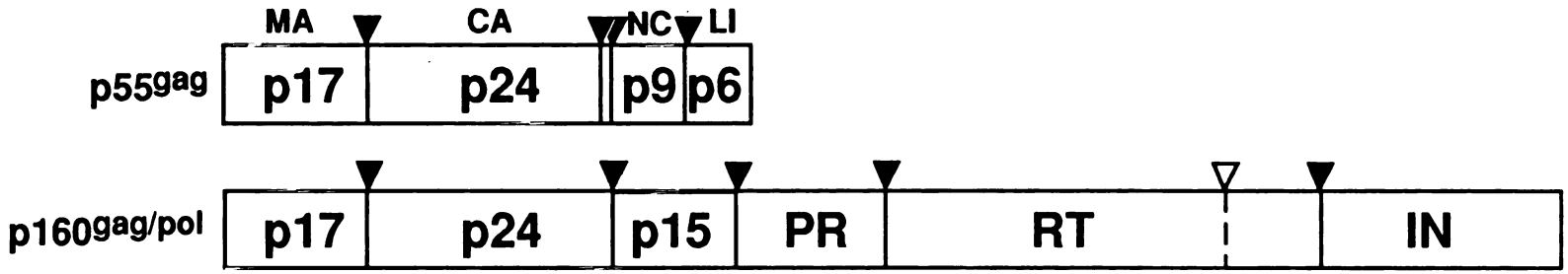
Analysis of the HIV-1 protease

When HIV-1 was originally sequenced, a protease-like sequence at the 5' end of the *pol* gene was readily identified by comparison with the sequences of other retroviruses (Ratner, et al., 1985). The prediction that this was the viral protease was borne out by the observation that deletion of nucleotide sequences at the 5' end of the HIV-1 *pol* gene blocked processing of recombinant $p160^{gag\text{-}pol}$ in a yeast expression system (Kramer, et al., 1986). Expression of smaller fragments of the *gag-pol* reading frames in *Escherichia coli* resulted in the production of a 10kDa protein which was competent to authentically process the $p55^{gag}$ precursor protein (Debouck, et al., 1987; Giam and Boros, 1988; Graves, et al., 1988) Protease was purified from virus and shown to contain the same amino terminus as material expressed in recombinant expression systems (Lillehoj, et al., 1988), indicating that the protease could be efficiently and authentically processed in heterologous organisms as had been seen for other retroviral proteases. The processing sites for the HIV-1 protease have been determined and are shown in Figure i-4.

With the addition of the amino acid sequences from the human retroviruses HTLV-I, HTLV-II, and HIV-1, as well as a number of retroviruses from other species, Toh, *et al.* (Toh, et al., 1985) were able to identify the consensus sequence Asp-Thr/Ser-Gly which aligned with the active site of the cellular aspartyl

Figure i-4: Protease cleavage sites in the HIV-1 polyproteins.

The sites identified as cleavage sites for the viral protease in the $p55^{gag}$ and $p160^{gag\text{-}pol}$ precursor proteins are shown by arrows. The white arrow in reverse transcriptase (RT) identifies a site between the reverse transcriptase and ribonuclease H domains which is cleaved only once in the dimeric reverse transcriptase/RNAse H precursor to form a heterodimer with two reverse transcriptase domains and one RNAse H domain. MA = matrix (p17); CA = capsid (p24); NC = nucleocapsid (p9); LI = linker (p6); PR = protease; IN = integrase.



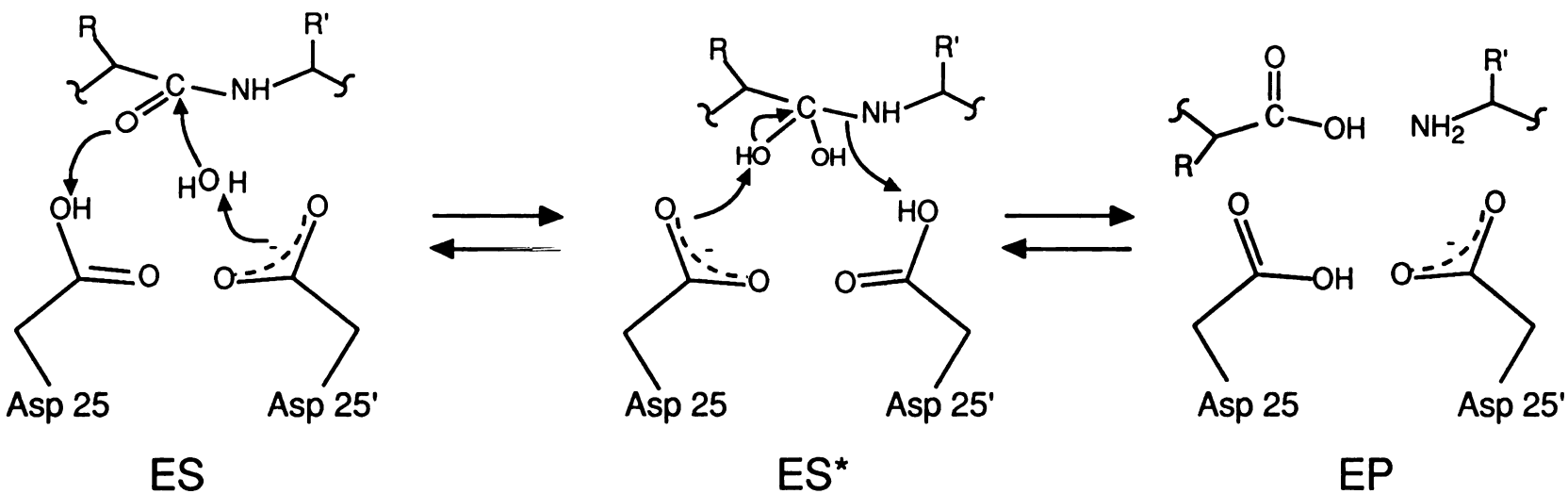
		P4	P3	P2	P1	↓	P1'	P2'	P3'	P4'
gag	p17/p24	Ser	Gln	Asn	Tyr		Pro	Ile	Val	Gln
	p24/internal	Ala	Arg	Val	Leu		Ala	Glu	Ala	Met
	internal/p9	Ala	Thr	Ile	Met		Met	Gln	Arg	Gly
	p9/p6	Pro	Gly	Asn	Phe		Leu	Gln	Ser	Arg
gag-pol	p15/protease	Ser	Phe	Asn	Phe		Pro	Gln	Ile	Thr
	protease/RT	Thr	Leu	Asn	Phe		Pro	Ile	Ser	Pro
	RT/RNAse H	Ala	Glu	Thr	Phe		Tyr	Val	Asp	Gly
	RT/Integrase	Arg	Lys	Ile	Leu		Phe	Leu	Asp	Gly

protease pepsin (Toh, et al., 1985). Based on the assumption that the *pol* encoded protease was an aspartyl protease, a structural model was derived suggesting that the protease would function as a dimer of two 99-amino acid monomers, each contributing one of the two catalytic aspartic acid residues (Pearl and Taylor, 1987). Using recombinantly produced HIV-1 protease it was demonstrated that the enzyme was, in fact, functional as a dimer. It was inhibited efficiently by 1,2-epoxy-3-(4-nitrophenoxy) propane (Meek, et al., 1989), a specific inhibitor of aspartyl proteases (Tang, 1971), confirming its identity as a member of this family. The X-ray crystallographic solution of the structures of virally purified RSV (Miller, et al., 1989) and recombinant HIV-1 protease (Navia, et al., 1989; Wlodawer, et al., 1989) confirmed both the dimeric nature of the retroviral enzymes and clearly demonstrated their similarity to the mammalian and microbial aspartyl proteases. It has since been demonstrated that metal ions can bind at the active site of the HIV-1 protease and inhibit its activity (Zhang, et al., 1991). Although HIV-1 protease can be inactivated by specific modification of sulfhydryls remote from the active site (Karlström, et al., 1993), not all retroviral proteases contain these sulfhydryls, calling into question some of the data which originally suggested the proteases were papain-like cysteine proteases.

The aspartyl proteases are so-named by virtue of the two aspartic acid groups which serve the primary catalytic role during the cleavage of the peptide bond. The low pH optimum for these enzymes, which ranges from 1.5 to 5.5, creates a situation at the active site where one aspartic acid is protonated while the other is deprotonated. As shown in Figure i-5, a water molecule situated between the two catalytic aspartates serves as the nucleophile following abstraction of a proton by the negatively charged aspartic acid. Unlike the cysteine and serine proteases, the intermediate is not covalently associated with the enzyme.

Figure i-5: Proposed catalytic mechanism of the HIV-1 protease.

The binding of a peptide substrate in the binding pocket of the HIV-1 protease generates the enzyme-substrate complex (ES). In the presence of the peptide bond, Asp 25' acts as a general base to remove a proton from a water molecule positioned between the two catalytic aspartates. A simultaneous proton transfer from Asp 25 assists the hydration of the peptide bond to form the enzyme-bound amide hydrate intermediate shown in ES*. The cleavage process is completed through the concerted action of Asp 25', which now acts as a general acid to donate a proton to the amide nitrogen, and Asp 25, which acts as a general base and deprotonates one of the OH groups at the scissile bond. This results in the collapse of the hydrated carbonyl and elimination of the amine moiety to generate the enzyme-product complex (EP). Product release frees the enzyme to continue its catalytic function. Mechanism shown adapted from (Hyland, et al., 1991).


R
R'
C
NH
O
HO H
OH
O
O
O
Asp 25
Asp 25'
ES
HO
OH
HO
Asp 25
Asp 25'
ES*
OH
NH2
Asp 25
Asp 25'
EP

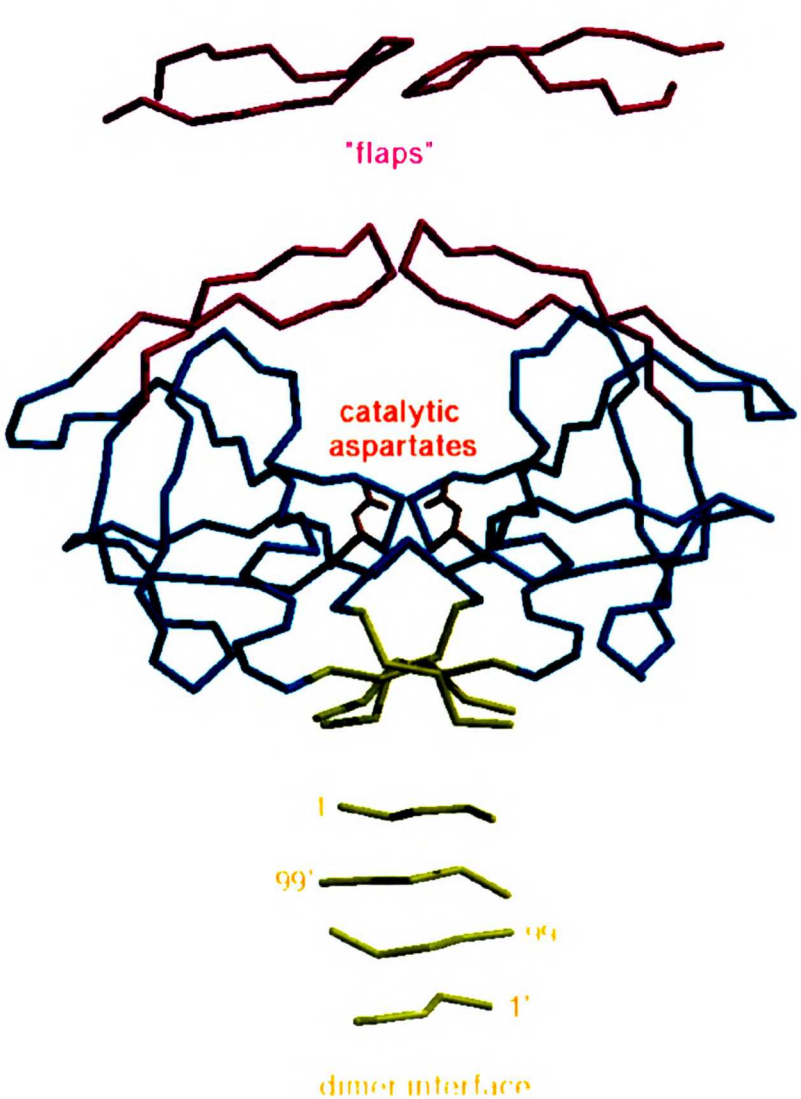
Donation of a proton to the amide nitrogen of the peptide bond results in the collapse of the hydrated intermediate and cleavage of the peptide bond, followed by product release (Hyland, et al., 1991).

The structure of the HIV-1 protease reveals some of the novel features of this enzyme (Figure i-6). The interface between the two monomers is a four-stranded anti-parallel β-sheet composed of two β-strands from each monomer. These strands represent the N- and C-terminal 5 amino acids of each polypeptide and interdigitate to form a specific and high-affinity interaction. The second novel feature of this enzyme is a pair of β-strands, comprising residues 43-58, which form a "flap" covering the top of the substrate binding site. Upon substrate binding these flaps exhibit a considerable conformational change, moving approximately 7 Å to close down upon the substrate, leading to a structure with a smaller active site cavity. The catalytic aspartic acid residues are located at the base of the active site cavity in a large hydrophobic channel which forms the substrate binding pocket.

The binding pocket of the HIV-1 protease can be conceived of as a hydrophobic tube with a length of approximately 25 Å and a diameter of 8 Å. This pocket is long enough to accommodate 8 amino acids in an extended conformation, which gives the protease a large specificity determinant (Darke, et al., 1988). In order to facilitate discussion of substrate binding and recognition, the nomenclature of Schechter and Berger will be used (Schechter and Berger, 1967). In this nomenclature system, the amino acids in a substrate are labeled P_n, where n indicates the distance from the scissile bond. Residues on the C-terminal side of the scissile bond are designated P_n'. Thus, as shown in Figure i-4, the substrate sequences extend from P4 to P4' around the scissile bond. The subsites

Figure i-6: Structural features of HIV-1 protease.

The central figure shows an α-carbon backbone trace of the HIV-1 protease dimer, with the C_2 symmetric axis of the dimer lying vertically in the plane of the page. The side chains for the catalytic aspartate residues are displayed, and lie in the base of the active site cavity. The dimer interface is shown in an end-on view below the dimer, and is composed of four β-strands, two from each monomer. The N- and C-terminal residues (1 and 99, respectively) are numbered, with primes (') indicating residues from one monomer. The flap regions from amino acids 43 to 58 are shown in end-on view above the dimer.


"flaps"
catalytic
aspartates
1
99'
99
1'
dimer interface

on the enzyme which recognize these residues are labeled S, where S_n corresponds to the pocket which binds the P_n residue. The interactions between a model peptide and the HIV-1 protease binding pocket are shown in Figure 1.7, with the residues colored differentially to indicate interactions with different subsites.

HIV-1 protease as a therapeutic target

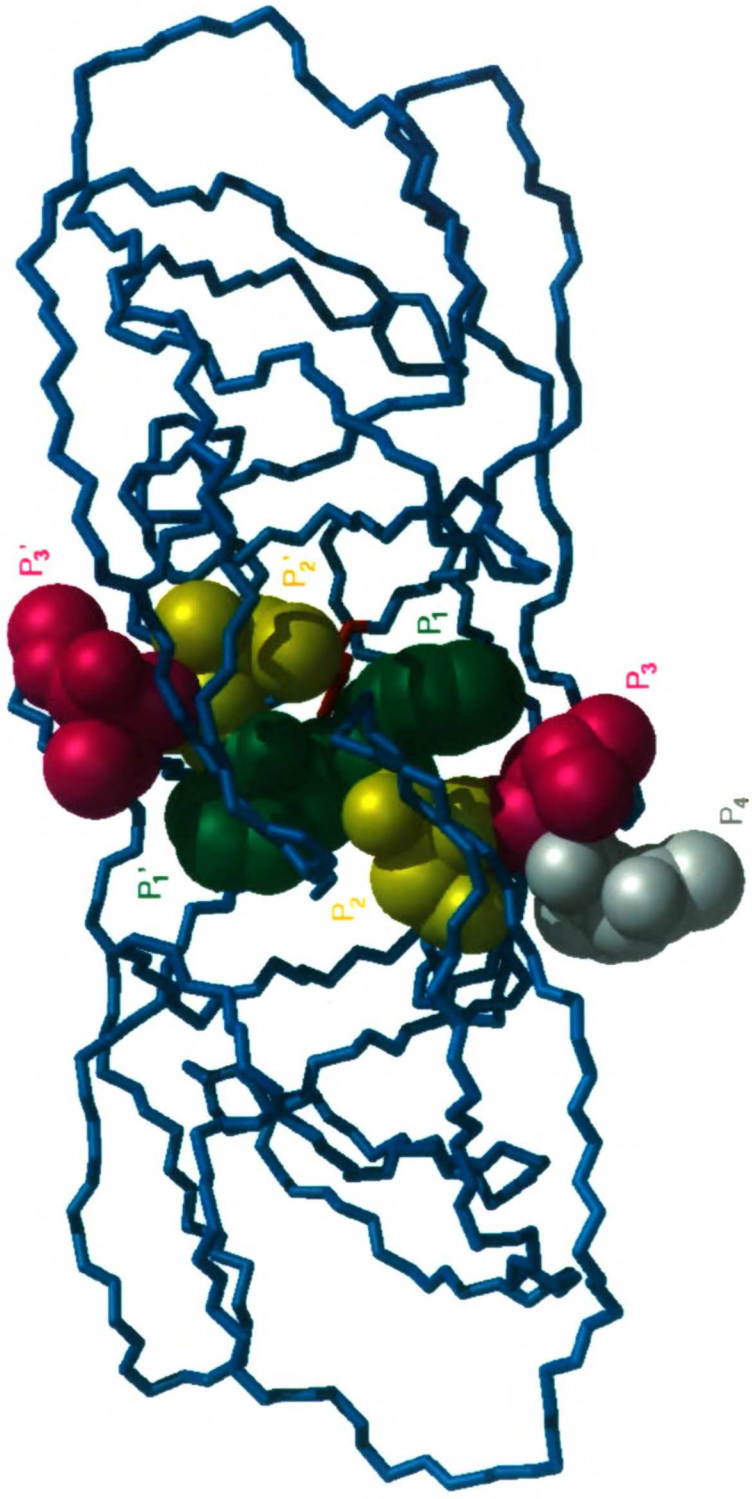
The essential role of the HIV-1 protease in producing infectious virus was demonstrated by specific mutagenesis studies, in which the active site aspartic acid was mutated to alanine or asparagine (Kohl, et al., 1988; Mous, et al., 1988). Viruses containing these mutations were nonviable and exhibited no processing of the viral polyproteins. Morphological analysis of viruses harboring protease mutants showed that virions displayed an aberrant morphology, in which the condensed central core was not visible (Göttlinger, et al., 1989; Peng, et al., 1989).

The essential nature of this protease makes it a possible target for chemotherapeutic intervention, and the development of inhibitors has been pursued successfully. At present, inhibitors exist which are capable of blocking HIV-1 replication in both *in vitro* and *in vivo* model systems (Martin, et al., 1994; Tomasselli, et al., 1991; Wlodawer and Erickson, 1993). Recent reports suggest, however, that the protease will tolerate a number of mutations in its active site which decrease inhibitor binding and affect proteolytic activity, but permit viral replication to continue (El-Farrash, et al., 1994; Ho, et al., 1994; Otto, et al., 1993; Sardana, et al., 1994).

Chemotherapy of HIV-infected individuals with anti-reverse transcriptase agents has revealed that mutation yields resistance to small molecule inhibitors rapidly in a clinical setting (Larder, 1993). This rapid mutation rate raises the

Figure i-7: Peptide binding site in the HIV-1 protease.

Structure of the peptidomimetic inhibitor JG-365 bound at the active site of the HIV-1 protease (Swain, et al., 1990). The sequence of the inhibitor is Ac-Ser-Leu-Asn-Phe-ψ[CH(OH)CH_2N]-Pro-Ile-Val-OMe. The inhibitor is shown as a van-der Waals surface representation. The residues are shown in colors as follows: P1, P1' in green; P2, P2' in yellow; P3, P3' in magenta, P4, in gray. Aspartates are shown in red at the base of the binding pocket. The protein backbone is shown as a blue chain tracing connecting the amide nitrogen, alpha carbon, and carbonyl carbon of each amino acid.


P3'
P2'
P1
P3
P4
P1'
P2

specter of resistance mutations arising in the viral protease during therapy with protease inhibitors. We have hoped to use the HIV-1, HIV-2 and SIVmac proteases as starting points from which to conduct structure-function analyses to understand more about the biology and specificity of the retroviral aspartyl proteases. With specific regard to the issue of enzyme inhibition, we intended to study the requirement of the human immunodeficiency virus for proteolytic processing: what level of protease activity is necessary and sufficient to produce an infectious virus? We feel that answers to this question will play a key role in determining the nature and application of inhibitors which target the protease as a means of preventing the maturation and spread of the virus.

To answer this question we have attempted to use current knowledge about the structure and function of the HIV-1 protease to engineer mutants with specific properties that could be used to study the maturation process. In order to do this, a heterologous expression system was developed which allowed the proteases from HIV-1, HIV-2, and SIV_{mac} to be expressed and purified at levels sufficient for biophysical and biochemical analysis. To facilitate biochemical studies of these proteases, a protein engineering approach was used to modify the enzymes to minimize autoproteolysis, a serious hindrance to both crystallographic and enzymatic analysis. Site-directed mutagenesis was further used to engineer protease variants in which overall catalytic activity was diminished without altering substrate specificity. The genes for these proteases were cloned back into viral vectors in order to quantitate the effects of diminished protease activity on viral maturation and infectivity. In the last part of this project we describe a genetic approach in which infectious virus production may be blocked in a highly efficient manner by expressing inactive

Chapter 1. Expression of the HIV-1, HIV-2 and SIV Proteases

Abstract

In order to study the structure and function of retroviral proteases an expression system was needed which could produce reagent quantities of wild type and mutant proteases. This precluded the use of a fusion protein expression system, in which high level translation of the fusion partner directed the production of a polyprotein from which the protease was freed by autoproteolytic processing. A bicistronic gene arrangement was constructed to taking advantage of the high translation rates of fusion partners without requiring a translational fusion between the fusion partner and protease. In this arrangement, ribosomes translating the highly expressed fusion partner (primary cistron) terminate and are recaptured by a synthetic ribosome binding site (RBS) at the end of the gene. These ribosomes then translate the second gene, which consists of the selected protease gene. This arrangement was shown to be capable of directing high levels of expression of the HIV-1 protease as a 99-amino acid polypeptide. Expression levels could be varied five to seven-fold by varying the distance between the RBS and the protease gene, and two to three-fold by changing the expression level of the primary cistron. The expression system could be extended to the expression of the related proteases from HIV-2 and SIVmac, although it was discovered that the A:T content of the 5' ends of these genes could affect expression levels as much as four-fold. The bicistronic expression system should not only provide a means for expressing wild type and mutant forms of the retroviral proteases, but also other proteins which would ordinarily require post-translational processing to yield native material for structure-function studies.

Choice of an organism for heterologous expression.

Biochemical studies of the HIV-1, HIV-2, and SIV proteases require that amounts of protein in the milligram range be available for kinetic analysis and X-ray crystallography. Since it is not feasible to purify these proteases from infected cells or purified virions, it is necessary to pursue methods of expressing the enzymes recombinantly. A number of host systems currently exist for the production of recombinant proteins: *Escherichia coli*, the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*, and the insect-cell based baculovirus.

We chose to attempt expression in *Escherichia coli* for several reasons. Detailed work has been done in E. coli regarding genetic manipulation. Over 150 promoters have been identified in the course of studying prokaryotic expression (Hawley and McClure, 1983) and the molecular mechanisms of regulation have been extensively characterized for several of them. By exploiting past research in *E. coli* on promoter function, genetic background, and growth conditions, it is possible to produce proteins at levels as high as 10 - 40% of the cell mass. This amplification greatly improves our ability to obtain reagent quantities of pure protease for study.

From the standpoint of optimizing protein expression, *E. coli* has the advantage of being easy to engineer at the genetic level using the technology currently available. The rapidity with which recombinant technology can be applied to this organism allows for the consideration and manipulation of a wide array of genetic factors influencing protein expression. As an organism, *E. coli* has advantages which make it a preferred expression host. Bacteria are not only easy to grow, but their maintenance is very inexpensive relative to eukaryotic hosts. Its simple growth requirements also mean that scale-up to high level production is feasible. Many fermentation devices are available which can be

used to maximize the growth and expression potential of the organism in high density fermentations. Since these allow for the production of very large quantities of bacteria in small volumes, even a low-level expression system can provide significant amounts of material for study.

Factors influencing expression in *Escherichia coli*.

In the course of planning the production of the HIV-1, HIV-2, and SIV proteases we considered a number of factors which affect the level of expression of heterologous genes in *E. coli*. Listed below are a number of these factors which appear to be most important for expressing a heterologous gene, and current opinions as to how they can be manipulated to enhance production of a desired protein. Some or all of these factors could initially be assumed to play a role in affecting the level of protease production in our system.

Regulation of transcription. Prokaryotic promoters are simpler than their eukaryotic counterparts, and generally consist of two consensus sequences at -10 (Pribnow box) and -35 relative to the start of transcription. These sequences interact with the bacterial polymerase complex, and consequently their exact sequence regulates the affinity of the polymerase for the promoter and the ability of the bound polymerase complex to initiate transcription. The identification of high-affinity promoters has been a major step in developing heterologous expression systems.

The *trp* and *lac* promoters are endogenous *E. coli* promoters which are considered "strong" in that they are capable of directing the synthesis of a gene product to levels between 10 and 40% of total cellular protein. The primary difference between the two lies in the ability of the *lac* promoter to be tightly

repressed by virtue of a nucleotide sequence located downstream of the transcription start site. This sequence, the *lac* operator, functions by binding the *lac* repressor, coded by the *lacI* gene, in the absence of lactose and blocking the polymerase complex (Straney and Crothers, 1987). Repression can be relieved either by the addition of lactose or other synthetic chemical inducers. Although the *trp* promoter also contains repressor sequences, its activation is mediated by tryptophan starvation which is slower and not as readily inducible as the *lac* promoter. The high inducibility of the *lac* promoter has been conferred upon hybrid promoters by combining the -10 and operator sequences of *lac* with the -35 regions of other promoters. The *tac* promoter is one such hybrid which contains the -35 region of *trp*, and has become widely used in heterologous expression (deBoer, et al., 1983).

Host background may also play a significant role in affecting promoter function. Bacterial strains such as JM-101 and DH-1 (Hanahan, 1983; Yanisch-Perron, et al., 1985) have been developed which contain an overexpressor mutation of the *lacI* gene, known as $lacI^q$. The presence of this gene provides increased repression of promoters containing the *lac* operator, and additional copies may be added to plasmids to achieve higher levels of repression.

Post-transcriptional regulation of messenger RNA levels. One of the simplest means of regulating translation is by affecting the population of a messenger RNA (mRNA) in the cell. This seems to be accomplished primarily by affecting the half-life of the transcript in question. The half-life of an average mRNA in *E. coli* is on the order of 2-6 minutes but genes have been identified whose transcripts have much longer half-lives (~15-20 min). Some messages exhibit cell growth control; having long half-lives under conditions where

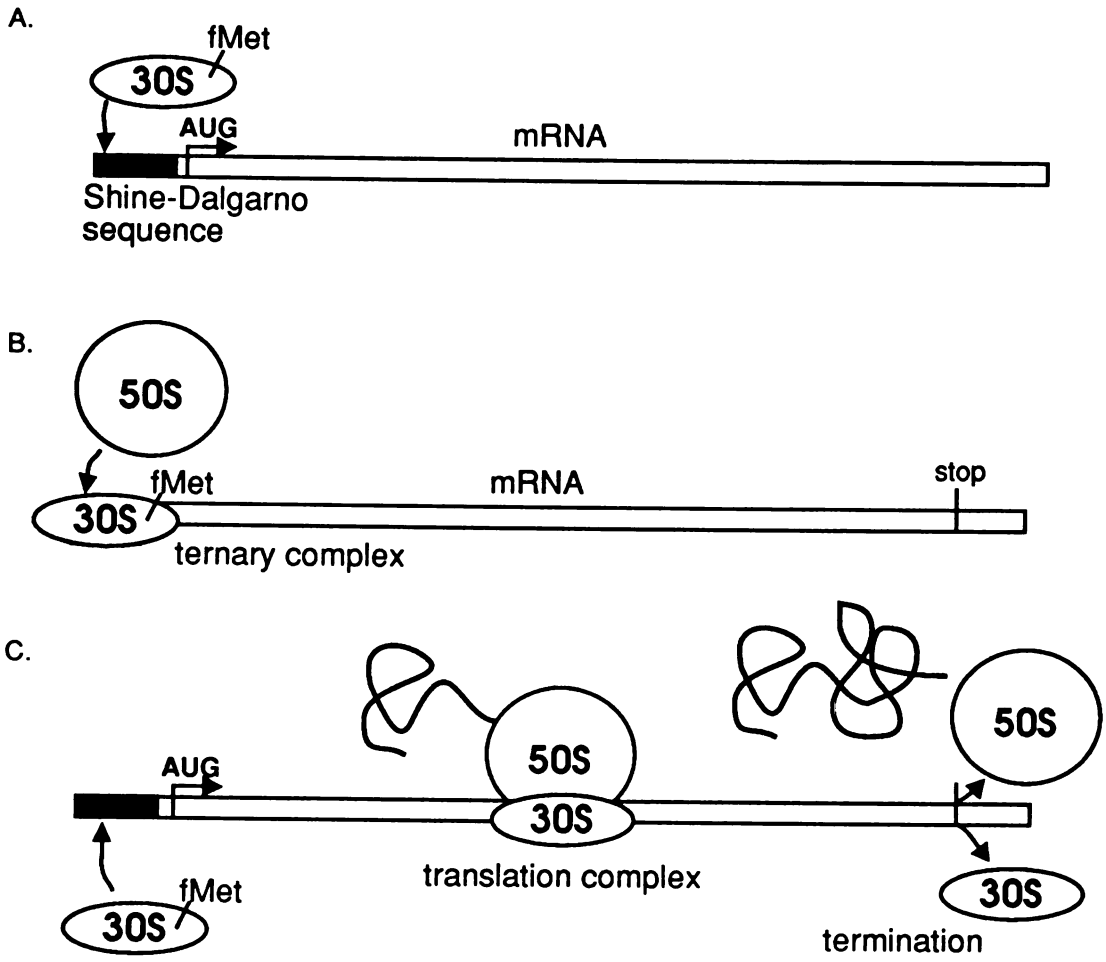
growth is rapid and short half-lives when growth is slow (Nilsson, et al., 1984). Messenger RNA degradation occurs by exo- and endonucleolytic attack, which occurs at structured regions in the 5' or 3' ends, as well as at internal sites in the transcript. Degradation promoting sites detected at the 5' and 3' ends of the transcript can be disrupted by simple site-directed mutants. In addition, Kennell and Riezman have demonstrated that the rate of message degradation of the three cistrons of the *lac* operon is inversely related to the frequency of translation (Kennell and Riezman, 1977). From this, it seems that the best way to stabilize mRNAs which contain internal endonuclease cleavage sites is by an indirect method such as the maximization of ribosome binding and translation.

Regulation of translation. Translation is initiated by a series of binding events which assemble the ribosomal subunits on a messenger RNA. The mRNA initially binds to the 30S ribosomal subunit in the presence of a formyl-Methionine charged transfer RNA (f-Met-tRNA) to form the 30S ternary complex. (see Figure 1-1) This complex then associates with the 50S ribosomal subunit to form the 70S initiation complex. Following the loss of various initiation factors the 70S complex is competent to carry out the initiation and elongation steps of polypeptide synthesis.

Initial association of the mRNA with the 30S subunit is proposed to occur by means of a nucleotide sequence (5'-GAUCACCUCCUUA-OH-3') on the 3' end of the 16S ribosomal RNA (Shine and Dalgarno, 1974), a structural RNA found in the prokaryotic ribosome. This sequence has a complementary consensus sequence on the mRNA known as the Shine-Dalgarno (SD) sequence, which makes up part of the ribosome binding site (RBS). The SD sequence is located at

Figure 1-1: Diagram of the prokaryotic translation process.

A. The mRNA associates with the 30S ribosomal subunit and f-Met tRNA complex by virtue of interactions between the 16S-rRNA and Shine-Dalgarno sequence. B. The 50S ribosomal subunit associates with the ternary complex to form a competent translation complex. C. Peptide synthesis by the translation complex is terminated at the stop codon and the polypeptide chain is released. The translation complex dissociates following termination.


A.
fMet
30S
AUG
mRNA
Shine-Dalgarno
sequence
B.
50S
fMet
mRNA
stop
30S
ternary complex
C.
AUG
50S
50S
30S
fMet
30S
translation complex
30S
termination

approximately -10 relative to the start AUG and typically consists of 3-5 contiguous bases from the consensus sequence AGGAGGU.

a. Regulation of Ribosome Binding / pre-initiation complexes. Genes cloned from organisms other than *E. coli* will most likely not contain optimal SD sequences. Ribosome binding site optimization should take into account the observation that even well-expressed *E. coli* genes do not contain long, highly complementary SD sequences. Ribosome binding in *E. coli* is optimized for both capture of pre-initiation complexes and release of translation-competent ribosomes. The current perception in the literature is that very long SD sequences hinder ribosome release and therefore decrease the frequency of polypeptide chain initiation.

Additional sequences around the ribosome binding site appear to be important regulators of expression. A systematic analysis of the nucleotide spacing between the SD sequence and the initiating AUG (Shepard, et al., 1982) indicates that the optimal spacing is 9 ± 2 nucleotides. Deviation from this spacing can affect expression as much as 200-500 fold. This might be expected, given the constrained structural environment of the ribosome-mRNA initiation complex. As discussed by Stormo, the sequence of the spacer region appears to be biased heavily in favor of A and U, perhaps in order to reduce self-complementarity and unambiguously define the SD sequence (Stormo, et al., 1982).

Ribosome binding to the SD sequence may also be affected by the formation of local secondary structures which occlude the site. A number of well-studied mRNAs contain predicted stem-loops which encompass the SD and/or initiator AUG and are, in general, poorly translated. Mutations which reduce the self-complementarity in the stem-loop can increase the translation of these genes by as much as 20-100 fold (Hall, et al., 1982). Consequently, genes cloned into

expression vectors should be examined for possible secondary structures which may be formed by nucleotide sequences around the SD site and start codon which could inhibit ribosome binding.

b. Regulation of polypeptide elongation. Once a ribosome initiates a polypeptide chain there are a number of factors which appear to affect its ability to efficiently complete the elongation process. The first of these is the nucleotide bias in codons immediately downstream of the initiator AUG. Taniguchi originally demonstrated that mutation of the second codon from GCA to ACA had a dramatic effect on translation (Taniguchi and Weissmann, 1978), suggesting that the codons immediately downstream of the start codon could affect translation efficiency. In keeping with this suggestion, it has been observed that G:C rich genes, such as those cloned from parasites, are poorly translated in *E. coli* (Button, et al., 1991). Maximizing the first four codons after the AUG for A:T usage without affecting the amino acid sequence resulted in a 20-30 fold increase in expression. Although the presence of G:C rich codons may promote the formation of secondary structures which could block elongating ribosomes, no general correlation has been made between the effect of A:T substitutions in the 5' region on protein expression and disruption of predicted stem loops.

c. Regulation during polypeptide elongation. The choice of codons which are most heavily used for a given amino acid differs between prokaryotes and eukaryotes, an observation which is thought to be tied to the relative levels of aminoacyl-tRNAs inside the host organism. Codons which are used infrequently by an organism are referred to as "rare". A good deal of information has been published with respect to the role of rare codons in affecting translational efficiency. In many cases, changing codon usage of a foreign gene to match that of the expression host leads to an improvement in expression levels of at least 2-

20 fold (Kotula and Curtis, 1991; Williams, et al., 1988). Certain regulatory genes of *E. coli* have provided evidence that rare codons are used to modulate expression levels (Konigsberg and Godson, 1983). In many other cases, however, heterologous genes with high percentages of rare codons (15-25%) are able to be expressed to very high levels (Devlin, et al., 1988; Nassal, et al., 1987). It is likely that rare codons modulate gene expression by affecting the rate at which proteins are synthesized, rather than the final levels which may accumulate ina cell. This represents a level of regulation which is subordinate to those mentioned previously.

Environmental factors.

Although much is known about the specific mechanisms by which protein expression can be manipulated at the transcriptional and translational level, there are other factors which may significantly influence the ability to recover recombinant proteins. Heat shock proteases in *E. coli* can be stimulated by expression of recombinant proteins, reducing overall expression levels. Strains in which these proteases have been deleted (known as *lon*$^-$) have been constructed and found to be more tolerant to overexpression (Goff and Goldberg, 1985). Variations in the temperature at which bacteria are grown may not only affect expression level, but also the partitioning of a recombinant protein between the soluble and insolube fraction of the cell (Browner, et al., 1991).

It has been observed that the same plasmid may vary widely in its ability to direct recombinant protein production when tested in several different strains of E. coli. This variation in host background is often unanticipated, and in most cases is not linked to a known genetic feature. For this reason, it is often a good

strategy to test several strains of *E. coli* for expression when initially attempting to express a protein (Browner, et al., 1991; Frorath, et al., 1992)

Specific considerations for HIV and SIV proteases.

Expression of the HIV-1 protease has generally been pursued by means of expressing a precursor polypeptide from which the protease is liberated by autoproteolytic processing (Debouck, et al., 1987; Giam and Boros, 1988; Graves, et al., 1988). In the Craik lab, HIV-1 protease expression has been achieved by fusion of human superoxide dismutase (hSOD) to a fragment of the HIV-1 *pol* gene (Babé, et al., 1990; Pichuantes, et al., 1989). Gene transcription was directed by the *tac* promoter, which permitted tight regulation in the presence of the *lac* repressor. Upon induction by the chemical inducer isopropyl-thio-β-d-galacto-pyranoside (IPTG) the fusion protein was produced within 5 to 10 minutes. Mature protease was produced from this fusion by autoproteolytic processing and reached a peak level in two to four hours.

Given this past history and the considerations for expression in *E. coli* described previously, we were faced with three specific requirements for optimizing expression. First, it was important that the protease gene be highly expressed to facilitate purification. Second, that expression be both tightly controlled to prevent leakage of the protease, which is toxic to *E. coli* (Baum, et al., 1990; Darke, et al., 1989; Korant and Rizzo, 1991) in the absence of induction, and rapidly expressed following induction. Third, we wanted to be able to design an expression system which would not rely upon the protease activity itself for processing to the mature 99-amino acid protein, thus allowing us to make variant proteases with altered activities or specificities.

Materials and Methods

Strains, plasmid constructions, and oligonucleotides. Restriction enzymes, T4 DNA ligase, T4 DNA polymerase, and T4 polynucleotide kinase were purchased from New England Biolabs Inc. and used according to the conditions recommended by the supplier. DNA fragments for ligations were purified after separation in 1% agarose gels using the Geneclean® Kit (Bio 101 Inc., La Jolla, CA).

The *E. coli* strains used for cloning and mutagenesis were JM-101 (*sup*E, *thi*-1, Δ(*lac-pro*AB), [F' *tra*D36, *pro*AB, *lac* $I^qZ\Delta M15$] and LE112 (F' *lac* I^q, *lacZ*::Tn5, *proAB/dut1*, *ung1*, *relA1*) (15). Vectors containing the *tac* promoter were constructed and tested for HIV protease activity in *E. coli* X90 (F' *lac* I^q, *lac*ZY, *pro*AB/Δ(*lac-pro*), *ara*, *nal*A, *arg*E(am), *thi*, *rif*^{r}).

The gene for the HIV-1 protease (HIV-99Y) was obtained as a fusion between hSOD and a synthetic gene encoding the 99-amino acid protease using codons biased for yeast expression (Pichuantes, et al., 1989). The HIV-2 protease gene was obtained from pSOD/HIV2PR113 (Pichuantes, et al., 1990) as an *Nco* I - *Sal* I fragment coding for a 113-amino acid pol precursor. The SIV protease gene (strain mac_{239}) was obtained from plasmid SOD/SIVPR (Rosé, et al., 1993) as an *Nco* I - *Sal* I fragment coding for the 99-amino acid protease with a Met-Ala extension on the amino terminus. The gene for CheY was obtained from the plasmid pCHEY15LOX (Sigal, et al., 1990).

Protease genes were subcloned into a Bluescript KS (-) plasmid from Stratagene (La Jolla, CA) to generate single-stranded templates for mutagenesis. Oligonucleotides used were synthesized on a 308B DNA synthesizer (Applied Biosystems, Inc., Foster City) and purified on Nensorb prep columns (NEN

Research Products, Boston). Site directed mutagenesis was carried out using the protocol of Kunkel (Kunkel, et al., 1987). Phosphorylated oligonucleotides were annealed to single stranded uracil-laden template in 20 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$ by boiling for 2 minutes and cooling to room temperature over 30 minutes. Complementary strands were synthesized and ligated by T4 DNA polymerase and T4 DNA ligase in the presence of 20 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 35 mM NaCl, 0.4 mM dNTPs, 1 mM ATP and 1.5 mM DTT. Following transformation, mutants were screened by restriction endonuclease digestion and dideoxy sequencing from double stranded templates using the Sequenase 2.0 enzyme and protocols (United States Biochemical Corp., Cleveland, OH).

Induction and Analysis of Expression. *E. coli* harboring expression plasmids were diluted ten-fold from saturated cultures into 3 mL of LB/AMP (60 μg/ml) and allowed to recover at 37°C for 1.5 hours before induction with 500 μM IPTG. Cells were pelleted by centrifugation at 2 hours post induction and sonicated in lysis buffer (100 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT) to generate crude lysates. Samples were centrifuged at 14,000 rpm in an Eppendorf 5415 C microfuge to clarify lysates. Volumes of clarified lysate corresponding to 0.045 OD_{600} for the HIV-1 protease or 0.1 OD_{600} for the SIV and HIV-2 proteases were analyzed for protease activity using a discontinuous HPLC assay described previously (Pichuantes, et al., 1990). Lysates were incubated in activity buffer (50 mM NaOAc, pH 5.5, 1 M NaCl, 1 mM EDTA, 1 mM DTT) at 37°C in the presence of 250 μM substrate peptide NH_2-ATLNFPISPW-COOH (Peptide I). Reaction products were separted by reverse-phase HPLC and detected at 280 nm. Percent substrate cleavage was determined by division of the area under the product peak by the total area under the product and substrate peaks.

SDS-PAGE and immunoblot analysis. Lysates of induced cells representing 0.2 OD_{600} of cells were resuspended in 40 µL of 1X Laemmli buffer and boiled for 5' prior to separation by SDS-PAGE (Laemmli, 1970). Proteins were separated on 15% polyacrylamide gels and either stained with Coomassie Brilliant Blue or electroblotted onto 0.2 µm nitrocellulose. Proteins were immunblotted using polyclonal rabbit antisera against the HIV-1, HIV-2, or SIV proteases. The secondary antibody used was a goat anti-rabbit antibody conjugated to horseradish peroxidase (Pierce, Rockford, IL). Blots were developed using 4-chloro-1-naphthol or 3,3'-diaminobenzidine in the presence of hydrogen peroxide.

Results

Fusion and native constructions for increasing HIV-1 protease expression. As an initial attempt to increase expression, the hSOD-HIV-1 PR gene fusion was cloned into the pTacTac expression vector to generate the plasmid pT2SOD HIV-1. The pTacTac vector (Appendix I) has been used for the high level expression of lysozyme and contains a unique promoter box, consisting of two serial *tac* promoters preceded by a *lac UV5* promoter. The presence of multiple promoter elements presumably allows for the attraction of large numbers of transcription complexes, leading to high level production of mRNA. Three copies of the *lac* operator in this plasmid ensure strong repression until transcription is induced. Although the vector contains a gene for the *lac* repressor (*lacI^q*), it is necessary to use strains of *E. coli* which also contain the *lacI^q* to compensate for the multiple copies of the *lac* repressor binding site.

Induction of the plasmid T2SOD HIV-1 led to the accumulation of large amounts of fusion precursor, but no mature, soluble HIV-1 protease (Figure 1-2).

The precursor and some intermediate processing forms were present in the insoluble fraction of the cell, suggesting that high-level expresion had led to aggregation during processing. When the gene for the 99 amino acid protease (HIV-99Y) alone was cloned into the vector no expression of the protease was observed by Western blot or activity assay. A synthetic protease gene containing optimized *E. coli* codons, as defined in (Reznikoff and Gold, 1986), was also tried (HIV DG - provided by AIDS Research and Reagent Program from Michael Hauser and Eun Hwang) but also did not direct the production of any HIV-1 protease.

The lack of protease expression was puzzling in light of the previous success with the fusion construct. Since the gene could be expressed by fusion to hSOD it was possible that a block to translation existed at the 5' end of the protease genes. By comparison to other studies on protein expression, it was predicted that the problem involved a secondary structure which blocked ribosome binding or translation initiation. RNA folding predictions generated by the computer program RNAfold (Biocomputational Laboratory, UCSF) from the sequences at the 5' end of the HIV-1 gene indicated a potential stem-loop structure with an energy of -12.3 kcal encompassing nucleotides +2 to +23 (Figure 1-3). Such a stem loop would prevent a ribosome from initiating translation, even though it could bind to the Shine-Dalgarno sequence (Baim, et al., 1985; Liebhaber, et al., 1992; Schoner, et al., 1984; Tomich, et al., 1989). Due to amino acid codon restrictions, it was not possible to alter the nucelotide sequence in such a way as to disrupt the predicted stem loop. Although the hSOD-fusion construction was useful for overcoming the translation initiation problems, it

Figure 1-2: Expression of HIV-1 protease in pTacTac.

E. coli harboring the fusion plasmids pSOD/PRT5 and pT2SOD HIV-1 and the bicistronic vector pT2b(4)HIV-1 were grown to approximately 1 OD_{600} and induced with 500 μM IPTG. After two hours, cells were harvested and fractionated by sonication into soluble (S) and pellet (P) fractions. Proteins were analyzed on 12.5 % SDS-PAGE and immunoblotted with anti-HIV-1 protease antibodies. Total lysate from noninduced cells (-) was run as a negative control.

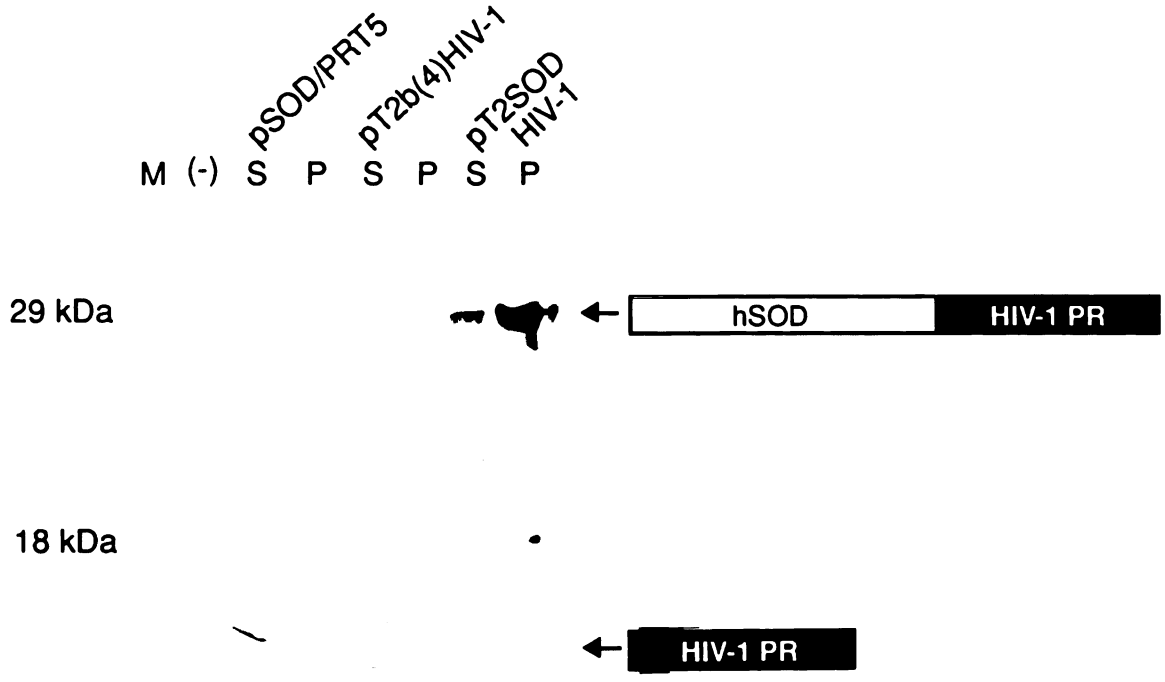
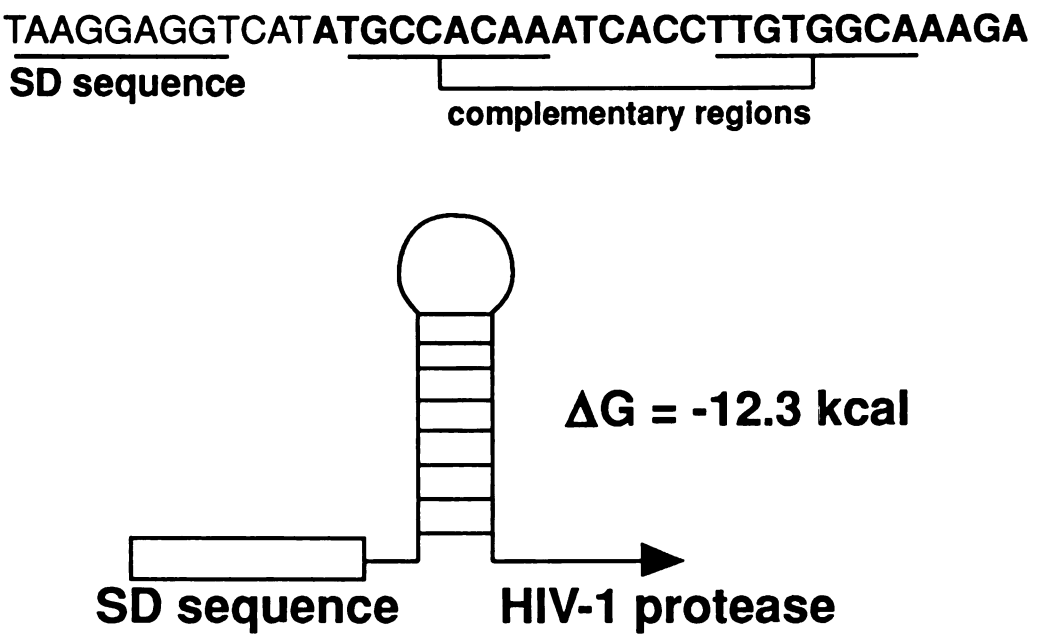

M (-) S P S P S P
pSOD/PRT5
pT2b(4)HIV-1
pT2SOD
HIV-1
29 kDa
hSOD
HIV-1 PR
18 kDa
HIV-1 PR

Figure 1-3: Diagram of predicted stem-loop in HIV-1 protease gene.

A stretch of 38 nucleotides, including the Shine-Dalgarno sequence and the first 9 codons of the HIV-1 protease gene (in bold) were analyzed for self-complementarity using the RNAfold program (Biocomputational Laboratory, UCSF). Complementary sequences in the HIV-1 protease gene were predicted to form a strong stem loop, ΔG = -12.3 kcal, which would encompass the start codon.


TAAGGAGGTCATATGCCACAAATCACCTTGTGGCAAAGA
SD sequence
complementary regions
ΔG = -12.3 kcal
SD sequence
HIV-1 protease

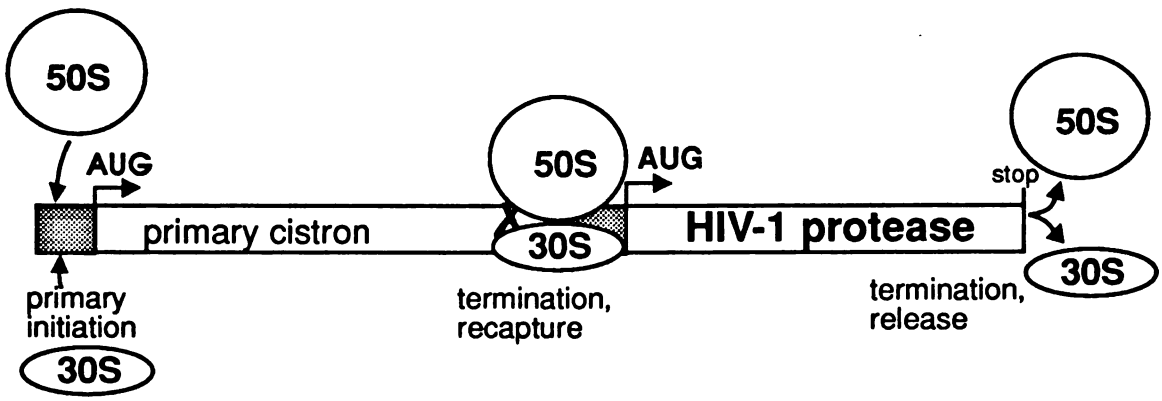
could not be expressed in a soluble form at high levels, and would complicate the expression of proteins which could not autocatalytically release themselves from the fusion protein. An alternative method had to be identified to prevent secondary structure formation around the start of the protease gene.

The theory behind bicistronic expression. A method has been described which uses a bacterial regulatory process to affect secondary structures at the 5' ends of mRNAs (Schoner, et al., 1986). In this method, genes are fused transcriptionally in a bicistronic arrangement, such that the gene of interest is downstream of a highly expressed protein: the primary cistron. This gene is chosen for efficient translation initiation, and contains a termination codon and second Shine-Dalgarno sequence (SD2) at its 3' end. The gene of interest is placed to take advantage of the SD2. As shown in Figure 1-4, ribosomes terminating from the primary cistron are recaptured by the SD2 and directed to translate the gene of interest. Since ribosomes cover approximately 30 nucleotides of an mRNA when paused (Steitz, 1979), it is thought that the ribosome can melt out inhibitory secondary structures in the 5' end of the second cistron, freeing it for translation (Berkhout, et al., 1987). In this way the protein of interest can be produced without making a true translational fusion which would require further processing. This system has been used by a number of investigators to overcome difficulties associated with poor intrinsic translation initiation (Guzman, et al., 1992; Makoff and Smallwood, 1990; Schoner, et al., 1986; Spanjaard, et al., 1989).

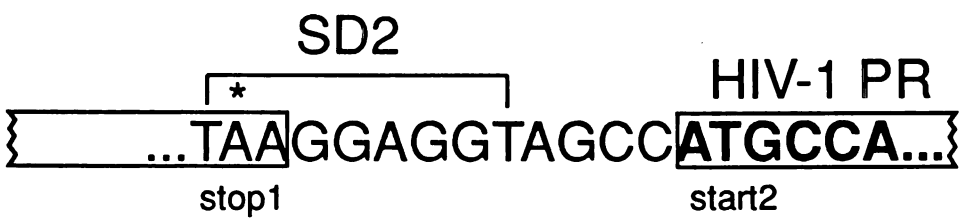
Figure 1-4. Function of a bicistronic expression system.

A. Initiation at the primary cistron is followed by translation and release of the polypeptide chain. Terminating ribosomes are recaptured by a second Shine-Dalgarno sequence at the site of termination, and directed to initiate translation of the HIV-1 protease gene. Termination of the HIV-1 protease results in dissociation and release of the translation complex. B. Sequence inserted into hSOD-HIV-1 PR fusion to generate a bicistronic expression plasmid. The second Shine-Dalgarno sequence (SD2) is complementary to 9 nucleotides at the 3' end of the 16S rRNA and contains the stop codon for the primary cistron.

A.


50S
AUG
primary cistron
primary
initiation
30S
50S
AUG
30S
termination,
recapture
50S
stop
HIV-1 protease
30S
termination,
release

B.


SD2
*
HIV-1 PR
...TAAGGAGGTAGCCATGCCA...
stop1
start2

Bicistronic constructions for HIV-1 protease expression. Using the hSOD-HIV-1 PR gene fusion as a starting point, the sequence in Figure 1-4B was inserted between the hSOD and HIV-1 protease sequences. The chosen ribosome binding site contained a 9 nucleotide SD2, which was a perfect complement to the end of the 16S rRNA, and a 4 nucleotide spacing between the SD2 and the ATG of the HIV-1 protease. This construction was cloned into the pTacTac vector, generating the bicistronic expression vector pT2b(4)HIV-1.

Cells harboring the bicistronic vector, the parent fusion vector pT2SOD HIV-1, and the original vector pSOD/PRT5 were induced and harvested at two hours post-induction. Analysis of the soluble and insoluble cellular fractions yielded the results shown in Figure 1-2. The bicistronic construction expressed levels of soluble, mature HIV-1 protease which were comparable to those obtained from pSOD/PRT5, while the pTacTac-based fusion vector produced only the 28kDa hSOD-HIV-1 PR fusion protein. By activity analysis of crude lysates of *E. coli*, the bicistronic construction expressed approximately 60% the amount of active protease as the pSOD/PRT5 vector.

Manipulation of intergenic spacing. Improvement of expression from the bicistronic construction was attempted by altering the sequences directing re-initiation. Based on the observation that SD-ATG spacing could significantly affect expression (Shepard, et al., 1982), a series of vectors, pT2b(n)HIV-1, was constructed in which the spacing (n) was varied from two to nine nucleotides. Efficiency of expression was analyzed by quantitating protease activity in induced bacterial lysates (Table 1-1). Detectable protease activity was maximal at a spacing of 3 nucleotides, and dropped off quickly below this number. Expression decreased slowly as the spacing increased from 3 nucleotides, finally

Table 1-1. Effect of spacer length variations on HIV-1 protease expression from hSOD-bicistronic constructions.

spacer	plasmid	rel. activity
none	pSOD/PRT5	1.0
AC	pT2b(2)HIV-1	0.11
ACC	pT2b(3)HIV-1	0.71
AGCC	pT2b(4)HIV-1	0.64
AGCAT	pT2b(5)HIV-1	0.58
AGCATAT	pT2b(7)HIV-1	0.40
AGATAACAT	pT2b(9)HIV-1	0

Activities determined by HPLC assay of clarified lysates are reported relative to the initial fusion construction.

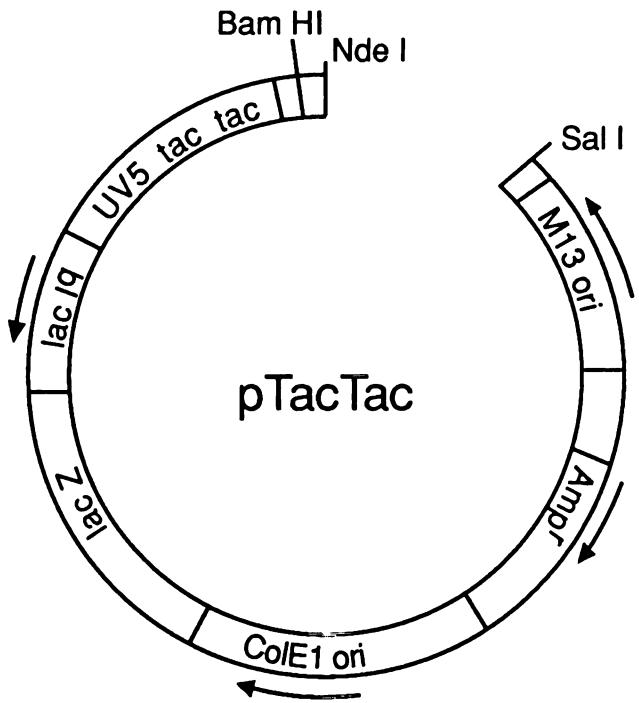
reaching undetectable levels at 9 nucleotides. In the best construction, pT2b(3)HIV-1, protease activity reached 70% of that observed from the pSOD/PRT5 vector.

Influence of the primary cistron on protease expression. A second strategy was adopted to maximize HIV-1 protease expression in the bicistronic arrangement (Figure 1-5). Since the level of expression of a downstream cistron is ultimately determined by the number of ribosomes which are recaptured (i.e. the translation level of the primary cistron), a more highly expressed protein was sought to replace hSOD. The CheY protein, a bacterial protein which functions as part of the chemotactic machinery (Matsumura, et al., 1984) had been used previously as a high-level fusion partner in our lab (Eakin, et al., 1992; Sigal, et al., 1990) and was readily available.

Using the polymerase chain reaction (PCR) a stop codon and Shine-Dalgarno sequence were inserted after the 91st amino acid of the CheY gene and the fragment was cloned into pT2b(4)HIV-1 to replace the hSOD gene, generating pT2Cb(4)HIV-1. A series of spacer mutants was constructed in which the spacing was changed from 2 to 5 nucleotides in order to assess the sensitivity of this construction to SD-ATG distance relative to the hSOD-bicistronic constructions. The results obtained with the pT2Cb(n)HIV-1 series of plasmids are shown in Table 1-2. The optimal spacing was 3 nucleotides, at which point expression was nearly twice that of the comparable pT2b(3)HIV-1 vector, and 40% higher than that of the pSOD/PRT5 vector. Expression dropped off as SD-ATG spacer distances varied, in a similar manner to the hSOD-bicistronic constructions, although overall expression remained higher. It was apparent by Western blot that some of the protease expressed from the CheY-bicistronic

Figure 1-5: Summary of vectors used to develop bicistronic expression system.

Nde I - *Sal* I cassettes used for probing efficiency of bicistronic plasmids were cloned into the pTacTac vector as shown. Black bars indicate the position of the ribosome recapture sequences. A. Bicistronic constructions based on human superoxide dismutase (hSOD), in which the spacing between the second Shine-Dalgarno sequence (SD2) and start ATG (n) varied from 2-9 nucleotides (as detailed in Table 1-1). B. CheY-based bicistronic constructions in which SD2-ATG spacing (n) varied from 2-5 nucleotides (as detailed in Table 1-2). C. Bicistronic construction based on a synthetic gene fragment cloned between the *Nde* I and *Pst* I sites of the CheY constructions. SD2-ATG spacing was held at 3 nucleotides.


Bam HI
Nde I
Sal I
UV5 tac tac
lac Iq
M13 ori
pTacTac
lac Z
Amp r
ColE1 ori

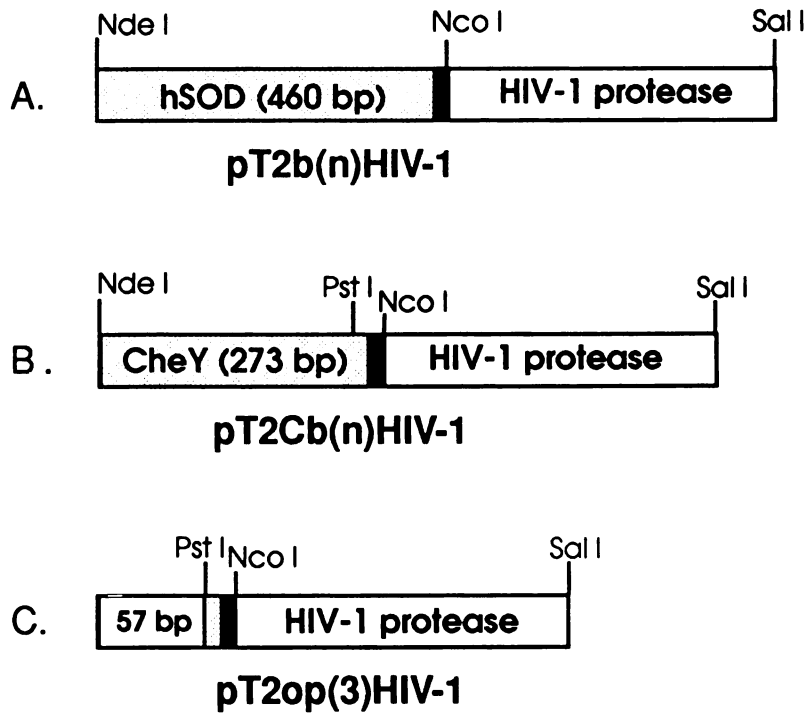

A.
Nde I
Nco I
Sal I
hSOD (460 bp)
HIV-1 protease
pT2b(n)HIV-1
B.
Nde I
Pst I
Nco I
Sal I
CheY (273 bp)
HIV-1 protease
pT2Cb(n)HIV-1
C.
Pst I
Nco I
Sal I
57 bp
HIV-1 protease
pT2op(3)HIV-1

Table 1-2. Effect of spacer length variations on HIV-1 protease expression from CheY-bicistronic constructions.

spacer	plasmid	rel. activity
none	pSOD/PRT5	1.0
AC	pT2Cb(2)HIV-1	0.21
ACC	pT2Cb(3)HIV-1	1.40
AGCC	pT2Cb(4)HIV-1	0.96
AGCAT	pT2Cb(5)HIV-1	0.71

Activities determined by HPLC assay of clarified lysates are reported relative to the initial fusion construction.

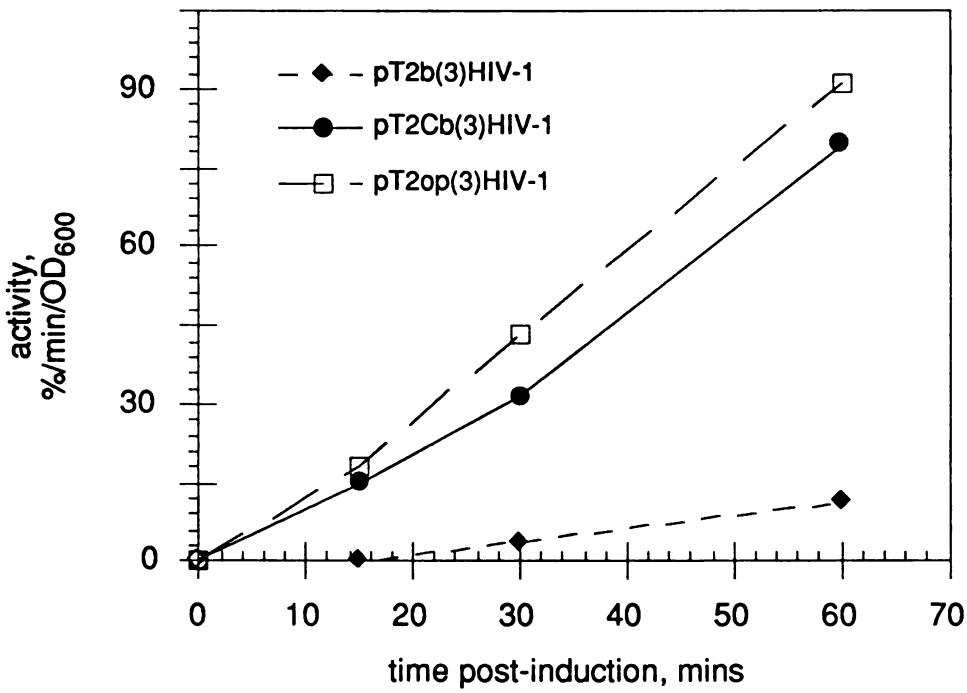
vectors had precipitated. Thus the activity measurement gives an indication of the amount of protease that remained soluble, which is an underestimation of the total expression level. The insoluble fraction also contained an immunoreactive band which corresponded in size to a CheY-HIV-1 PR fusion protein. This may have been due to some low-level read-through of the stop codon in the second ribosome binding site, leading to the synthesis of an insoluble CheY-HIV-1 PR fusion. Athough this effect could be corrected through the addition of more efficient stop codons, or stop codons in other reading frames, it represented only a small amount of the expressed protein and we chose not to pursue it.

Changing the identity of the primary cistron appeared to have an effect on the rate at which protease activity accumulated in induced cells (Figure 1-6). Western blots of induced pT2b(3)HIV-1 lysates with anti-hSOD and anti-HIV-1 PR antibodies, showed that although hSOD was present in the cell at 15 minutes, no protease was present (data not shown). Thus, the fifteen minute lag observed in the appearance of HIV-1 protease in the hSOD bicistronic construction was likely due to a slow delivery of ribosomes to the re-initiation site.

A final change was made in the primary cistron in an attempt to improve expression levels. A synthetic 57 base oligonucleotide was designed which obeyed the rules for high level expression mentioned previously: 1) optimal *E. coli* codon usage, 2) high A:T usage, and 3) absence of predicted mRNA secondary structure. This synthetic operon was cloned into the pT2b(3)HIV-1 vector in place of hSOD, generating the plasmid pT2op(3)HIV-1 (Figure 1-5). Upon induction in small scale cultures this plasmid was capable of directing the expression of approximately 20% more protease than the pT2Cb(3)HIV-1 construction (data not shown). Further analysis, however, indicated that bacteria harboring the plasmid were poorly viable and rapidly lost the ability to express

Figure 1-6: Effect of primary cistrons on the rate of protease expression.

E. coli harboring the bicistronic vectors pT2b(3)HIV-1, pT2Cb(3)HIV-1, and pT2op(3)HIV-1 were grown to approximately 1 OD600 and induced with 500 μM IPTG. At 0, 15, 30, and 60 minutes post-induction samples were taken and crude lysate generated by sonication. Protease activity in crude lysates was measured as described in Materials and Methods. Data was plotted as % substrate cleaved per minute per OD_{600} to normalize activity to cell density.


pT2b(3)HIV-1
pT2Cb(3)HIV-1
pT2op(3)HIV-1
activity, %/min/OD600
90
60
30
0
0
10
20
30
40
50
60
70
time post-induction, mins

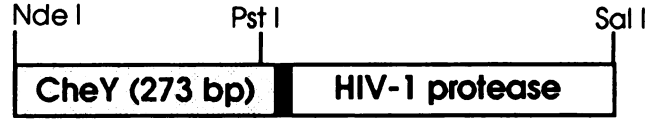
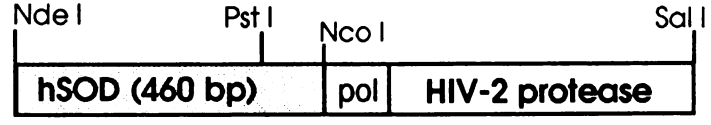
the protease. This was presumably due to low level expression of the protease in the absence of induction, which selected for non-expressing bacteria. Because of the instability of this construction, plasmids containing the synthetic operon were abandoned.

Bicistronic expression of SIV protease. Having developed an effective system for HIV-1 protease expression, we turned to test its applicability to expressing other viral proteases. A summary of the constructions obtained during the development of SIV and HIV-2 expression vectors is presented in Table 1-3.

The protease from SIVmac239 was expressed efficiently from a construct pSODSIVPR built by Sergio Pichuantes, in which the SIV protease gene replaced the HIV-1 protease gene in pSOD/PRT5. By measuring activity, protease concentrations in induced cells were calculated to be approximately 5 mg/L/OD_{600}. To examine the potential for bicistronic expression, the HIV protease genes in pT2Cb(3)HIV-1 and pT2op(3)HIV-1 were replaced by the SIV protease gene. By virtue of the cloning sites used to transfer the SIV gene, the protease cistron encoded the protease with a Met-Ala extension on the amino terminus. Expression of SIV protease activity from pT2Cb(3)SIV and pT2op(3)SIV was compared to that from pSODSIVPR. Activity levels were considerably lower than those observed in the fusion system, being decreased nearly two- and four-fold in the synthetic operon and CheY bicistronic constructions, respectively.

The poor performance of the bicistronic system expressing the SIV protease had not been anticipated. A possible explanation for the effect was that the SIV protease gene itself contained sequences which hindered expression from the

Table 1-3. Summary of current expression vectors for retroviral protease

	purpose	N-terminus	construct name, gene organization
A.	expresses active and inactive HIV-1 protease mutants	M-PQITL...	 **pT2Cb(3)S/HIV-1**
B.	expresses active and inactive SIV protease mutants	M-PQFS...	 **pT2op(3)S/SIV**
C.	expresses active HIV-2 protease. Requires autoprocessing	(pol) ...GLAA - PQFS...	 **pT2HIV2PR115**
D.	expresses active or inactive HIV-2 protease mutants	M-PQFS...	 **pT2Cb(3)S/H2**

bicistronic system. It was observed upon closer analysis that the 5' end of the protease gene contained a large number of G and C nucleotides immediately downstream of the start codon, a factor which has been observed to inhibit translation initiation (Devlin, et al., 1988; Tomich, et al., 1989). In our case, this effect did not appear to correlate with predicted stem-loop structures, suggesting that nucleotide usage itself played a role in limiting translation initiation.

Analysis of nucleotide usage effects at the 5' end of the SIV protease gene. In order to express the protease with only a methionine extension on the amino terminus, and to test the effect of 5' nucleotide sequences on expression, a mutant (S/SIV) was constructed[†]. In this construct, the N-terminal codons were altered to increase A:T usage from 9 to 11 out of 18 positions. A second mutant, in which Ser4 was changed to His (S/SIV S4H), also contained 11/18 A or T nucleotides although they were positioned in a different arrangement. Further mutations to increase A:T usage were not possible due to codon usage restrictions. The sequences of these genes and corresponding protease activity levels are shown in Table 1-4. The increase in A:T usage resulted in a two to three-fold increase in expression of the SIV protease in the context of the CheY-bicistronic arrangement and three to four-fold increase in the synthetic operon constructions. Unlike the HIV-1 protease, the synthetic operon construction was stable and was maintained for the expression of the SIV protease. With all constructions, the expressed protease partitioned only into the soluble fraction of induced cells,

---

All S/ constructs were built by cutting the original bicistronic vector with Nco I and blunting with T4 ligase to generate a blunt ended ATG, which served as the start codon. The A:T optimized protease genes were cut with Stu I at the 5' end to yield a blunt end which started with CCT, the codon encoding Pro1. The ligation of these two fragments yielded protease genes in bicistronic constructs with only Met amino-terminal extensions.

Table 1-4. Effect of 5' nucleotide sequences on SIV protease expression

clone	sequence of 5' end (aa 1-7)	% clvg[a]
pT2Cb(3)SIV	ATGGCCCCTCAATTCTCTCTT	3.1 ± 0.5
pT2Cb(3)S/SIV	ATGCCTCAATTCTCTCTTTGG	7.6 ± 0.2
pT2Cb(3)S/SIV S4H	ATGCCTCAATTCCATCTTTGG	10.5 ± 1
pT2op(3)SIV	ATGGCCCCTCAATTCTCTCTT	4.7 ± 0.7
pT2op(3)S/SIV	ATGCCTCAATTCTCTCTTTGG	14.4 ± 0.6
pT2op(3)S/SIV S4H	ATGCCTCAATTCCATCTTTGG	16.1 ± 0.2

[a] % clvg measured by incubating 0.1 OD600 equivalents of clarified lysate with 250 μM substrate for 15 minutes at 37°C in activity buffer.

suggesting that activity assays were an accurate measure of total expression level.

Expression of the HIV-2 protease. Expression of the HIV-2 protease was initially attempted in the pTacTac vector using a fusion between hSOD and a precursor of the HIV-2 protease which included 13 amino acids of the pol gene upstream of the HIV-2 protease (Pichuantes, et al., 1990). The fusion product was well expressed without becoming insoluble, and mature HIV-2 protease was released by autoproteolytic processing. Levels of mature protease were calculated to be approximately 1 mg/L/OD_{600} (Rafael Salto, personal communication).

The pol-precursor form of the HIV-2 protease (115 amino acids in length) and a mutant in which the precursor region had been deleted (S/H2) were cloned into the pT2Cb(3)HIV-1 construction to generate the plasmids pT2Cb(3)H2/115 and pT2Cb(3)S/H2. These plasmids expressed active HIV-2 protease at levels approaching those seen for the fusion construction. A second set of plasmids based on the synthetic operon were also constructed (see Table 1-5). The pattern of protease expression was very similar to what had been observed for SIV, in that the genes containing a less G:C rich amino terminus were more highly expressed. The use of the synthetic operon increased expression of the 115-amino acid precursor form of the HIV-2 protease, but all attempts to isolate clones expressing the mature, 99-amino acid form were unsuccessful because of toxicity problems. The construction pT2Cb(3)S/H2 was retained for expression of inactive mutants (Table 1-3).

Table 1-5. HIV-2 protease expression from fusion and bicistronic constructions

clone	relative activity
pT2HIV2	1.0
pT2Cb(3)H2/115	0.78
pT2Cb(3)S/H2	0.92
pT2op(3)H2/115	1.54
pT2op(3)S/H2	unstable

Discussion

The expression of the HIV-1 protease proceeded by two major steps. The first was the use of the hSOD-HIV-1 protease gene fusion to assess the strength of the pTacTac promoter relative to the promoter in the parent pSOD/PRT5 vector. Although results were positive in that larger amounts of precursor protein were expressed from the pTacTac promoter, overexpression led to precipitation of the precursor protein and thus prevented the liberation of native HIV-1 protease. Attempts to express the HIV-1 protease gene as a single entity, rather than a gene fusion, were unsuccessful even in the context of the stronger promoter system.

Our intention to study the effects of site-directed mutants on the activity and specificity of the HIV-1 protease required that we be able to produce the enzyme in the absence of autoproteolytic processing. The second step of the expression involved determining a means of expressing the HIV-1 protease free of fusion partners, taking into account the initial failure of attempts to express the gene by itself. Among the known factors which might be influencing prokaryotic translation we felt that rare codon usage could be ruled out, since both protease genes contained codons optimized for *E. coli*, or yeast, which are efficiently utilized by *E. coli*. Furthermore, the presence of a sequence which rendered the mRNA unstable was unlikely, as messages containing these genes in the context of a fusion construction were efficiently expressed. We assumed that some block to translation was present in the HIV-1 protease gene when it was placed by itself behind a strong promoter; placing the protease gene in the context of a fusion to hSOD relieved this block.

In order to utilize the apparent ability of a gene fusion to overcome translation-inhibiting factors, we turned to the development of a bicistronic expression system. Bi- or multi-cistronic expression is widely used in bacteria,

and involves translating several independent polypeptides from a single mRNA. Synthesis of discontinuous polypeptides is achieved by an arrangement in which termination of one polypeptide chain leaves the ribosome at the initiation site of the second polypeptide. Presumably, the presence of the ribosome at the second translation-start site facilitates binding and initiation of translation by melting out inhibitory secondary structures in the RNA at the ribosome binding site.

Two factors were studied and found to be crucial in modulating protease expression: the identity of the primary cistron and the spacing between the second Shine-Dalgarno (SD) sequence and the protease gene. The ribosome termination and recapture sequences were derived from the sequences used by Schoner, et al. (Schoner, et al., 1986). We decided to use a maximally complementary SD sequence (9 nucleotides) to promote efficiency of ribosome recapture. In fact, truncation of this sequence by 2 nucleotides on the 5' or 3' end dramatically decreased translation of the protease gene (data not shown).

The distance between the SD and the start codon is thought to be involved in fine-tuning the positioning of the AUG at the "A" site of an initiating ribosome. A spacing of 3 nucleotides was optimal for expression of the protease, although constructs with spacings of up to 7 nucleotides still generated some protease activity. When the spacing was less than 3 nucleotides, however, expression decreased drastically. This suggests that flexibility exists in the ribosomal structur or the mRNA which will accomodate longer than optimal spacings but not shorter ones.

Altering the primary cistron had as much as a two to three-fold effect on protein expression levels. The kinetics of protease expression were dramatically altered by changing the primary cistron. Whereas protease activity could not be detected in hSOD-based bicistronic constructions until nearly 15 minutes after

induction, CheY based constructions produced protease almost immediately. Our interpretation of this result is that the CheY gene is more efficiently translated and was able to deliver ribosomes to the secondary cistron with greater speed. Although an optimized synthetic cistron appeared to improve upon this delivery rate and increased expression yet again, the constructs were unstable and cells harboring them appeared to suffer from the effects of protease toxicity when expanded to large scale cultures. This was probably due to a low level of leakage of *lac* repression, which allowed the production of a small number of mRNA transcripts from which protease was efficiently synthesized.

This observation highlighted an important issue when dealing with the pTacTac vector. The toxicity of the expressed gene products played a major role in isolation of clones, requiring that a high *lacI* background strain of *E. coli* be used for all cloning manipulations. Upon induction, protease expression increased until approximately 2 hours after induction. At this point the measurable levels of protease activity were at their peak, and beyond this point activity declined steadily. The growth rate of cells harboring active protease constructs declined (as measured by OD_{600}) and generally began to fall between 1 and 2 hours post induction relative to cells which had not been induced. Induced cells were harvested at 2 hours post induction to maximize protease accumulation and minimize cell death. Constructs which harbored inactive protease mutations (HIV-1 PR D25N) were highly stable and induced cells exhibited no decline in growth rates.

The ability of a bicistronic system optimized for the HIV-1 protease to direct the expression of related proteases was examined. The genes for the HIV-2 and a macacque strain of the SIV protease were both expressed from this system, but at somewhat lower levels than the parent fusion constructions. The increase in

protease expression with variation of the primary cistron was similar to that observed for the HIV-1 protease. By manipulation of the 5' sequences in the HIV-2 and SIV protease genes, it was possible to demonstrate that the identity of the nucleotides at the 5' ends of the genes was another important determinant of how well the proteases were expressed. Increasing A and T nucleotide usage increased expression as much as three to four-fold. Although this manipulation was limited by codon usage restrictions, it suggests that this is an important factor to be considered when expressing other genes in this system.

It was clear from this analysis that the bicistronic arrangement was a useful and efficient method for producing retroviral proteases. Expression levels were mostly limited by the toxicity of the protease, which rendered some constructs unstable and impaired growth of *E. coli* following induction. Each protease gene cloned into this system had slightly differing characteristics and required some systematic fine-tuning to optimize expression. The method developed here minimized both insolubility problems and post-translational modification requirements by expressing only the mature forms of the proteases. The N-terminal methionines were removed efficiently by endogenous methionine amino peptidase activity (Ben-Bassat, et al., 1987). These features rendered the system ideal for the expression of mutants with altered activities or substrate specificities.

Chapter 2. Engineering Retroviral Proteases for Increased Stability

Part of this material has been published. See Rosé, J.R., Salto, R., and Craik C.S. 1993. Regulation of Autoproteolysis of the HIV-1 and HIV-2 Proteases with Engineered Amino Acid Substitutions. J. Biol. Chem. **268:** 11939 - 11945.

Abstract

Autoproteolysis of the retroviral aspartyl proteases is a major obstacle to purification and analysis of these enzymes. A mutagenic approach to rendering autolytic cleavage sites less labile was applied to the primary cleavage site between Leu5 and Trp6 in HIV-1 PR. From predictions based on known substrates it was concluded that amino acids Lys or Ser in place of Gln at position 7 would prevent cleavage at the Leu5-Trp6 peptide bond, therefore stabilizing the protein. Autoproteolytic stability was enhanced at least 100-fold by these mutations. At longer time points the protease was degraded at secondary sites which contained adequate substrate sequences but were conformationally restricted. Conversely, a mutation in HIV-2 PR which changed Lys7 to Gln rendered the protein 3-fold less stable and shifted the position of the initial autoproteolytic cleavage from Phe3-Ser4 to Leu5-Trp6. The HIV-2 PR contained too many additional cleavage sites to allow stabilization with a small number of amino acid substitutions. The SIV protease, however, could be stabilized by approximately 4-fold through the replacement of Ser4 with His in the amino terminus. The effects of these mutations demonstrate that small changes in protein sequence can have a major impact on their autoproteolytic stability. The work described here suggests a general method for stabilizing proteases, and perhaps other recombinantly produced proteins, to autolysis.

General issues in proteolysis

Uncontrolled proteolysis is a problem frequently encountered when purifying proteins. In most cases, however, specific protease inhibitors or other compounds may be employed to prevent unwanted proteolysis during purification (North, 1989). When purifying and storing proteases, however, these strategies may complicate subsequent analysis of the enzyme. Proteases are often more susceptible to autoproteolysis when purified, being free of contaminating polypeptides that can serve as alternate substrates. In such a case, a protein may be stored at a sub-optimal pH to reduce proteolytic activity in a reversible fashion as can be done with the digestive enzymes trypsin, chymotrypsin, and elastase. Alternatively, a recombinant approach can be used to regulate or to stabilize the proteolytic activity. A metal activated switch has been engineered into trypsin to permit control of proteolysis by varying the concentration of copper in solution (Higaki, et al., 1990). Conformationally flexible loops in subtilisin have been immobilized and stabilized to autolysis through the creation of calcium binding sites in the loops (Braxton and Wells, 1992).

Autoproteolysis in the HIV and SIV proteases.

It was noted as early as 1989, that the retroviral aspartyl proteases were prone to activity loss during purification and storage (Strickler, et al., 1989), and a major contributing factor to the loss of activity was autoproteolytic degradation. Loss of activity and increased sample heterogeneity due to autoproteolysis can be a significant problem when performing kinetic and crystallographic analysis. The amino terminus of the HIV-1 protease appears to contain an efficiently cleaved sequence: PQITLWQRP, in which the Leu5-Trp6 bond is the bond cleaved.

Cleavage at this site results in the formation of a truncated species (HIV-1 PR Δ1-5) which has no intrinsic activity (Babé, et al., 1991). The crystal structure of HIV-1 PR reveals that this site occurs on an exposed surface loop following the first β-strand of the dimer interface. We observed that the HIV-2 and SIV proteases were more stable to autolytic degradation than HIV-1 protease and were not cleaved at the L5-W6 bond. Although there should be no change in accessibility of the site between the three enzymes, changes in primary sequence around this bond may affect the ability of the HIV-2 and SIV proteases to recognize the site.

Simple options to overcoming autoproteolysis, such as those described previously for stabilizing other proteases, were not readily available for the retroviral proteases. These enzymes exhibit a wide pH range for proteolytic activity (Hyland, et al., 1991) and the HIV-1 protease in particular is sensitive to extreme pH conditions. Furthermore, the highly compact structure of the protease homodimers make them less amenable to the mutations required for the aforementioned engineering approaches. It was crucial for us to identify a means of stabilizing these enzymes in order for us to perform biochemical analysis on mutants engineered for the purpose of studying structure-function relationships.

The proteolytic stabilization of HIV-1 protease was explored using a general method that sought to render known cleavage sites less labile by altering their amino acid sequences. This process was greatly facilitated by a knowledge of the substrate specificity of the HIV-1 and HIV-2 enzymes. Based upon data from many proteins and peptide substrates, algorithms have been constructed to predict the ability of the proteases to cleave substrates containing specific amino acids (Poorman, et al., 1991). Poorman's predictive algorithms and comparisons between HIV-1, HIV-2, and SIV protease sequences were used to design single-site mutants with the intent of preventing autoproteolysis in all three enzymes.

Materials and Methods

Site-directed mutagenesis. Mutagenesis was performed using the procedure described previously (see Chapter 1). The oligonucleotide 5'-CACCTTGTGG AAGAGACCACT AG-3', was used to replace Gln7 in HIV-1 protease with Lys. Mismatches are indicated by underlined nucleotides. Mutants were screened by sequencing through the site of mutation. The oligonucleotide 5'-CACCTTGTGG TCTAGACCACTAG-3', replaced Gln7 in HIV-1 protease with Ser and introduced a unique *Xba* I site (TCTAGA) which was used to screen mutants. The oligonucleotide 5'-CTCTCTTTGGCAGCGCCCGGTGGTCACCGC-3' replaced Lys7 in HIV-2 protease with Gln and introduced a unique *Nci* I site (CCCGG) which was used to screen mutants. The oligonucleotide 5'-CCTCA ATTCCATCTTTGGAGAGACC-3' was used to replace His4 of SIV protease with Ser, and mutants were identified by sequencing through the site of mutation. All mutants were later confirmed by sequencing the entire protease gene. Genes with the desired mutations were cloned into the bicistronic vectors pT2Cb(3)HIV-1, pT2op(3)S/SIV, and pT2op(3)H2/115 for the expression of the HIV-1, SIV, and HIV-2 proteases, respectively.

Production and purification of recombinant proteins. Saturated cultures of *E. coli* (strain X90) were diluted 10-fold into fresh LB medium plus 50 μg/ml ampicillin and incubated for 1.5 hours at 37°C. IPTG was added to 500 μM to induce expression. Cultures were allowed to grow for two hours and then harvested by centrifugation at 7,000 rpm in a GSA rotor. Cell pellets were resuspended in 5 volumes of lysis buffer (50 mM HEPES, pH 7.5, 1 mM EDTA, 1mM DTT) and lysed by sonication. Lysates were clarified at 4°C by centrifugation at 14,000 rpm for 30 minutes in an SS-34 rotor.

a. HIV-1 protease: The clarified supernatant was treated with 0.5% (w/w cell pellet) protamine sulfate and spun at 42,000 rpm for 30 minutes in a Ti-45 rotor to remove nucleic acids. After adjusting the conductance of the supernatant to 2.8 millisieverts (mS) it was loaded onto a DEAE Sepharose column at pH 7.5. The protease was collected in the flow through. This material was adjusted to pH 6.0 with 2 M MES and a conductance of 2.2 mS by dilution with water, and loaded onto an S-Sepharose column equilibrated in 50 mM MES, pH 6.5, 1 mM EDTA, 1 mM DTT. The protease was retained on the column and eluted with a gradient of 0 to 1 M NaCl. Active fractions from this elution were pooled and loaded onto a Pepstatin-A Agarose affinity column equilibrated in 50 mM HEPES, pH 7.5, 1 mM EDTA, 1mM DTT. Protease was eluted with a buffer containing 100 mM ε-amino caproic acid, pH 10.5, 50 mM NaCl, 5% glycerol, 5% ethylene glycol, 1 mM EDTA, 1 mM DTT. Active fractions were neutralized with 2 M HEPES and frozen for storage. These fractions were analyzed by SDS-PAGE on 13% polyacrylamide gels and stained with Coomassie Blue. Purified protease was shown to contain a small amount (<2%) of a 14kDa contaminant. Yields were approximately five milligrams of purified protease from 10 grams of *E. coli.*

b. HIV-2 protease: Clarified supernatant generated by sonication was precipitated with 57.5% ammonium sulfate overnight, followed by centrifugation for 45 minutes at 12,500 rpm in an SS-34 rotor. The pellet was resuspended in 50 mM HEPES, pH 7.5, 1 mM EDTA and loaded on a column of Pepstatin-A Agarose. The protease was eluted with a buffer containing 250 mM ε-amino caproic acid, pH 10.5, 5% glycerol, 5% ethylene glycol and 1 mM EDTA. Active fractions were pooled and loaded on a DEAE Sepharose column at pH 10.5. Following extensive washing, the protease was eluted with 50 mM HEPES, pH 7.5, 1 mM EDTA. HIV-2 PR K7Q was purified in essentially the same way as

HIV-2 protease except that elution from the DEAE Sepharose column required a gradient of 0 to 0.5 M NaCl in 50 mM HEPES, pH 7.5 buffer following the change of pH. Yields were between three and ten milligrams of purified protein from 17 grams of *E. coli.* The purified protease was shown to be of single-band purity by Coomassie Blue stained SDS-PAGE (13% polyacrylamide).

c. SIV protease: This enzyme was purified using essentially the same protocol as HIV-2 protease. Protease eluted from the DEAE Sepharose column with 50 mM HEPES, pH 7.5, 1 mM EDTA was found in fractions containing the pH change from 10.5 to 7.5. Material with the highest activity:A280 ratio was frozen at -70°C. These fractions were analyzed by SDS-PAGE on 13% polyacrylamide gels and shown to be of single-band purity by Coomassie Blue staining. This protocol yielded eight milligrams of purified protease from five grams of *E. coli.*

Kinetic analysis of proteases. Protease activity was measured by the method of Pichuantes *et al.* (Pichuantes, et al., 1990). Samples were incubated in activity buffer (50 mM sodium acetate, pH 5.5, 1 M NaCl, 1 mM DTT, 1 mM EDTA) in the presence of the decapeptide substrate NH_2-ATLNFPISPW-COOH (Peptide I). Cleavage products were separated on a reversed-phase C_{18} Pecosphere column (Perkin Elmer). k_{cat} and K_M determinations were performed in triplicate using Peptide I concentrations up to 2.5 mM in order to bracket K_M. Purified protein concentrations were determined by active site titration (Tomasselli, et al., 1990) using inhibitor U85548 (a gift of Dr. A. Tomasselli, Upjohn Laboratories).

Cleavage site identification. The identity of the first cleavage site in each protease was determined as follows. Approximately 5 μg (HIV-1 protease) or 10 μg (HIV-2 protease, SIV protease) of purified protease was incubated in activity

buffer at 37°C as in the time course assays mentioned previously. At a time corresponding to one half life, the sample was heated briefly at 75°C to inactivate the enzyme and frozen at -70°C. The contents of the solution were bound to a PVDF membrane by spinning the digested solution through a Pro-Spin column (ABI). Sequential Edman degradation was performed on the membrane using an Applied Biosystems model 470A gas-phase protein Sequencer equipped with an on-line phenylthiohydantoin-derivative analyzer. The phenylthiohydantoin-derivatives were separated by reverse phase chromatography on a Brownlee C-18 column (Hunkapillar, et al., 1983). Sequencing was carried through five cycles and the identified sequences matched to the known protease sequence. In order to identify all of the cleavage sites in each protease, reactions were set up as before, but digestions were extended to five half-lives before the reaction was stopped and transferred to PVDF for analysis as previously described.

Autoproteolysis. Time courses of protease decay were performed as follows. Purified proteases were incubated at 37°C in activity buffer at concentrations of five μg/ml (HIV-1 protease) or 10 μg/ml (HIV-2, SIV protease). At periodic intervals 10 μl aliquots were withdrawn and assayed for proteolytic activity as described. Decay curves were generated by comparing the remaining activity at a given time (A_t) to the initial activity (A_0). Curves were fit to either a first order rate equation $A_t = A_0e^{-kt}$ or a simple second order rate equation $A_t = A_0/(1 + A_0*k_2*t)$, where t is time and k, k_2 are the rate constants for inactivation.

Results

Identity of Cleavage Sites. N-terminal sequencing of endpoint digestions of HIV-1, HIV-2 and SIV proteases yielded the identities of the cleavage sites listed

in Table 2-1. These sites are shown graphically in Figure 2-1. The HIV-1 protease was cleaved in three locations: the N-terminal site L5-W6 and two additional sites at L33-E34 and P63-V64. The SIV protease was also cleaved at an N-terminal site, F3-S4, and at two additional sites G35-I36 and G52-F53. The HIV-2 protease was cleaved at the same positions as SIV protease, as well as at three additional positions at Y14-I15, M76-T77 and L90-T91.

By using N-terminal sequence data from early time points in the digestion it was possible to identify sites which were clearly cleaved first, labeled "primary sites" in Table 2-1. Cleavage at these sites is correlated directly with the inactivation of the enzyme. The sites which are listed as secondary in Table 2-1 are sequences which were detected only after the primary cleavage event had occurred. Although cleavage at these sites appears to be sequential, the order of the autoproteolysis reaction has not been extensively analyzed.

Design of autoproteolysis mutants. A statistical algorithm has been written for predicting the tendency of a protein sequence to be a substrate of HIV-1 protease based on data accumulated from natural protein substrates (Poorman, et al., 1991). This algorithm assumes that the preference of a given pocket of the enzyme for a given amino acid in a substrate is independent of the identity of the other residues in the substrate. By observing the frequency with which residues appear at given positions in the substrate, tables of probabilities have been constructed which indicate the apparent specificity of the enzyme for a given residue at a given position. The probability (h) that a peptide will be a substrate is determined as a ratio which compares the products of the specificity values (s) for the eight amino acids in a substrate to the same value plus a scaling factor (1/PS) which corrects for occurrences where a random octapeptide is a substrate.

Figure 2-1: Autoproteolysis sites in retroviral proteases.

Cleavage sites were identified by incubating 5 μg of HIV-1 proteases or 10 μg of HIV-2 or SIV proteases in activity buffer for a time approximately equal to five half lives. Arrows indicate sites identified by N-terminal sequencing of digested samples transferred directly to PVDF membranes.

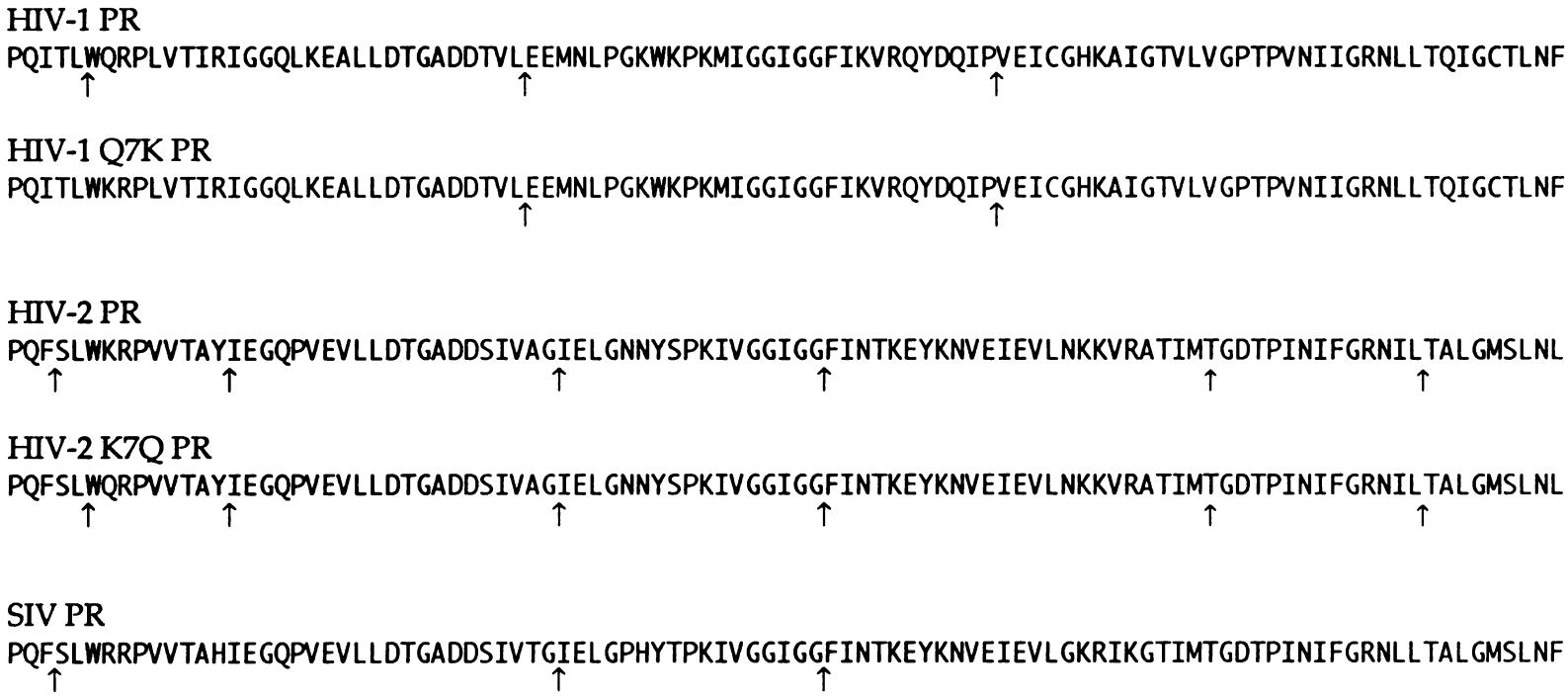

HIV-1 PR
PQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYDQIPVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
HIV-1 Q7K PR
PQITLWKRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYDQIPVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
HIV-2 PR
PQFSLWKRPVVTAYIEGQPVEVLLDTGADDSIVAGIELGNNYSPKIVGGIGGFINTKEYKNVEIEVLNKKVRATIMTGDTPINIFGRNILTALGMSLNL
HIV-2 K7Q PR
PQFSLWQRPVVTAYIEGQPVEVLLDTGADDSIVAGIELGNNYSPKIVGGIGGFINTKEYKNVEIEVLNKKVRATIMTGDTPINIFGRNILTALGMSLNL
SIV PR
PQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGGFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNF

Table 2-1. Cleavage sites identified in retroviral proteases.

HIV-1 PR	HIV-1 Q7K PR	HIV-2 PR	HIV-2 K7Q PR	SIV PR
L5-W6	L33-E34	**F3-S4**	**L5-W6**	**F3-S4**
L33-E34	P63-V64	Y14-I15	Y14-I15	G35-I36
P63-V64		G35-I36	G35-I36	G52-F53
		G52-F53	G52-F53	
		M76-T77	M76-T77	
		L90-T91	L90-T91	

Sites at which cleavage is associated with initial activity loss (primary sites) are shown in bold.

The values for PS are enzyme- and case-specific.

$$\text{(Equation 2-1)} \quad h = \frac{\prod_{1}^{8} S_{ij}}{\prod_{1}^{8} S_{ij} + 1/PS}$$

The specificity data for the twenty amino acids in an HIV-1 or HIV-2 protease substrate is reproduced in Appendix B. The PS values used for HIV-1 and HIV-2 predictions were 0.006 and 0.004, respectively. Using this data, the Poorman algorithm predicts the probability of cleavage (h) at the L5-W6 bond to be 0.46. The analogous HIV-2 sequence (QFSLWKRP), which is not cleaved by the HIV-2 protease, is calculated at h = 0.020, suggesting that it would also not be cleaved by HIV-1 protease. The reason for this block of cleavage would appear to be the presence of Lys at position 7 (P_2' relative to L5-W6), which has a probability of occurrence in an HIV-1 substrate of 0.09 and causes the single greatest drop in cleavage probability for this sequence. Previous mutagenesis data suggested that lysine at P_2' in an otherwise native processing site is sufficient to prevent proteolytic processing (Margolin, et al., 1990).

It was our intention that autoproteolysis stabilization mutations should have minimal effects on protein function. Mutation of Gln7 to Lys in HIV-1 protease should have minimal effects on protein structure, as this residue is fully solvent accessible and distant from the vicinity of the active site. Concern for the effect of increased positive charge on the activity of HIV-1 PR Q7K prompted the construction of HIV-1 PR Q7S. The presence of Ser7 in the N-terminal cleavage site also reduces the cleavage probability to 0.020.

The mutant HIV-2 PR K7Q was designed to test the effect of an HIV-1 protease sequence substitution at position 7 on the rate of autodegradation and

position of the initial cleavage site for HIV-2 protease. Given the number of cleavage sites in the HIV-2 protease we decided not to attempt stabilization of this enzyme. A mutant was designed for the stabilization of the SIV protease, however, which appeared to have a manageable number of sites. The mutation S4H was designed based on the observation that a point substitution of His in SIV protease for Tyr in HIV-2 protease was sufficient to prevent cleavage at an otherwise identical sequence from amino acids 11-18.

Autoproteolysis of wild type and mutant HIV-1 protease. To make accurate comparisons among the mutant enzymes it was important to assess any effect of the mutations on their kinetic characteristics. As shown in Table 2-2, HIV-1 PR Q7K differed from the wild type enzyme by increases of 20 - 25% in k_{cat} and K_M, but the overall specificity as measured by k_{cat}/K_M was identical within experimental error. HIV-1 PR Q7S had kinetic values which varied considerably from wild-type enzyme, showing an overall 14% decrease in specific activity as measured by k_{cat}/K_M. Although there was a moderate (7%) decrease in K_M, the majority of the activity difference arose from a 30% decrease in k_{cat}. This loss of activity implies that even though position 7 is remote from the active site, the choice of amino acid substitution may have a significant effect on activity. HIV-1 PR Q7K displayed a lower solubility in *E. coli* than wild type protein and HIV-1 PR Q7S, presumably due to its increased positive charge. The decreased solubility of this mutant was also evident with purified material.

The stability of the wild type and mutant enzymes is shown in Table 2-3. Replacing Gln7 with Lys or Ser resulted in a stabilization of the proteolytic activity at 37°C of at least 15-30-fold. Time course SDS-PAGE of wild type HIV-1 protease correlated the loss of proteolytic activity with formation of a truncated

Table 2-2. Kinetic parameters for wild type and mutant proteases.

protein	k_{cat} min^{-1}	K_M μM	k_{cat}/K_M $min^{-1}\mu M^{-1}$
HIV-1PR	815 ± 20	990 ± 50	0.82 ± 0.04
HIV-1 PR Q7K	1030 ± 30	1280 ± 60	0.81 ± 0.03
HIV-1 PR Q7S	580 ± 50	820 ± 40	0.70 ± 0.07
HIV-2 PR	180 ± 10	740 ± 70	0.24 ± 0.02
HIV-2 PR K7Q	260 ± 15	950 ± 100	0.26 ± 0.02
SIV PR	220 ± 20	1030 ± 50	0.22 ± 0.02
SIV PR S4H	277 ± 30	1500 ± 250	0.18 ± 0.04

k_{cat} and K_M were determined using substrate I (NH_2-ATLNF-PISPW-COOH) in standard assay buffer (50 mM NaOAc, 1 M NaCl, 1 mM EDTA, 1 mM DTT).

Table 2-3. Inactivation kinetics for wild type and mutant proteases.

protein	sequence	k_2 $ng^{-1}min^{-1}$	$t_{1/2}$ (10ng/μl)
HIV-1PR	PQITL↑WQRP	$2.6 \pm 0.1 \times 10^{-5}$	38 ± 3
HIV-1 PR Q7K*	PQITLWKRP	$2.28 \pm 1 \times 10^{-4}$	3100 ± 200
HIV-2 PR	PQF↑SLWKRP	$5.9 \pm 0.2 \times 10^{-6}$	170 ± 3
HIV-2 PR K7Q	PQFSL↑WQRP	$2.4 \pm 0.1 \times 10^{-5}$	42 ± 2
SIV PR	PQF↑SLWRRP	$1.9 \pm 0.1 \times 10^{-6}$	530 ± 30
SIV PR S4H*	PQFHLWRRP	$5.0 \pm 0.1 \times 10^{-4}$	1300 ± 100

Rates of inactivation determined in standard assay buffer (50 mM NaOAc, 1 M NaCl, 1 mM EDTA, 1 mM DTT) at 37°C. k_2 represents the apparent second order rate constant for activity loss taken from the equation describing the decay of each protease. Arrows indicate sites of autocleavage associated with the loss of activity. Asterisks indicate probable first order decay processes, rate constants reported are therefore k_1 with units of min^{-1}.

species (Figure 2-2). N-terminal sequencing of digestion products of HIV-1 protease identified this truncated species as being identical to HIV-1 PR Δ1-5.

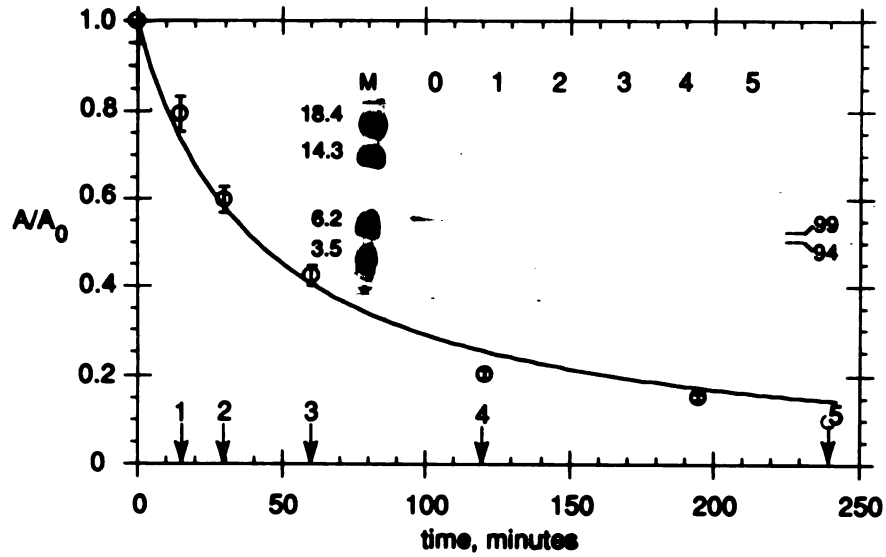
When attempting to identify the site of degradation responsible for the loss of activity in HIV-1 PR Q7K, it was discovered that the loss of activity did not correlate with the appearance of a truncated form of the protease, as only native N-terminal sequences were identified. Due to the length of the assays needed for examining the stability of HIV-1 PR Q7K and HIV-1 PR Q7S, reversible or irreversible oxidation of cysteine 67 and/or 95 may have occurred. This has been observed to cause a loss of proteolytic activity (Strickler, et al., 1989). When additional DTT was added during the digest, the protein recovered nearly 80% of its initial activity following 88 hours of incubation at 5 ng/μl (Figure 2-2). Our initial estimation of the protein stability probably represented the rate of inactivation by cysteine oxidation under the conditions reported. N-terminal sequencing of HIV-1 PR Q7K digestions carried out with higher concentrations of reducing agent for 4 days (~90 hours) identified fragments generated by cleavage at L33-E34 and P63-V64.

Autoproteolysis of wild type and mutant HIV-2 protease. HIV-2 protease is initially inactivated by a cleavage event taking place at the F3-S4 bond, as shown in Table 2-1 and Figure 2-3. Presumably this cleavage event occurs because the N-terminal region is still accessible for cleavage but the Lys at position 7 prevents cleavage at L5-W6. To demonstrate that the L5-W6 bond cleavage was affected by sequence and not a change in accessibility, the Lys at position 7 was substituted with Gln. The presence of an HIV-1 protease-like sequence at position 7 on HIV-2 protease was expected to destabilize the enzyme. The k_{cat} of HIV-2 PR was increased 23 - 30% relative to the wild type enzyme but again the

Figure 2-2: Activity decay curves for HIV-1 protease (A) and HIV-1 PR Q7K (B).

HIV-1 protease was assayed at 5 ng/μl, HIV-1 PR Q7K at 10 ng/μl. A/A_0 represents the ratio of active protein at a given time (A) to the total active protein at time = 0 (A_0). Decay curves were generated by assaying remaining protease activity at various time points (o). Numbered arrows indicate samples analyzed by SDS-PAGE and stained with Coomassie Blue (see figure insets). Sample times for inset in (A) are 0 (0), 45 (1), 75 (2), 130 (3), 240 (4) and 480 (5) minutes. For HIV-1 PR Q7K, 5 mM DTT was added to the reaction mixture at 4000 minutes prior to final sampling. A/A_0 at this time point was 0.8 following addition of DTT (data not shown). Sample times for inset in (B) are 0 (0), 1450 (1), 3400 (2), and 4380 (3) minutes.

A.


1.0
0.8
0.6
0.4
0.2
0
A/A0
M 0 1 2 3 4 5
18.4
14.3
6.2
3.5
99
94
1 2 3 4 5
0 50 100 150 200 250
time, minutes

B.

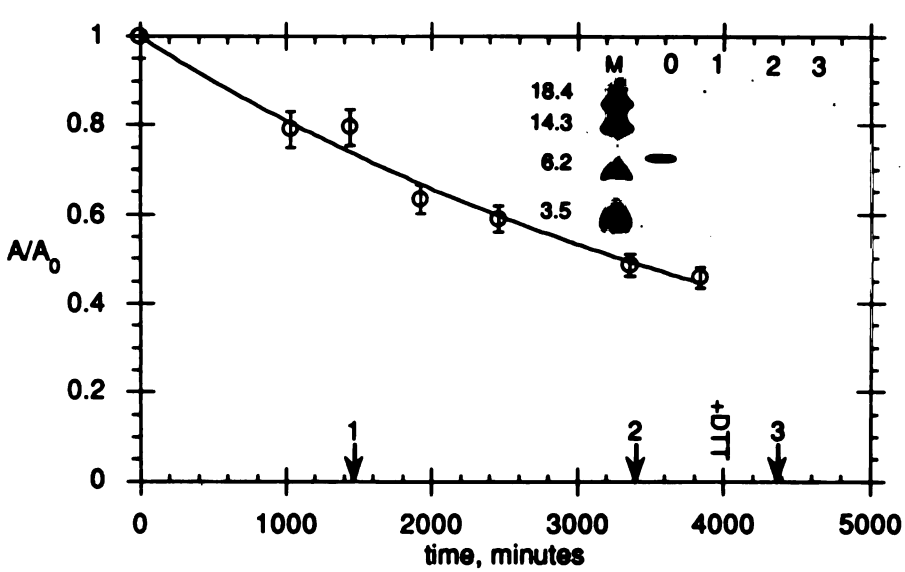
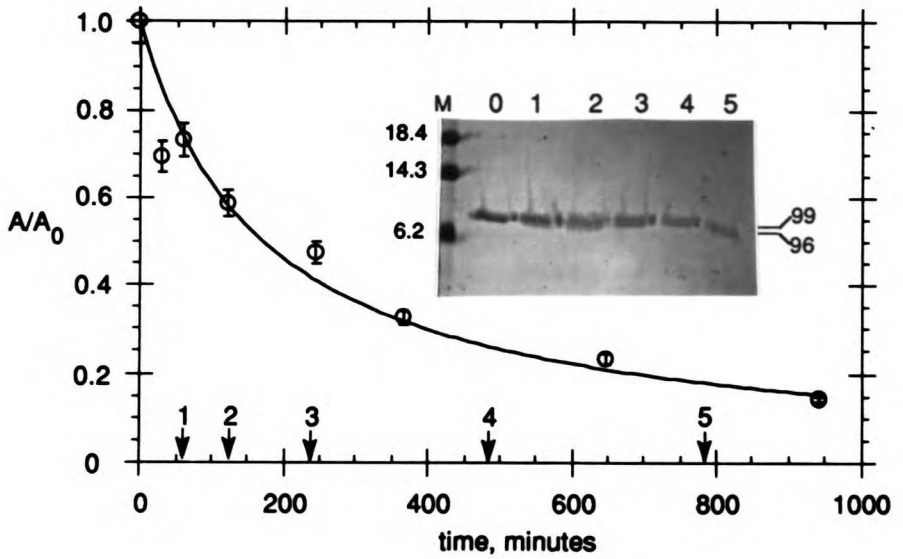

1
0.8
0.6
0.4
0.2
0
A/A0
M 0 1 2 3
18.4
14.3
6.2
3.5
1 2 +DTT 3
0 1000 2000 3000 4000 5000
time, minutes

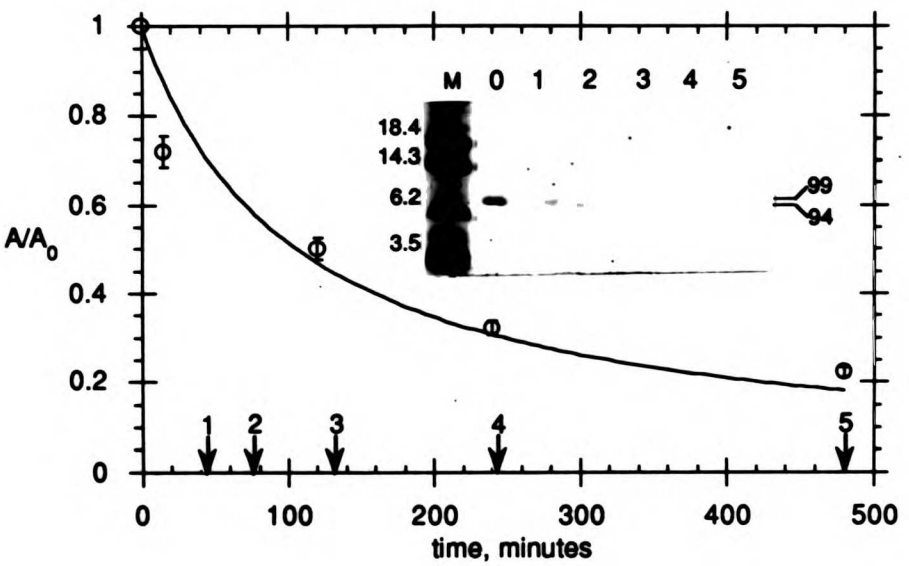
Figure 2-3: Activity decay curves for HIV-2 protease (A) and HIV-2 PR K7Q (B).

Starting concentrations of HIV-2 protease and HIV-2 PR K7Q were 10 ng/μl. A/A_0 represents the ratio of active protein at a given time (A) to the total active protein at time = 0 (A_0). Decay curves were generated by assaying remaining protease activity at various time points (o). Numbered arrows indicate samples analyzed by SDS-PAGE and stained with Coomassie Blue (see figure insets). Sample times for inset in (A) are 0 (0), 60 (1), 120 (2), 240 (3), 480 (4), and 780 (5) minutes. Sample times for inset in (B) are 0 (0), 15 (1), 30 (2), 60 (3), 120 (4), and 240 (5) minutes.

A.


1.0
0.8
0.6
0.4
0.2
0
A/A0
0
200
400
600
800
1000
time, minutes
M 0 1 2 3 4 5
18.4
14.3
6.2
99
96
1
2
3
4
5

B.


1
0.8
0.6
0.4
0.2
0
A/A0
0
100
200
300
400
500
time, minutes
M 0 1 2 3 4 5
18.4
14.3
6.2
3.5
99
94
1
2
3
4
5

overall specificity was identical within experimental error (Table 2-2). The half lives of the two enzymes showed that the mutation destabilized HIV-2 PR K7Q three-fold relative to HIV-2 protease (Table 2-3). SDS-PAGE analysis of samples of the proteases taken during a time course showed the formation of a truncated species with a mobility very similar to the initial material (Figure 2-3). N-terminal sequencing showed that the truncated species of HIV-2 PR K7Q was degraded exclusively at the L5-W6 bond; no species suggesting cleavage at F3-S4 was detected. Complete digestion of the protease showed that changing the N-terminal autolysis site had no effect on downstream sites.

Autoproteolysis of the SIV protease. The expression and purification system developed for the SIV protease allowed for the production of this protease at considerably higher levels than the HIV-1 and HIV-2 proteases. Since this enzyme differs from HIV-2 protease by only 9 amino acids, its kinetic parameters and degradation profile were examined relative to HIV-2 protease. This enzyme has kinetic constants very similar to those of HIV-2 protease, shown in Table 2-2, as suggested by previous analysis (Grant, et al., 1991). The half life for this enzyme, however, is three-fold longer than the half life of HIV-2 protease (Table 2-3). A time course of decay for this enzyme on SDS-PAGE identified several truncated species resulting from autoproteolytic degradation. N-terminal sequencing of the earliest observed species indicated that it was generated by cleavage at the F3-S4 bond.

The mutant SIV PR S4H was made and analyzed for stability to autolysis. The mutation had only minor effects on the kinetic parameters of the enzyme, decreasing k_{cat}/K_M by approximately 20 percent. A comparison of the decay curves for the wild type and mutant enzymes was made at 40 ng/μl due to the

very slow decay rate of the SIV PR S4H mutant (Figure 2-4). As had been observed for HIV-1 PR Q7K, the S4H enzyme lost activity considerably more slowly than wild type SIV PR, with an approximately six-fold increase in half life. Activity loss was not directly correlated with the detection of cleavage products, suggesting that another process was responsible for the observed loss of activity.

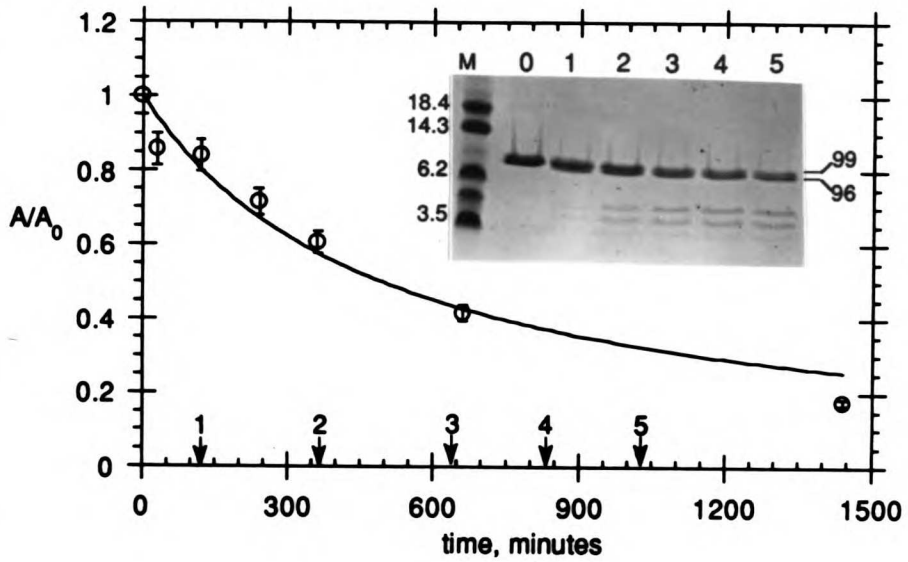
Analysis of Reaction Order. A mechanism for inactivation of these enzymes can be proposed based on the order of the reaction. Autoproteolysis could involve an intermolecular cleavage of one protease by another, a second order process, or one protease cleaving itself in a first order process. In the first case, the half life of the active species would be sensitive to protease concentration, whereas an intramolecular cleavage event would be insensitive to concentration. Wild type SIV protease was used to examine the order of the autoproteolytic reaction because of its relative ease of production and comparable activity to the other two enzymes. Decay curves were generated using three different concentrations of protease, and the half lives measured for each concentration (Figure 2-5). The observation that the half life varied as a function of concentration suggested that the decay process is second order, and therefore represents an intermolecular cleavage event. At lower concentrations, an apparent first order component to activity loss became more significant. It is probable that this component represents activity loss due to dimer dissociation. This first order component was seen at concentrations predicted to be 50-60 nM for HIV-1 protease and 15 - 30 nM for HIV-2 and SIV proteases, which is within the range of the dimer dissociation constant observed under these assay conditions (Babé and Craik, 1992).

The inactivation of HIV-1 PR Q7K and HIV-1 PR Q7S did not follow the

Figure 2-4: Activity decay profile for SIV protease (A) and SIV PR S4H (B).

Starting concentration of SIV protease was 10 ng/μl. A/A_0 represents the ratio of active protein at a given time (A) to the total active protein at time = 0 (A_0). Decay curves were generated by assaying remaining protease activity at various time points (o). Numbered arrows indicate samples analyzed by SDS-PAGE and stained with Coomassie Blue (see figure inset). Sample times are 0 (0), 120 (1), 360 (2), 645 (3), 840 (4), and 1040 (5) minutes. In panel (B) both SIV protease and SIV PR S4H were assayed at 40 ng/μl.

A.


1.2
1
0.8
0.6
0.4
0.2
0
A/A0
M 0 1 2 3 4 5
18.4
14.3
6.2
3.5
99
96
1
2
3
4
5
0
300
600
900
1200
1500
time, minutes

B.

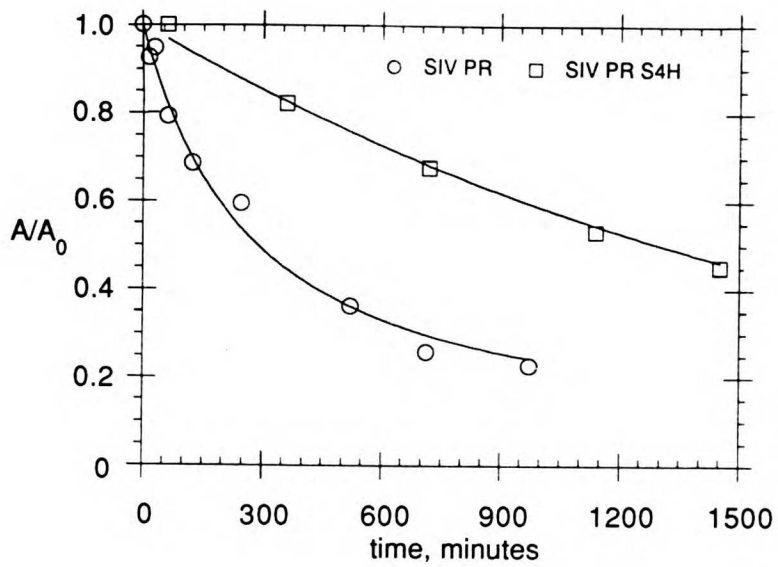
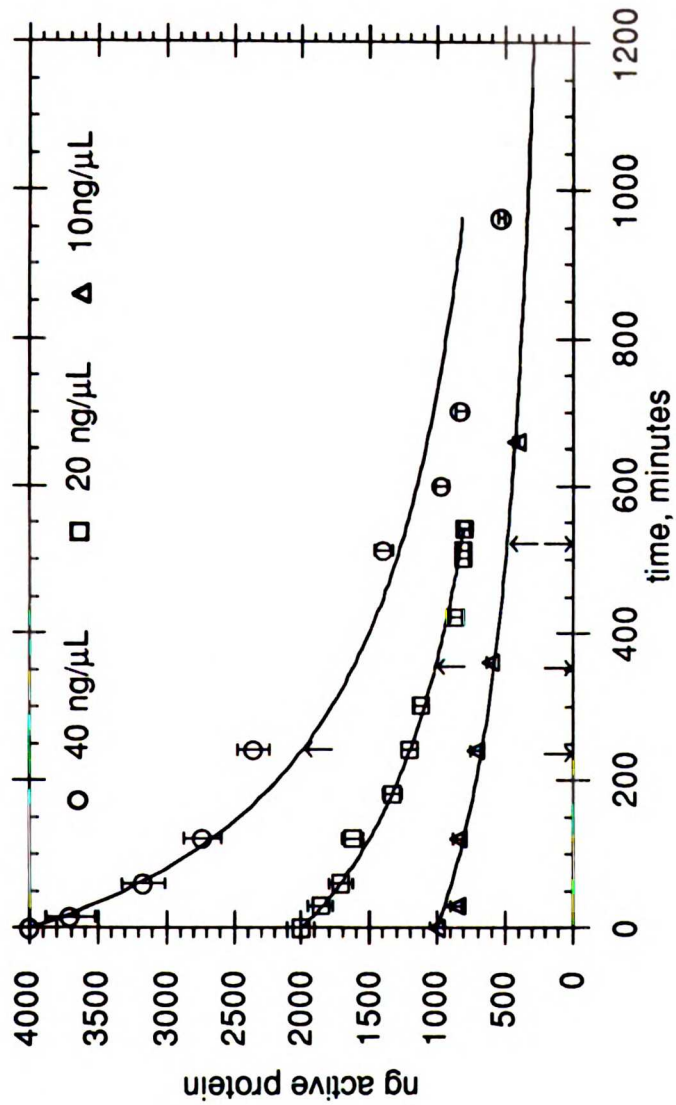

1.0
0.8
0.6
0.4
0.2
0
A/A0
SIV PR
SIV PR S4H
0
300
600
900
1200
1500
time, minutes

Figure 2-5: Effect of SIV protease concentration on decay curve.

Initial concentrations of SIV protease were 10 (Δ), 20 (❑) and 40 (O) ng/μl. Active protein concentration was determined by assaying activity at the indicated time points and normalizing it to the activity at T = 0. Half lives are indicated by arrows on each curve and corresponding intersections at the X-axis. The endpoint of the 10 ng/μl curve was fit to a data point at 1440 minutes which was omitted from the graph.


40 ng/µL
20 ng/µL
10ng/µL
ng active protein
time, minutes
0
500
1000
1500
2000
2500
3000
3500
4000
0
200
400
600
800
1000
1200

second order pattern described above. Samples at 5 and 10 ng/μl displayed approximately the same rate of activity loss (data not shown). This concentration independence and the ability of additional DTT to restore activity suggest that these enzymes lose activity primarily through a first order process, probably cysteine oxidation. It is also likely that during the course of an extended assay (>72 hours) the protein undergoes irreversible thermal denaturation. This would be a likely explanation for the loss of activity of the SIV PR S4H mutant, which has no cysteine residues.

Discussion

The regulation of protease activity is often achieved by autoproteolytic cleavage events. In the case of bovine chymotrypsin, the precursor chymotrypsinogen undergoes activation to α-chymotrypsin through a series of as many as nine proteolytic intermediates which are stable and have been isolated and characterized (Sharma and Hopkins, 1981). These intermediates represent pathways for either slow or rapid activation of this enzyme. Autoproteolytic cleavages also play a role in inactivating these digestive enzymes. Fully activated bovine α-trypsin undergoes autolytic cleavage at K176-N177 which destroys the affinity of the enzyme for its substrates (Smith and Shaw, 1969), although disulfide bonding keeps most of the native structure intact.

Total autoproteolytic degradation of the HIV-1, HIV-2, and SIV proteases is restricted to cleavage at three to five sites whose identities are dictated by the particular sequence specificity of these proteases. The initial cleavage events occur at an N-terminal strand or loop which is exposed and likely to be accessible to an attacking protease. Mutation of amino acids around these cleavage sites may render them non-optimal substrates, thus stabilizing the enzymes to autolysis. Of the three proteases, HIV-1 protease is an ideal model for type of manipulation because it has been extensively studied. A large database of mutants of HIV-1 protease (Loeb, et al., 1989) allows for the identification of amino acid positions which are sensitive to engineered stabilization mutations. These positions can be avoided when selecting residues around a known cleavage site as targets for substitution.

In addition, extensive studies on substrate specificity of HIV-1 protease have provided information on substrate specificity determinants. This includes data

on nonviral protein substrates summarized by Poorman (Poorman, et al., 1991), as well as directed mutants of natural viral substrates (Jupp, et al., 1991; Margolin, et al., 1990; Partin, et al., 1990). These studies provide information as to which amino acids can be positioned at defined sites in a peptide sequence to render a substrate less labile. In this study we showed that the identity of residues which will prevent cleavage at a given position could be determined by which amino acids caused the greatest drop in the cleavage probability of the target sequence.

Cleavage at the HIV-1 PR L5-W6 bond in the N-terminal sequence QITLWQRP occurs very rapidly and is sufficient to inactivate the enzyme. This sequence, consequently, was an initial target for stabilizing mutations. The HIV-2 protease, which is not cleaved at the L5-W6 bond, differs significantly from HIV-1 protease in having a Lys at position 7. This residue was predicted by the Poorman algorithm to make the N-terminal sequence of HIV-1 protease a poor substrate. A mutation from Gln to Ser at position 7 was predicted to have a similar effect without affecting the isoelectric point of the enzyme. In HIV-1 protease, replacement of Gln7 (P_2' relative to the 5-6 bond) with Lys or Ser decreased the rate of proteolytic inactivation by greater than 100-fold at 37°C. Under normal reaction conditions HIV-1 PR Q7K and HIV-1 PR Q7S lost activity in a concentration independent manner (i.e. a first order process). SDS-PAGE analysis indicated that this loss of activity did not correlate with the generation of truncated forms of the protease. It was concluded that oxidation of cysteine residues, which has been reported to inactivate the enzyme (Strickler, et al., 1989) was the major cause of inactivation. Autolysis experiments performed under conditions of low oxygen and high reducing agent concentrations generated truncated forms of the protease at very long time points (>72 hrs) identified as

the results of proteolysis at either the L33-E34 or P63-V64 bond. Sequencing data was sufficient to identify the sites, but not the order in which they were cleaved. Both of these sites may be structurally occluded: L33-E34 by two β-strands formed from amino acids 76 - 79 and 82 - 88, and P63-V64 by a strand formed by amino acids 69-79. The length of time required to generate these species suggests that cleavage at such "secondary sites" may be restricted by conformational accessibility and may only occur as a function of denaturation of the enzyme. Recently, it has been reported that engineering of the two secondary sites as well as the primary site to include non-optimal residues stabilizes the protease even further than the Q7K mutation alone (Mildner, et al., 1994).

HIV-2 PR K7Q was made to confirm that the lack of cleavage at the L5-W6 bond of HIV-2 protease was not due to conformational differences between HIV-1 and HIV-2 proteases. The half life of HIV-2 PR K7Q was approximately the same as that of HIV-1 protease, having been destabilized three-fold relative to HIV-2 protease. Furthermore, the site of the original cleavage event shifted from F3-S4 to the better substrate L5-W6, recapitulating the pattern of cleavage in the N-terminus of HIV-1 protease.

The stabilization of SIV protease was effected by manipulation of the P1' residue at the cleavage site F3-S4. The P1' residue, Ser4, pointed away from the dimer interface and was not involved in any interactions involved in holding the dimer interface together. Although kinetic analysis suggested that the mutation affected the specific activity of the enzyme by approximately 20%, part of this may have arisen from difficulties which occurred during kinetic analysis. We were unable to measure substrate concentrations substantially above K_M, thus increasing the error in the measurement. SIV PR S4H was stabilized approximately 3-fold by this mutation, although the loss of activity which was

observed during time courses with the mutant enzyme was not consistent with a second order autoproteolytic degradation process. The absence of detectable degradation products by N-terminal sequencing suggested that the SIV PR S4H might also have lost activity through thermal denaturation or methionine oxidation. This stabilized mutant was highly useful in providing stable protease for crystallization trials, and eventually yielded diffraction quality crystals (Rose, et al., 1993). It was noted that the pI of the SIV PR S4H enzyme was shifted to 6.3, relative to the wild type pI of 5.2 (Todd Richmond, personal communication), which may also have facilitated crystallization. Analysis of the crystal structure indicated that, as expected, the histidine at position 4 made no contacts with the other residues of the dimer interface (Robert Rose, personal communication).

The SIV and HIV-2 proteases have amino terminal sequences which differ only by R or K, respectively, at position 7 (P_4' relative to the 3-4 bond). The observation that the rate of cleavage at F3-S4 differs by three-fold between these proteases suggests that the P_4' residue may affect substrate specificity. This implies that there are sites other than P_2' which may be useful sites for mutations which affect the lability of a cleavage site. The final autodigestion products of the HIV-2 and SIV enzymes differ in another manner. The efficient cleavage of HIV-2 protease at Y14-I15 is not observed in SIV protease, which is most likely due to the substitution of His for Tyr at position 14 in SIV protease. The absence of two other sites identified in HIV-2 protease (M76-T77, L90-T91) may be due to insensitivity of the assay and is being investigated. This type of comparison between related HIV and SIV proteases may identify other potential cleavage sites and provide independent confirmation of predicted stabilizing mutants.

The approach to stabilization outlined above may be pursued along two routes, reflecting two different mechanisms for overall degradation. In the first, a

primary cleavage event (for instance, in the N-terminus) results in a change in conformation of the protein. This allows for exposure of other cleavage sites which, in the native form, may be conformationally inaccessible. In this case, identification and elimination of the primary cleavage site should have a considerable impact on stabilization of the protein. This strategy may, however, expose a new primary cleavage site which would ordinarily not be observed due to a more rapid cleavage at the original site. This is the case observed with HIV-2 PR K7Q and HIV-2 protease. There are two potential cleavage sites available at the N-terminus in HIV-2 PR K7Q: F3-S4 and L5-W6. Autolysis occurs rapidly and exclusively at L5-W6, unless the site is made non-labile by a Lys at position 7 as is seen in the wild type enzyme. Autolysis of HIV-2 then occurs at the F3-S4 bond, although at a slower rate. Elimination of all cleavage sites in the amino terminus should stabilize the enzyme to a large extent.

Within the limits of detection of our experiments the N-terminal cleavage site was the only cleavage site associated with the initial loss of activity. If a low rate of cleavage is occurring simultaneously at other sites, then the rate of degradation at these sites also contributes to the observed rate of inactivation. This suggests a second general mechanism of degradation, where all potential cleavage sites are accessible in the native protein and the timing of cleavage at each site represents how well each sequence is recognized as a substrate. Thus, the observed half life for loss of protease activity represents the rate of cleavage at the bond which is the best substrate for the protease. In this case, elimination of autoproteolytic degradation requires mutation of all sites shown to exist by total degradation analysis. Although this is a more thorough approach, the number of mutations required may result in a dramatic effect on the structure or activity of the enzyme.

The effect of autoproteolytic cleavage on the activity of HIV-1 protease *in vitro* is quite dramatic, and the stabilized enzyme has provided a useful reagent for biophysical analysis. HIV-1 PR Q7K has consistently yielded large, diffraction quality crystals (Rutenber, et al., 1993). Although current efforts have focused on regulating the autoproteolysis of retroviral proteases, a similar strategy may be applied to the targets of a protease. Degradation of recombinant proteins can be a serious complication in expression and purification. By identifying the sites which are cleaved and determining the substrate specificity of the degrading protease, methods similar to those described in this report may be used to produce a more stable form of the expressed protein.

It is intriguing to consider the possibility that the autoproteolytic cleavage sites are actually utilized as a means of regulating protease activity. Following processing of the viral proteins and complete maturation of the virion the protease may undergo autoproteolytic cleavage as a means of downregulating its own activity. The situation may be considered analogous to the regulation of the activity of digestive serine proteases. It is possible that the presence of high levels of active protease inside a virion may be detrimental, as it increases the chances for degradation of crucial viral proteins. Using the vector system described in the next chapter we introduced the HIV-1 Q7K mutant protease into the viral genome. We did not observe any effects on the efficiency of virus production, or any decreases in viral infectivity, which suggests that the stablized protease activity is not detrimental to the virus. The ability to engineer proteases with variable half lives for *in vivo* as well as *in vitro* studies, however, may allow us to address the role of protease activity at later stages in the life cycle.

Chapter 3. Structure and Function and Virology of Catalytic Mutants of the HIV-1 Protease

Portions of this chapter have been submitted for publication. Rosé, J.R., Babé, L.M., and Craik, C.S. (1995) Defining the Level of HIV-1 Protease Activity Required for HIV-1 Particle Maturation and Infectivity. J. Virol., *in press*.

Abstract

The human immunodeficiency virus type-1 (HIV-1) protease is the enzyme required for processing of the gag and gag-pol polyproteins to yield mature, infectious virions. Although the complete absence of proteolytic activity prevents maturation, the level of activity sufficient for maturation and subsequent infectivity has not been determined. Amino acid substitutions have been engineered into the active site of the HIV-1 protease that reduce catalytic activity without affecting substrate recognition. The catalytic efficiency (k_{cat}) of the HIV-1 protease is decreased 4-fold when threonine 26 is replaced by serine, and approximately 50-fold when alanine 28 is replaced by serine. Genes containing these mutants were cloned into a proviral vector for analysis of their effect on virion maturation and infectivity. The results show that virions containing the T26S protease variant (T26S PR), in which only 25% of the protease is active, are very similar to wild type virions, although slight reductions in infectivity are observed. Virions containing the A28S protease variant (A28S PR) are not infectious, even though a limited amount of polyprotein processing does occur. There appears to be a linear correlation between the level of protease activity and particle infectivity. Our observations suggest that a threshold of protease activity exists between a 4-fold and 50-fold reduction, below which processing is insufficient to yield infectious particles. Our data also suggests that reduction of protease activity by 50-fold or greater is sufficient to prevent formation of infectious particles.

Proteolysis and the maturation process

The protease encoded in the HIV-1 genome plays a crucial role in the life cycle of the virus, cleaving the $p55^{gag}$ and $p160^{gag\text{-}pol}$ precursors into their mature and functional forms (Kohl, et al., 1988). Viruses in which the protease has been mutated to remove the active site aspartic acid residues will assemble in the absence of proteolytic activity, but the resultant virions are non-infectious and morphologically aberrant (Göttlinger, et al., 1989; Kohl, et al., 1988; Peng, et al., 1989). The development of inhibitors of the HIV-1 protease has been pursued as a means of preventing viral replication, and many inhibitors now exist which can block viral maturation in tissue culture (McQuade, et al., 1990; Tomasselli, et al., 1991; Wlodawer and Erickson, 1993) and slow viral replication *in vivo* (Martin, et al., 1994).

The exact requirements of the maturation step for protease activity are not clear. Although it is known that elimination of activity prevents maturation, detailed information regarding the minimal proteolytic activity required for formation of an infectious virion is not available. This is an important issue in determining how potent an inhibitor of the protease must be to prevent viral replication. It has been reported that in the presence of limiting concentrations of protease inhibitors it is possible to partially inhibit the protease, resulting in virions with small defects in processing and large decreases in infectivity (Kaplan, et al., 1993). We wished to study the effect of specific decreases in protease activity on viral maturation and infectivity, and determine if a correlation exists between protease activity and infectivity.

Site-directed mutagenesis has been carried out on the HIV-1 protease to determine the effects of mutants on substrate specificity and activity. A general rule derived from saturation mutagenesis of the protease is that the residues

within the active site are highly sensitive to mutation and generally cannot be substituted without severely compromising protease activity (Loeb, et al., 1989; Louis, et al., 1989). More directed mutations within the active site have generally been able to affect the protease's ability to recognize substrates (Sardana, et al., 1994). Only one mutant (A28S) has been reported in which the primary effect has been to alter the rate of proteolytic cleavage without significant effects on substrate binding (Ido, et al., 1991).

We proposed to examine the requirement for protease activity in the maturation process by studying protease mutants with specifically diminished catalytic activity. Three mutants that resulted in reduced proteolytic activity were engineered in the active site region of the protease (Figure 3-1). The A28S mutant has been characterized previously (Co, et al., 1994; Ido, et al., 1991) and is thought to reduce proteolytic activity by making hydrogen bonding interactions with the catalytic aspartates. A second mutation, T26S, was made after comparison with the sequence of the active site of the RSV protease (Figure 3-2). Tis residue is involved in a crucial hydrogen bonding array known as the "fireman's grip" which is involved in holding the active site loops in their proper orientation (Blundell and Pearl, 1989). A third mutant, D25N, replaced the active site aspartic acids with asparagines and served as a negative control for protease activity (Kohl, et al., 1988). These enzymes were expressed and characterized using the systems described previously in this report.

For analysis of the effect of the protease mutants on viral function, genes containing the single protease mutations were cloned into an HIV vector, pHIV-gpt, which produced all of the viral gene products except envelope (Page, et al., 1990). Transfection of this vector into mammalian cells permitted the formation

Figure 3-1: Location of mutants in the active site of HIV-1 protease.

Model showing α-carbon backbone with side chains displayed at the sites of mutation. Protease backbone is shown in blue with catalytic aspartates (D) 25 in red, glutamines (Q) 7 in magenta, threonines (T) 26 and alanines (A) 28 in green.

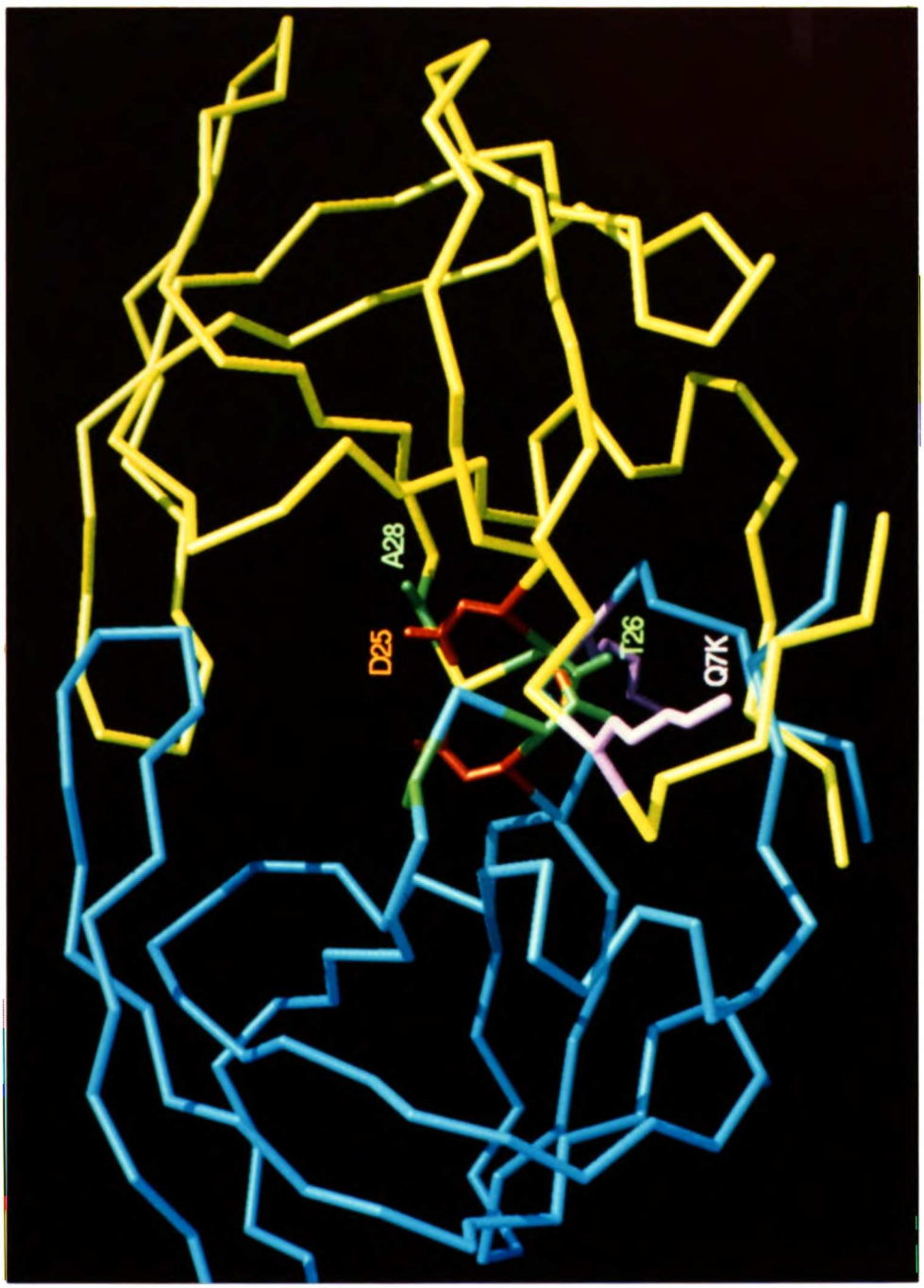
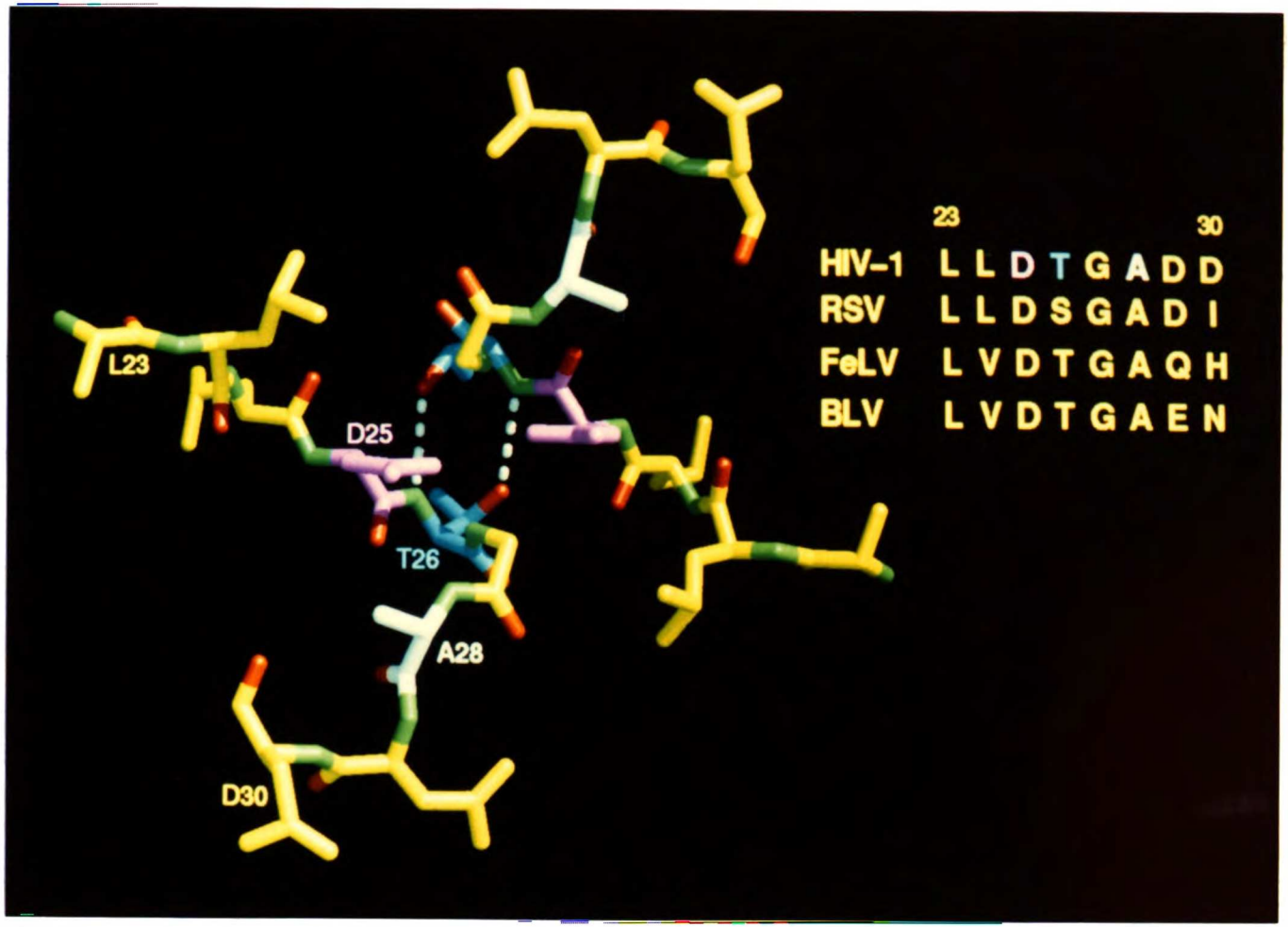

A28
D25
T26
Q7K

Figure 3-2: Alignment of viral protease active site sequences.

Model of the HIV-1 active site loop from residues 22-30. Backbone amide nitrogens are shown in green and carbonyl oxygens in red. Side chains for D25 are colored magenta, T26 in blue, and A28S in white. The hydrogen bonds between threonines 26 and 26', which form the fireman's grip, are shown as dashed lines. Sequences of Rous sarcoma virus (RSV), feline leukemia virus (FeLV) and bovine leukemia virus (BLV) are shown to the right to show sequence variation at the active sites of several distantly related retroviruses.


L23
D25
T26
A28
D30
23
30
HIV-1 L L D T G A D D
RSV L L D S G A D I
FeLV L V D T G A Q H
BLV L V D T G A E N

of viral capsids which were analyzed with regard to RNA content, polypeptide composition, and degree of maturation. By co-transfection of envelope-encoding plasmids, it was possible to generate infectious particles in order to analyze the effects of altered polyprotein processing on viral infectivity. This plasmid-based system was used for increased safety when dealing with viruses, since enveloped particles could only undergo one round of infection and would not pose as great a danger as virions in which the envelope gene was left intact. Our results indicate that although lowering the protease activity does not appear to affect the assembly process, reduction of activity by 50-fold or greater is sufficient to prevent the formation of infectious virions.

Materials and Methods

Site-directed mutagenesis. Mutagenesis of the bacterially and viral encoded HIV-1 protease genes was performed using standard methods (Kunkel, et al., 1987). The olignucleotides used to make these mutations in both bacterial and viral genes are listed in Table 3-1. Mutants were sequenced after identification by restriction-site screening to ensure no other mutations had occurred in the protease gene.

Purification of mutant proteases. Proteases carrying the HIV-1 Q7K mutation had an increased pI relative to wild type (9.3 to 9.5 vs 8.7 to 9.0, Todd Richmond, personal communication) and were purified using a protocol which was significantly modified from that used for wild type protease. Cells were grown and protein expression induced as described previously. For 10 grams of

Table 3-1. Oligonucleotides used to introduce mutations into HIV-1 protease genes.

mutation	gene	oligonucleotide sequence	restriction site
D25N	bacterial	5'-GAAGCTTTGTTG**A**ACA**CAGGTGCTG**ACG-3'	*Alw*N I
D25N	viral	5'-GAAGCTCTATTA**A**A**C**A**CCGG**AGCAGATG-3'	*Hpa* II
T26S	bacterial	5'-CTTTGTTGGAC**T**C**A**GGTGCTGACGA-3'	*Alw*N I
T26S	viral	5'-GCTCTATTA**GATTCT**GGAGCAGATGAT	*Hinf* I
A28S	bacterial	5'-GGACACC**GGATCT**GACGACACC-3'	*Bst*Y I
A28S	viral	5'-TTAGATACA**GGATCC**GATGATACAG-3'	*Bam*H I

Nucleotides in bold indicate those which make up the restriction sites used to screen mutations.

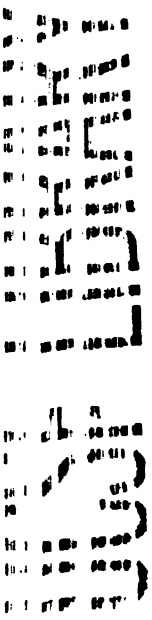
Underlined nucleotides indicate mismatches with the wild type genes.

E. coli, cells were resuspended in 125 ml of PCB (50 mM Tris, pH 7.5, 5 mM EDTA, 1 mM DTT, 1 mM PMSF) and sonicated in five minute bursts on ice until clear. Cell lysates were clarified by centrifugation at 42,000 rpm for 30 minutes in a Ti45 rotor. Clarified supernatant was added to 30 ml of P11 phosphocellulose (Whatman, Maidstone, England) pre-equilibrated in PCB. The slurry was shaken slowly at room temperature for thirty minutes and poured into a 2.5 cm column. The resin was washed with approximately 4-7 column volumes of PCB until the A_{280} of the column wash fell below 0.05. Protease was eluted with a 100 ml gradient from 0 to 750 mM NaCl in PCB. The active fractions were pooled and mixed with 10 ml of Pepstatin-A Agarose equilibrated in 50 mM HEPES, pH 7.5, 0.4 M ammonium sulfate, 1 mM EDTA, 1 mM DTT. Following a twenty minute incubation at room temperature the slurry was poured into a 1.5 cm column and washed until the A_{280} fell below 0.05. Protease was eluted with a buffer containing 100 mM ε-amino caproic acid, pH 10.5, 50 mM NaCl, 5% glycerol, 5% ethylene glycol, 1 mM EDTA, 1 mM DTT. Active fractions were neutralized with 2 M HEPES and frozen for storage or used fresh for crystallization trials. The final yield from this procedure was 5-7 mg of purified protease from 10 grams of *E. coli*.

Crystallization of HIV-1 protease. Purified HIV-1 protease was concentrated using Centricon-10 microconcentration units (Amicon, Beverley, MA) which had been pre-rinsed with 2 ml of distilled water. The buffer volume was reduced from 2 ml to approximately 500 µl and then diluted back to 2 ml with crystallization buffer containing 20 mM NaOAc, 50 mM NaCl, 1 mM EDTA and 1 mM DTT. This solution was concentrated to 500 µl again and the buffer exchange repeated with crystallization buffer to which the small molecule

inhibitor UCSF-8 (Rutenber, et al., 1993) had been added at a final concentration of 200 μM. After the buffer exchange process was completed the concentration of the protease was determined by Bio-Rad Bradford assay, using bovine serum albumin (BSA) as a standard. In the final step the protease was concentrated to a calculated concentration of 2.5 mg/ml. For further concentration, the protein solution was transferred to a siliconized Eppendorf tube and the volume reduced by passing a stream of dry nitrogen over the solution.

Crystals were grown using the hanging drop vapor diffusion method. The well buffer used contained 20 mM NaOAc, 1 M NaCl, 1 mM EDTA, and 1 mM DTT. Drops with a final volume of 5 μl were set up by mixing 2.5 μl of protein solution with 2.5 μl of well buffer. Drops were set up at protein concentrations between 2.5 and 3.5 mg/ml. Crystals appeared within 18 hours and grew over a period of 3 days. Data was collected by Earl Rutenber and Robert Rose in the laboratory of Dr. Robert Stroud (UCSF) on a multiwire detector (Siemens X-100) using Cu Kα X-rays from a rotating anode generator (Rigaku Ru200) with a graphite monochromator. Structure solutions were obtained as described elsewhere (Rutenber, et al., 1993).



Mammalian cells and viral constructs. COS-7 and HeLaT4 cells were maintained in Dulbecco's modified Eagle medium H21 (DMEM) supplemented with 10% fetal calf serum (GIBCO, Long Island, NY), 100 U/ml penicillin, 100 μg/ml streptomycin. COS-7 cells were obtained through the American Type Culture Collection. HeLaT4 cells from Prof. R. Axel and HIV-1 $_{IIIB}$ gp160 antiserum (HT3) from Repligen were obtained through the AIDS Research and Reference Program, Division of AIDS, NIAID, NIH. HeLaT4 cells were grown in the presence of G-418 (300 μg/ml) to maintain CD4 expression.

The HIV-gpt proviral vector consists of the HIV-1 HXB2 sequences (the *env* gene, nucleotides 6402 to 7620[1] replaced with the drug selectable *E. coli* guanine phosphoribosyl transferase (*gpt*) gene) cloned into the pBS plasmid (pBluescript, Stratagene, San Deigo, CA) (Page, et al., 1990). The HXB2-env vector encodes the gp160 envelope sequences from the HIV-1 HXB2 strain (nucleotides 5999 to 8896) cloned into a simian virus (SV40) expression vector (Page, et al., 1990). These plasmids were obtained from Prof. Dan Littman.

For introduction of mutants into the provirus, site directed mutagenesis was carried out with the Bluescript (Stratagene) based phagemid pSPR, which contains a 918 bp *Bgl* II - *Eco*R V fragment (nt 1639 to 2557) spanning the protease gene. All mutants were sequenced through the protease gene prior to cloning out of pSPR. Mutants were recovered from pSPR as shown in Figure 3-3. The *Bgl* II - *Eco*R V fragment containing the mutant gene was excised and cloned into the pSpol vector, which contains a 4315 bp *Spe* I - *Sal* I fragment encompassing half of *gag* and most of *pol*. For the final step, the *Spe* I - *Sal* I fragment was cloned back into HIV-gpt to generate proviruses harboring the protease mutant of interest.

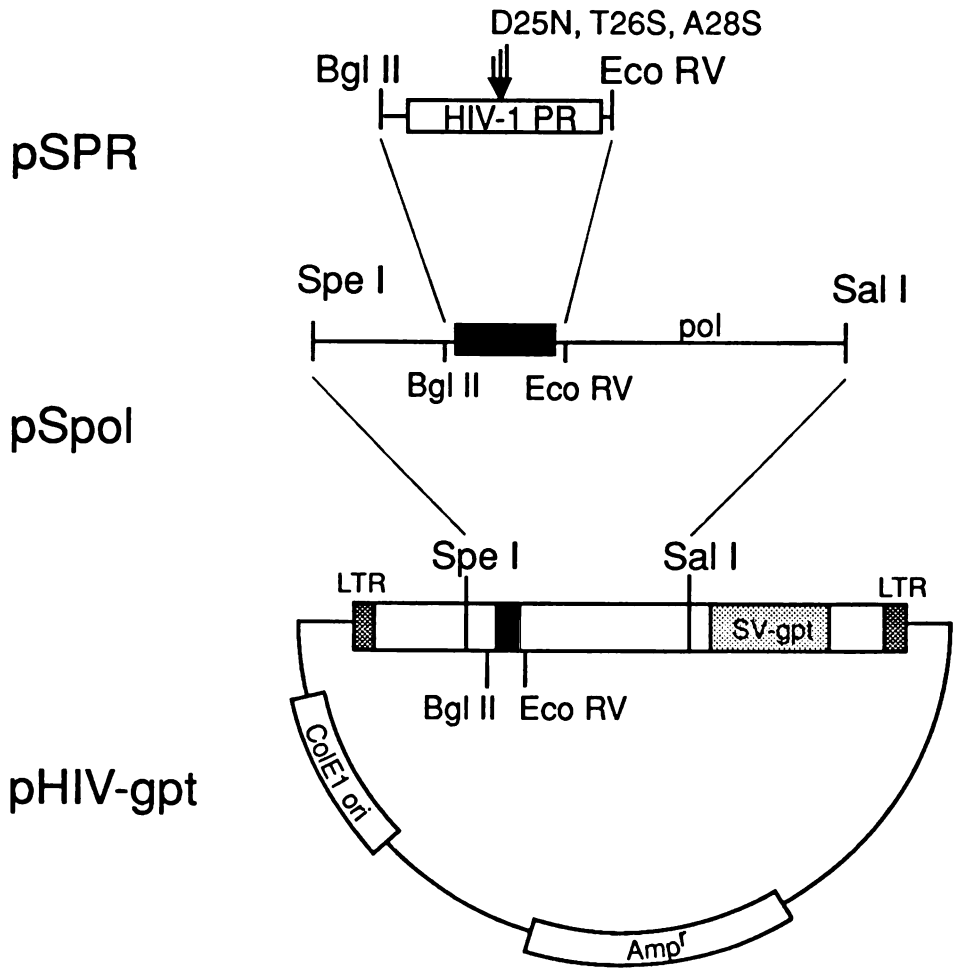
Transfection. Viral capsids were produced by transfection of COS-7 cells using the calcium phosphate procedure (Graham and van der Eb, 1973) with the following modifications. Thirty micrograms of HIV-gpt DNA was used per 10-cm dish of approximately 30% confluent cells, and 100 μM chloroquine added to increase transfection efficiency. After 14 hours, the cells were washed with phosphate buffered saline (PBS) and fresh medium was added. To produceinfectious particles, 60 μg of HXB2-env or 30 μg of pSV-A-MLV were included with the HIV-gpt DNA.

---

[1]Nucleotide designations are based on the numbering system of (Ratner, et al., 1985)

Figure 3-3: Vectors for constructing viral protease mutants.

Two shuttle vectors were built based on the pBluescript (Stratagene, San Diego, CA) backbone. The ~3.8 kb pSPR vector contains a 918 bp *Bgl* II - *Eco*R V fragment (nt 1639 to 2557) spanning the protease gene and was used for generating single stranded DNA. The second shuttle vector pSpol contains the 4315 bp *Spe* I - *Sal* I (nt 1051 to 5366) fragment of the HIV-1 HXB2 clone. The final viral vector, pHIV-gpt, contains the HIV-1 HXB2 clone described in (Ratner, et al., 1985) with a substitution of the envelope gene by an SV-40 driven gpt gene. The *Spe*-I - *Sa*l I fragment was then cloned back into HIV-gpt to generate mutant proviral vectors.


pSPR
D25N, T26S, A28S
Bgl II
Eco RV
HIV-1 PR
pSpol
Spe I
Sal I
pol
Bgl II
Eco RV
pHIV-gpt
Spe I
Sal I
LTR
LTR
SV-gpt
Bgl II
Eco RV
ColE1 ori
Amp^r

Infection. For determining viral titer, monolayers of HeLaT4 cells were seeded at 15% confluence in 6-well plates 24 hours before infection. Transfected cell supernatants were filtered through 0.45 μm cellulose acetate filters (Corning Glass Works, Corning, N.Y.) and diluted serially (final volume 750 μl) prior to addition to wells. After a one hour absorption at 37°C, 1 ml of fresh medium was added per well. At 18-24 hours post-infection, the medium was replaced with DMEM containing 50 μg/ml mycophenolic acid, 250 μg/ml hypoxanthine, and 14 μg/ml xanthine in addition to 10% fetal calf serum and antibiotics. The medium was changed every three or four days until drug resistant colonies appeared (10-14 days). Colonies were fixed and stained with PBS containing 1% formalin, 50% ethanol, and 0.5% crystal violet.

Isolation and Purification of Virions. Transfected cell supernatants were filtered through 0.45 μm cellulose acetate filters to remove cellular debris. Sucrose cushions in PBS were prepared by layering 4 ml of 20% (wt/vol) sucrose over 75 μl of 60% sucrose. Particles were concentrated from 10 ml of cleared supernatant by centrifugation at 220,000 x g for 1.5 hours. The bottom 500 μl of the cushion was collected and stored at -20°C.

Quantitation of p24. In order to determine particle-associated p24 levels, pelleted virions were diluted in PBS to correspond to dilutions of 1:10 to 1:500 of the initial supernatant. These samples were assayed using an enzyme-linked immunoassay kit (DuPont NEN, Wilmington, DE) which is sensitive to the processed forms of the major capsid antigen (p24 and p25) but not the $p55^{gag}$ or $p160^{gag\text{-}pol}$ precursors. For determination of relative particle numbers, 50 μl of purified capsids were diluted to 200 μl in HIV-1 protease activity buffer (50 mM

NaOAc, pH 5.5, 1 M NaCl, 1 mM DTT, 1 mM EDTA) containing 0.1% Triton X-100, and vortexed to disrupt capsids. The samples were processed *in vitro* by incubation at 37°C following the addition of 100 ng of recombinant HIV-1 protease. The total amount of p24 derived from all precursors was assayed by ELISA at various dilutions as described previously.

Metabolic Labeling of Cultures. For labeling of transient expression products, cells were grown for approximately 48 hours after transfection. Cells were starved for 20 minutes in DMEM deficient in methionine, supplemented with 2% dialyzed fetal calf serum. Cells were labeled for six hours using 0.1 mCi/ml of [^{35}S] Translabel (ICN, 70% methionine/30% cysteine, 1000 Ci/mmol specific activity) in 4 ml of DMEM deficient in methionine and supplemented with 2% dialyzed fetal calf serum. Chases were performed with DMEM + 10% fetal calf serum. Supernatants were collected and capsids purified as described previously.

RNA analysis. Capsids purified by sucrose cushion sedimentation were mixed with an equal volume of denaturation solution (0.4% SDS, 4 M formaldehyde, 50% formamide, 40 mM MOPS, pH 7) and heated to 65°C for 15 minutes. The samples were applied to nitrocellulose using a slot blot apparatus (Schleicher and Schuell, Keene, NH). The hybridization probe was synthesized by random priming from a 922 bp Spe I - Bcl I fragment spanning part of *gag* and half of the *pr* gene (nt 1051 - 2009). Probe specific activity was 5 x 10^5 cpm/µl. Hybridization was performed at 50°C.

Virion stability studies. Viruses were harvested at 50 hours post-transfection and purified over sucrose cushions. Samples were adjusted to 50 ng of total gag protein (assessed by ELISA) and incubated at 37°C for one hour in PBS with or without 0.5% Triton X-100 and 2.5 mM EDTA. Following treatment, the samples were layered over 3.5 ml cushions of 20% sucrose containing the same detergent and EDTA conditions as the experimental treatment, and centrifuged for one hour at 170,000 x g. Fractions of 500 µl from the top and the bottom of the gradient were collected, precipitated with TCA, and separated by SDS-PAGE on 12.5% acrylamide gels. Proteins were transferred to 0.2 µm nitrocellulose and visualized by immunoblotting with a primary rabbit polyclonal anti-p24 antibody (American BioTechnologies), followed by a secondary horseradish peroxidase-conjugated goat antibody to rabbit immunoglobulin-G (Pierce, Rockford, IL).

Reverse transcriptase assays. Reverse transcriptase activity was determined using the RT-Detect™ kit from DuPont NEN. Briefly, virions were purified directly from media by centrifugation at 100,000 x g for two hours. Virions were lysed with 0.3% Triton X-100 and the endogenous reverse transcriptase was used to extend a synthetic RNA template. Following alkaline hydrolysis of the RNA template, the cDNA was detected by sandwich hybridization. A biotinylated capture probe complementary to one region of the cDNA was used to immobilize synthesized strands, which were detected by a second probe, complementary to a different region of the cDNA, and conjugated to horseradish peroxidase. Detection was performed with 3,3',5,5'-tetramethylbenzidine (TMB). For comparison of results, the lysed virions were digested with HIV-1 protease

and p55 quantitated as described previously. The amount of cDNA synthesized was normalized to the amount of p55 present in the viral samples.

Results

Kinetic characterization of mutants. The Q7K, Q7K/T26S, Q7K/A28S, and Q7K/D25N protease variants were expressed using the bicistronic system previously described. The presence of the Q7K mutation caused an increase in pI of the protease, allowing the variants to be purified using a two column procedure which yielded approximately six to eight milligrams of purified protease from 10 grams of *E. coli*. Purification of the Q7K/D25N mutant was less efficient due to poor binding to the Pepstatin-A Agarose column, and protein concentration was estimated using Western blots against known amounts of wild type HIV-1 PR.

The kinetic parameters of the enzymes were determined using the peptide NH_2-ATLNFPISPW-COOH, which represents the cleavage site at the protease-reverse transcriptase junction. The results are shown in Table 3-2. It was apparent that only k_{cat} had been affected by the A28S and T26S mutations, being approximately 50- and 4-fold lower, respectively. The decrease in k_{cat} observed for the Q7K/A28S variant was similar to that observed by Ido, et al. (Ido, et al., 1991). No activity was detected with the Q7K/D25N PR, even following an overnight incubation with 1 μg of protease.

X-ray crystallography. Crystal structures of the Q7K/T26S and Q7K/A28S mutants complexed with the small molecule inhibitor UCSF-8 were compared with the Q7K-UCSF-8 complex determined previously (Rutenber, et al., 1993).

Table 3-2. Kinetics of wild type and mutant proteases

mutant	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\mu M^{-1}$)
WT	820 ± 20	990 ± 50	0.82 ± 0.04
Q7K	1030 ± 30	1300 ± 60	0.81 ± 0.03
Q7K,T26S	240 ± 10	900 ± 60	0.26 ± 0.02
Q7K,A28S	17 ± 2	1100 ± 140	0.015 ± 0.003
Q7K,D25N	0	0	0

k_{cat} and K_M were determined using the substrate NH2-ATLNFPISPW-COOH in standard assay buffer (50 mM NaOAc, pH 5.5, 1 M NaCl, 1 mM EDTA, 1 mM DTT).

Crystals were obtained for HIV-1 Q7K PR and HIV-1 Q7K/T26S PR which had the general morphology of orthorhombic plates and an average size of 600 x 250 x 100 μm (Figure 3-4A). Crystals of HIV-1 Q7K/A28S contained a mixture of orthorhombic plates and tetragonal bipyramids with an average size of 150 x 50 μm (Figure 3-4B). The tetragonal bipyramid crystals have been observed before for co-crystals of UCSF-8 with the wild type HIV-1 protease (Rutenber, et al., 1993) and indicate a conformation in which the inhibitor UCSF-8 lies in the peptide binding pocket perpendicular to the peptide binding axis. At the size obtained, these crystals did not diffract nearly as well as the orthorhombic plates and were not used for data collection.

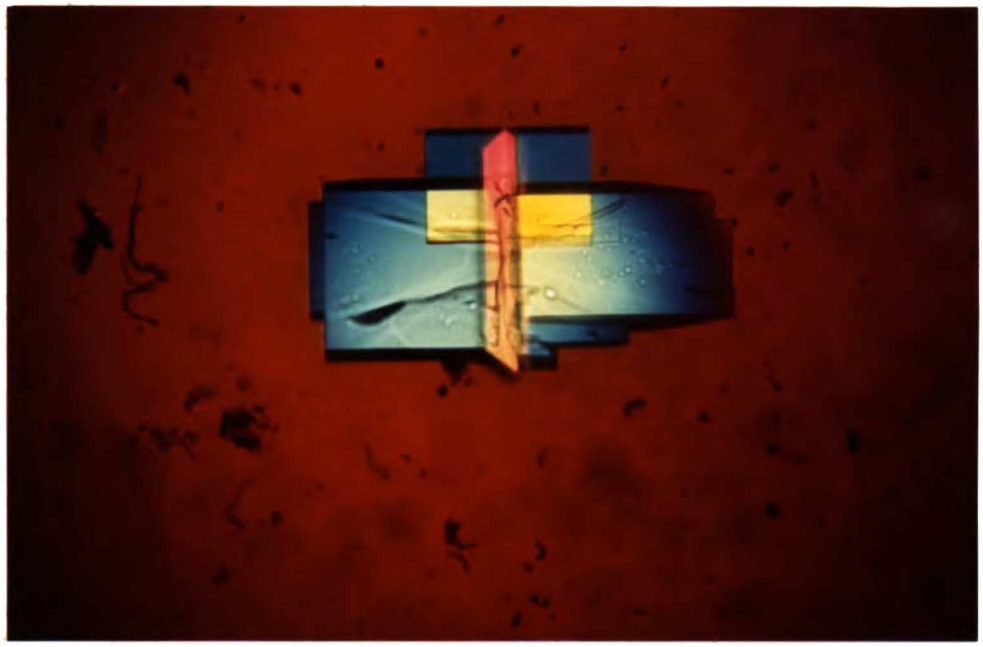
The statistics obtained on data collected for the crystal complexes are shown in Table 3-3. The structures of the Q7K and variant complexes were determined to resolutions ranging from 1.9 to 2.2 Å, and were refined to R-factors of approximately 18%. The only discernible deviation between the Q7K/T26S PR structure and the Q7K structure was the absence of the Cγ methyl groups at the site of the mutation, which showed clearly as negative density in an Fo-Fc map (Figure 3-5). Although some slight motion may have occurred in the side chains of Leu5, Leu95, and the related residues from the other monomer (Leu5', Leu95'), it was very slight (~0.3Å) and well within the margin of error for the structures. No discernible differences occurred anywhere in the structure outside of this region, confirming that the mutation had no effect on the binding pocket.

The Q7K/A28S complex structure also displayed no significant deviations in the binding pocket of the enzyme. The position of the side chain hydroxyl of Ser28 was found to be oriented away from the catalytic aspartate ("up"), towards a polar pocket containing the amide nitrogen, carbonyl oxygen, and one Oδ of Asp30 (Figure 3-6). Although this appears to conflict with previous conclusions

Figure 3-4: Crystals obtained with mutant proteases.

(A). Crystal obtained with HIV-1 Q7K PR + UCSF-8. Protein concentration was approximately 2.5 mg/ml in the presence of 200 μM UCSF-8. Dimensions are approximately 550 x 250 x 100 μm. (B). Crystals obtained with HIV-1 Q7K/A28S PR + UCSF-8. Protein concentration was approximately 2 mg/ml in the presence of 200 μM UCSF-8. Bipyramids are approximately 150 x 50 μm.

A.



B.

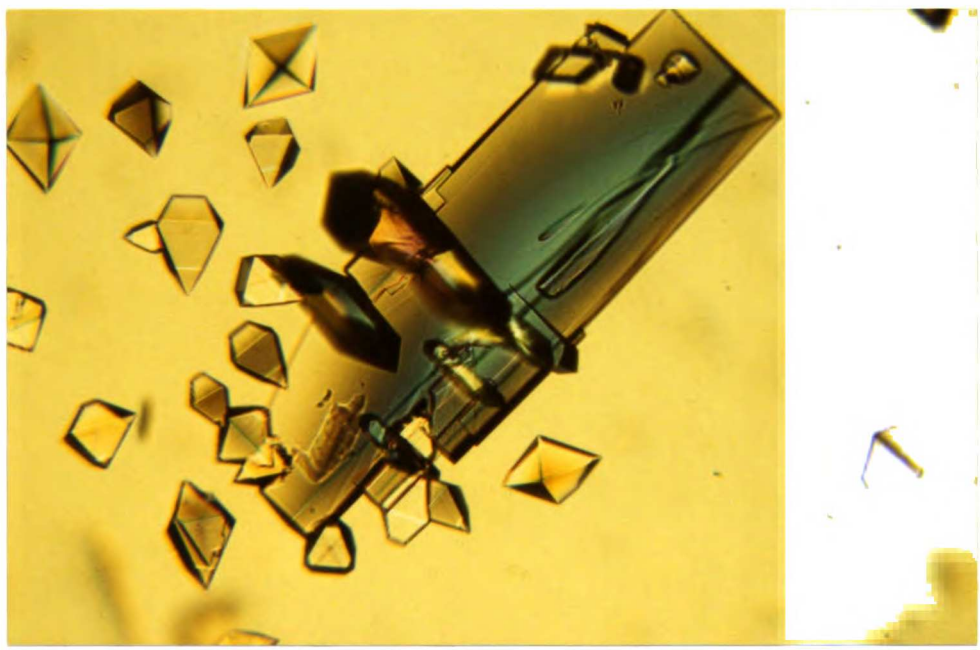


Table 3-3. Summary of Crystallization Results for Wild Type and Mutant Proteases

Protein: 2.5 - 3.5 mg/mL protease + 200 μM UCSF-8

in 50 mM NaOAc, pH 5.5, 50 mM NaCl, 1 mM EDTA, 1mM DTT

Precipitant: 50 mM NaOAc, pH 5.5, 1 M NaCl, 1 mM EDTA, DTT

Crystallographic data				
protease:	WT	Q7K	Q7K/T26S	Q7K/A28S
Space Group	$P4_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	50.2, 50.2, 100.4	51.5, 60.4, 62.6	51.5, 60.4, 62.6	51.5, 60.4, 62.6
Resolution (Å)	∞ - 2.2	∞ - 1.9	∞ - 2.1	∞ - 2.2
% completeness	93	77	85	~80
Refinement				
Resolution (Å)	7.0 - 2.2	7.0 - 1.9	7.0 - 2.1	7.0 - 2.2
R_{cryst} refined	17.4	17.9	18.0	< 19

Figure 3-5: Density map of HIV-1 Q7K/T26S PR-UCSF-8 complex.

Superimposition of the chain tracings for the HIV-1 PR Q7K structure (yellow) and the HIV-1 PR Q7K/T26S (in green). Density shown in purple represents a 2Fo-Fc map contoured at 1 sigma. The absence of density for the γ-methyl group in the T26S mutant is clearly observed by the protusion of the green side chain out of the calculated density. Slight motions are also observed at the ends of the four labeled leucine residues 5, 97, 105 (5'), and 197 (97') which abut the site of mutation.

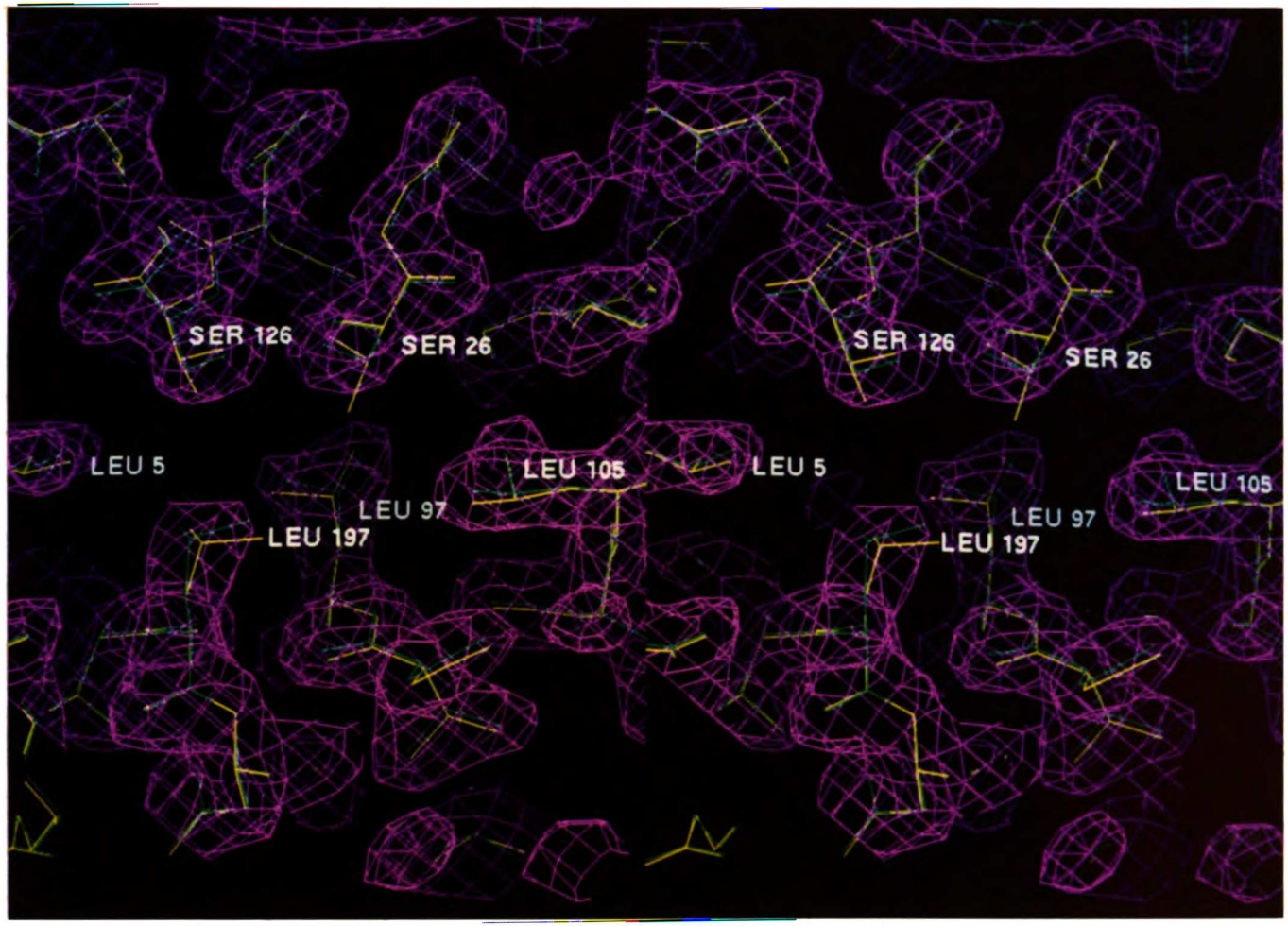
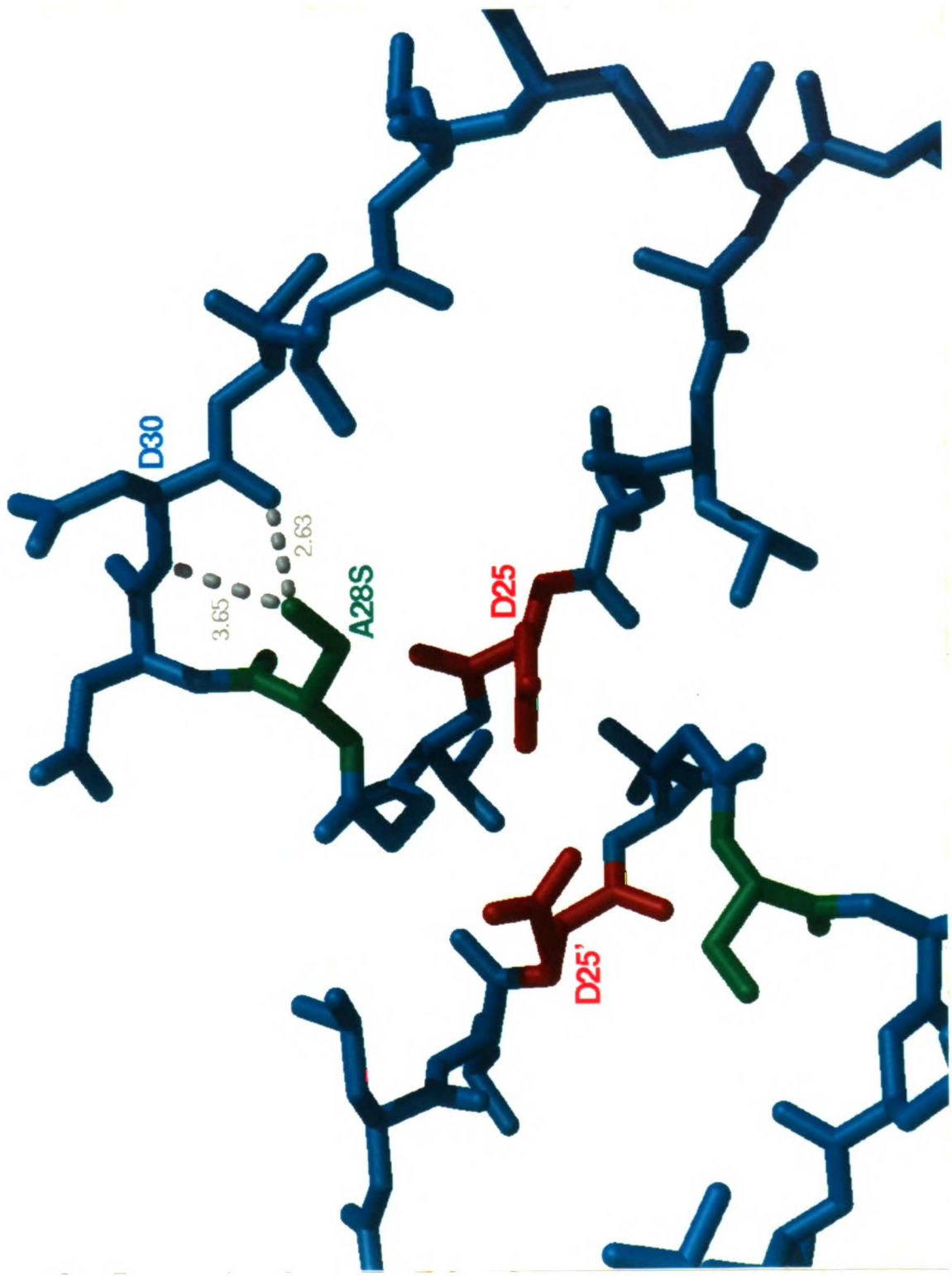

SER 126
SER 26
LEU 5
LEU 105
LEU 97
LEU 197
SER 126
SER 26
LEU 5
LEU 105
LEU 97
LEU 197

Figure 3-6: Active site model of HIV-1 Q7K/A28S PR-UCSF-8 complex.

Chain tracing for the HIV-1 PR Q7K/A28S structure showing the mutant side chain A28S in green. The active site aspartic acids are shown in red. Distances shown are those between the side chain hydroxyl (Oγ) and either the Asp30 amide nitrogen (3.65 Å), or carbonyl oxygen (2.63 Å).


D30
2.63
3.65
A28S
D25
D25'

that the catalytic aspartate and A28S hydroxyl interact, those conclusions had been reached using a model in which substrate was bound to the enzyme (Ido, et al., 1991). In the presence of a substrate, additional interactions may stabilize the Ser28 hydroxyl in the "down" orientation.

Effect of mutants on polyprotein processing. To assess the effect of reduced proteolytic activity on viral polyprotein processing, the mutants were introduced into a tissue culture system for analyzing HIV virion assembly. Protease genes containing the three single mutations T26S, A28S, and D25N were cloned into the HIV-gpt vector. Following transfection in COS-7 cells, capsids were labeled with ^{35}S-methionine and purified over sucrose cushions. The presence of the T26S, A28S, and D25N protease mutations had a dramatic effect on the degree to which viral polyproteins were processed (Figure 3-7). After a 12 hour pulse-chase, wild type virions contained predominantly mature capsid protein (p24), with a small amount of a less mature form (p25). In T26S PR capsids, the capsid protein existed largely in the p25 form, and additional precursors were visible at 39 and 55 kDa. The A28S PR and D25N PR capsids did not show any significant processing, and contained immature p55^{gag} precursor almost exclusively. Small amounts of processing intermediates could be observed in A28S PR samples only after extended periods of time.

Longer time periods were necessary to examine the ability of the less active proteases to process viral capsids. Accumulation of mature capsid protein was followed over 130 hours by p24-ELISA of particle-associated material (Figure 3-8A). Although the T26S PR particles display a short, 24 hour lag relative to WT particles, significant levels of p24 are present at later time points. Very low levels

Figure 3-7: Contents of viral capsids.

Cells were transfected with 30 μg of plasmid DNAs and grown for approximately 48 hours after transfection. For labeling, cells were starved for 20 minutes in DMEM deficient in methionine, supplemented with 2% dialyzed fetal calf serum. Cells were labeled for 6 hours, followed by a 6 hour cold chase in complete media. Capsids were isolated by sucrose cushion sedimentation as described and precipitated by treatment with trichloroacetic acid. Proteins were analyzed by 12.5% SDS-polyacrylamide gel electrophoresis and visualized by autoradiography.

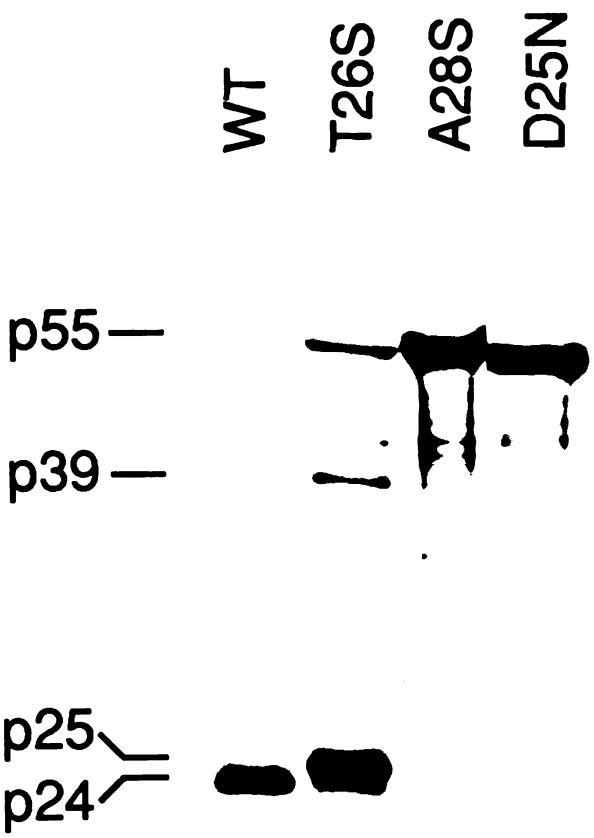
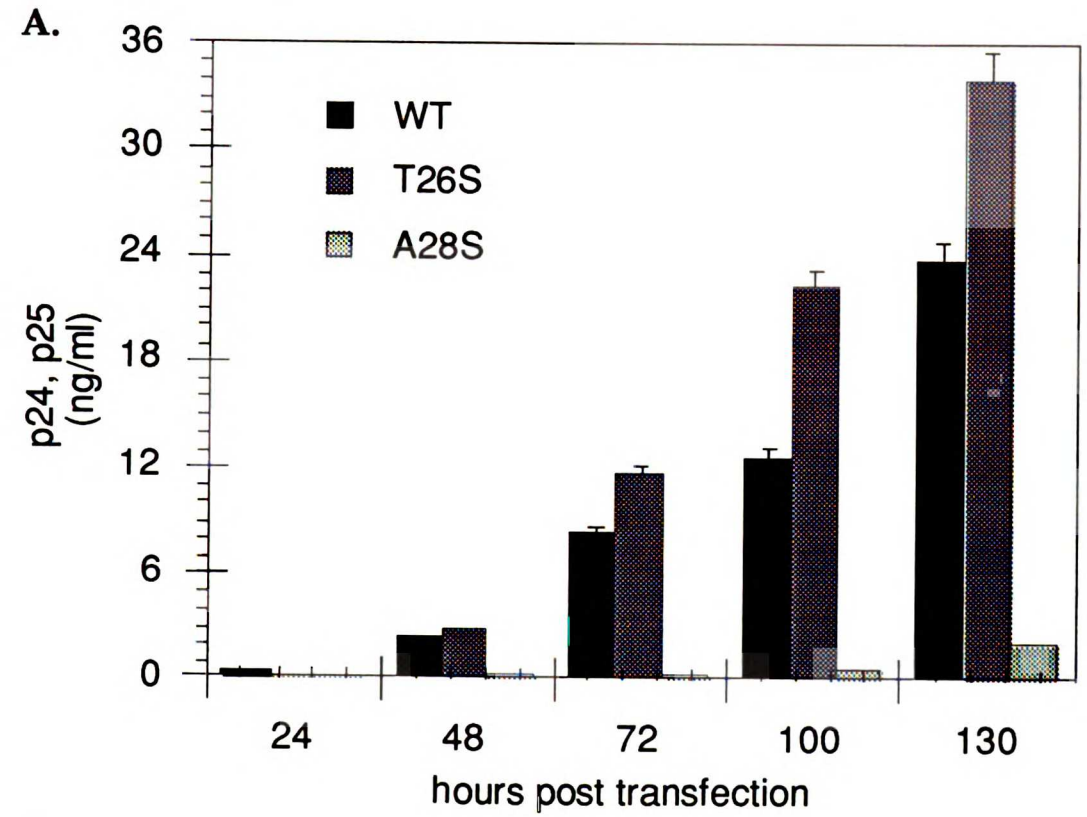
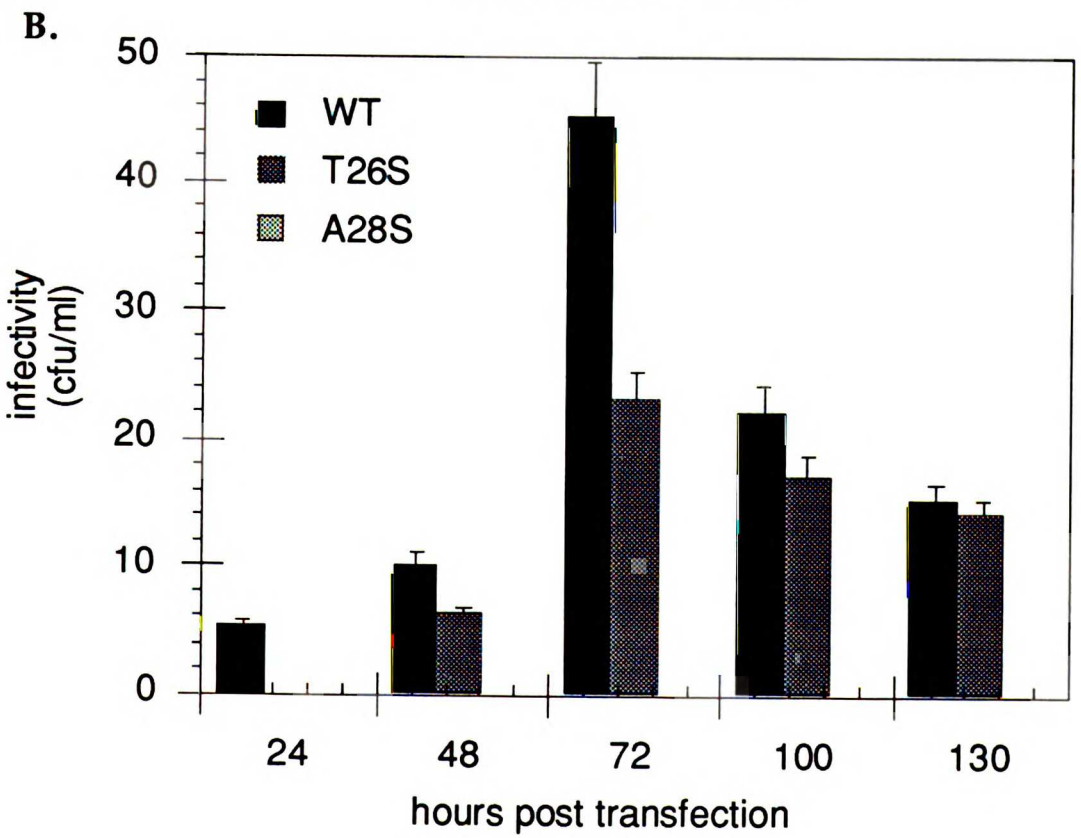

WT
T26S
A28S
D25N
p55
p39
p25
p24

Figure 3-8: Extended time course of mutant capsid processing.

COS-7 cells were transfected with 30 μg of plasmid DNA and culture supernatants harvested at 24, 48, 72, 100, and 130 hours post transfection. (A) Purified viral particles were disrupted by vortexing with 0.1% Triton X-100 and p24 levels of three serial dilutions measured by ELISA (DuPont NEN, Wilmington, DE). (B) Supernatants were filtered through 0.45 micron filters and used to infect HeLaT4$^+$ cells for determination of viral titer.


A.
36
30
24
18
12
6
0
p24, p25 (ng/ml)
WT
T26S
A28S
24
48
72
100
130
hours post transfection


B.
50
40
30
20
10
0
infectivity (cfu/ml)
WT
T26S
A28S
24
48
72
100
130
hours post transfection

of p24 are present in cells transfected with the A28S PR clone, indicating that processing of polyprotein precursors is dramatically diminished even after 130 hours. The order of appearance of gag processing intermediates in the T26S PR and A28S PR capsids over longer time periods was similar to the order in which they have been observed in other systems (Erickson-Viitanen, et al., 1989; Gowda, et al., 1989; Kaplan and Swanstrom, 1991), suggesting that selection of cleavage sites had not been altered in mutant capsids.

Effect of mutants on viral infectivity. The effect of altering the kinetics of polyprotein processing on viral infectivity was analyzed. Infectious particles were generated by co-transfecting the wild type and mutant-bearing HIV-gpt vectors with an HIV-1 HXB2 envelope-expressing plasmid, HXB2 env. Infectivities of samples from four independent transfections were analyzed at 48 hours post-transfection and gave the results shown in Table 3-4. At a point where p24 levels were approximately equal between WT and T26S PR virions, the T26S PR clones were consistently three to four-fold less infectious than wild type virions. A28S PR and D25N PR clones produced no infectious particles at this time point or later time points up to 130 hours (Figure 3-8B). It was observed that infectious titers of WT and T26S PR virions declined after 72 hours despite a continuous increase in p24 levels, thus limiting the use of the system to time points at 72 hours or earlier.

Effect of mutants on viral assembly. To determine if the protease mutations affected other aspects of the assembly process, RNA encapsidation and envelope association were analyzed. Samples of purified particles were normalized for the amount of p55 present, and dot blots performed with an HIV-1 specific probe to

Table 3-4. Infectivity of mutant viruses

	HeLaT4 colonies/ml/unit p24			
experiment	WT	T26S	A28S	D25N
#1	400 ± 15	131 ± 7	0	0
#2	176 ± 8	64 ± 5	0	0
#3	121 ± 11	31 ± 5	0	0
#4	106 ± 5	40 ± 3	0	0

For normalization of infectious particle data to total particle levels, purified particles were disrupted with 0.1% Triton X-100, 5 mM EDTA and digested with exogenous purified HIV-1 protease to convert all gag proteins to their processed forms. Total p24 in digested samples was determined by ELISA, and the colony numbers divided by the ratio of mutant to wild type p24 levels in order to generate HeLaT4 colonies/ml/unit p24.

measure RNA levels. Association of the envelope glycoprotein was assessed by separating purified virions on 4-12% gradient SDS-polyacrylamide gels and immunoblotting with anti-gp160 antibodies. The results, shown in Figure 3-9, indicate that levels of gp120/gp160 and viral RNA correlated well with the amount of total capsid antigen present. The levels of RNA present at 96 hours post transfection were higher than those observed at 72 hours post transfection, which agreed with the observation that particles continued to accumulate at longer time points (Figure 3-8A). This semi-quantitative analysis indicates that there are no defects in assembly or budding associated with the presence of the A28S and T26S mutant proteases.

Effect of mutants on virion stability. Since the previous data suggested that the viruses contained all the components necessary for replication, we surmised that inefficient uncoating of the viral nucleoprotein complex may be occurring in unprocessed particles. Thus, in particles which were processed to a lesser extent (D25N PR, A28S PR), the viral core would be expected to be more stable than in wild type virions. To examine relative virion stability, wild type and mutant virions which had been purified over sucrose cushions were treated with 0.5% Triton at 37°C in the presence of 2.5 mM EDTA. Treated and untreated samples were purified a second time by centrifugation through 20% sucrose (Figure 3-10). As has been observed for HIV and ALV capsids, wild type virions were totally disrupted by detergent treatment, whereas immature virions were unaffected (Park and Morrow, 1993; Stewart, et al., 1990). T26S PR capsids, which appeared to be 85-90% processed were similar to wild type virions in that they were completely disrupted by this treatment. Treated samples from A28S PR showed some dissociation, in that a small amount of p24/25 and p39/41 were found at

Figure 3-9: Envelope and RNA composition of mutant virions.

Viral particles were isolated from culture supernatants by sucrose cushion sedimentation at 72 and 96 hours post transfection. (A). Samples of purified virions were TCA precipitated and resuspended in loading buffer. Proteins were separated on 4-12% SDS-PAGE gels. After electroblotting onto 0.45 μm nitrocellulose, envelope proteins were detected by immunoblotting with sheep polyclonal antiserum to gp160. Lane 1, cell lysate from HTLVIII chronically infected CEM cells containing both gp160 and processed gp120; M, 97.4 kDa size marker; lanes 2, 5, WT virions, lanes 3, 6, T26S virions, lanes 4, 7, A28S virions. (B). RNA dot blot analysis of 10 and 50 μl samples of purified virions. Lane 1, no sample; lanes 2, 5, WT virions, lanes 3, 6, T26S virions, lanes 4, 7, A28S virions.

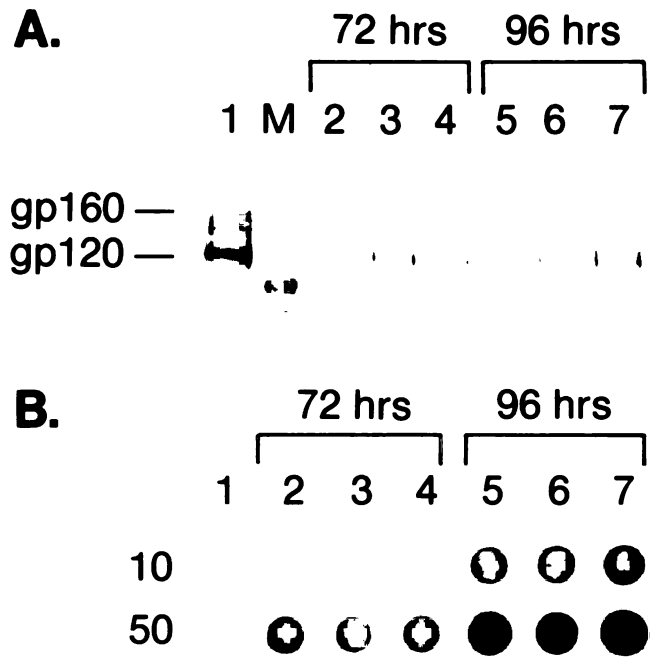

A.
72 hrs
96 hrs
1 M 2 3 4 5 6 7
gp160 —
gp120 —
B.
72 hrs
96 hrs
1 2 3 4 5 6 7
10
50

Figure 3-10: Virion stability to detergent disruption.

Viral particles were incubated at 37°C in the presence or absence of 0.5% Triton X-100 and 2.5 mM EDTA. Samples were loaded on 20% sucrose gradients in detergent and EDTA conditions which matched those of the incubation. Particles were purified again by centrifugation at 170,000 x g for 1.5 hours. Samples from the top (T) and bottom (B) of the gradients were TCA precipitated and separated by 12.5 % SDS-PAGE. Following transfer to 0.2 μm nitrocellulose proteins were visualized by p24 immunoblot. Lanes 1, 2, 5, 6, 9, 10, 13, and 14, no detergent. Lanes 3, 4, 7, 8, 11, 12, 15, and 16, 0.5% Triton X-100+ 2.5 mM EDTA.

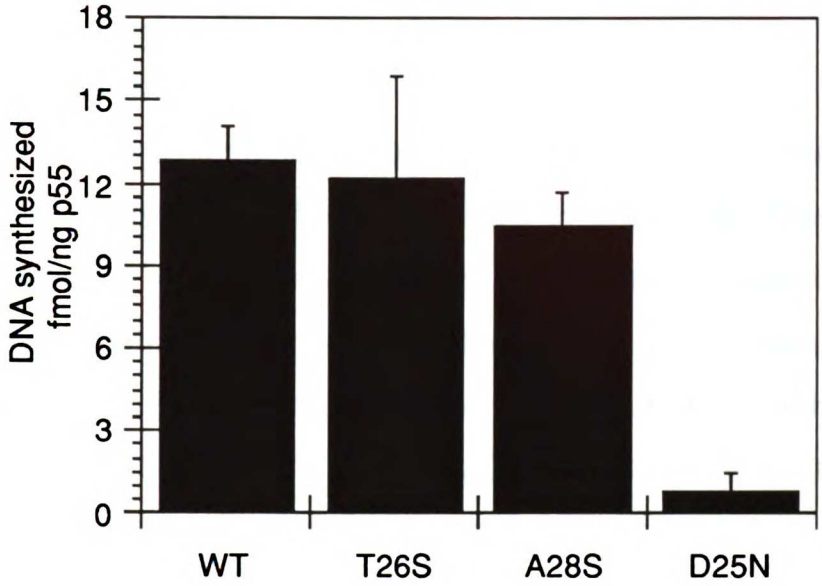

WT
T26S
A28S
D25N
− + − + − + − +
T B T B T B T B T B T B T B T B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
p55
p41
p39
p25
p24

the top of the gradient, while all of the protein precursors remained pelleted at the bottom of the cushion. This suggests that the A28S PR material contained a small number of particles which were sufficiently processed to allow disruption, while the majority of the particles were immature. The D25N PR particles, by contrast, were found to be completely resistant to detergent disruption under these conditions.

Reverse transcriptase activity in mutant virions. To characterize the degree of activation of the reverse transcriptase enzyme, virions were harvested at 56 hours post transfection and purified directly from tissue culture supernatants by centrifugation at 100,000 x g for two hours. Virion samples were assayed for p55 content by ELISA, and reverse transcriptase activity was quantitated using a non-radioactive polymerase assay (DuPont NEN). As shown in Figure 3-11, there was an approximately 10-fold difference observed in the level of RT activity in wild type versus D25N PR virions. T26S PR virions contained approximately the same amount of RT activity as wild type virions, whereas levels in A28S PR virions were only 75 - 80% of that observed in wild type virions.

Figure 3-11. Reverse transcriptase activity in mutant virions.

Virions were purified directly from media by centrifugation at 100,000 x g for 2 hours and cDNA synthesized using the endogenous reverse transcriptase. A biotinylated capture probe complementary to one region of the cDNA was used to immobilize synthesized strands, which were detected by a second probe, complementary to a different region of the cDNA, and conjugated to horseradish peroxidase. Detection was performed with 3,3',5,5'-tetramethylbenzidine (TMB). For comparison of results, the amount of cDNA synthesized was normalized to the amount of p55 present in the viral samples, as determined by ELISA.


DNA synthesized
fmol/ng p55
18
15
12
9
6
3
0
WT
T26S
A28S
D25N

Discussion

In vitro structure-function analysis.

A limited number of mutations can be made which affect only the catalytic activity of the HIV-1 protease. Since we wished to isolate the activity of the protease (k_{cat}) from its substrate specificity (K_M), we were restricted to the two mutants described here: T26S and A28S. The T26S mutation was made to correspond to a residue substitution found in the RSV protease, which has a lower specific activity than the HIV-1 protease (Grinde, et al., 1992). It has been reported that the reverse mutation in RSV, S38T, increased the overall activity of the enzyme (Moshe Kotler, unpublished results). The T26S mutation in HIV-1 protease lowered the specific activity of the enzyme by 4-fold, within the range of the RSV protease. Crystallographic analysis indicates the presence of two cavities in the variant enzyme at the site of mutation. These might be predicted to have a detrimental effect on the neighboring catalytic aspartic acid residues.

Our kinetic results with the A28S mutation confirm the observations of Ido, *et al.* that k_{cat} is decreased approximately 50-fold. The reason for this loss of activity is not clear from our crystal structure, since the serine hydroxyl is pointing away from the catalytic residues. A possible answer to this dilemma lies in the behavior of the Q7K/A28S variant enzyme during crystallization trials. Tetragonal bipyramids dominated the crystallization drops after 18 hours, but after three days large rhombohedral plates appeared in several drops. In the presence of the UCSF-8 inhibitor these two different crystal forms are associated with two different orientations of the inhibitor in the protease active site: perpendicular to the peptide binding axis in pyramidal crystals, and parallel to the peptide binding axis in the plate form (Rutenber, et al., 1993). It is possible

that the presence of the serine at position 28 interferes with the inhibitor binding in the peptide-like mode, thus leading to an increased frequency of crystals forming in a lattice which favored the alternative binding mode. The plate-form crystals which were diffracted show the inhibitor lying in the active site in an orientation that could prevent the positioning of the serine 28 side chain near the catalytic aspartic acids. Solution of a structure with a peptidomimetic inhibitor in the active site would help to clarify this issue and more clearly define the interactions of Ser28 with the catalytic machinery.

Virion maturation

The presence of the protease mutants had a dramatic effect on the rate at which viral proteins were processed following the assembly of virions. In pulse-chase studies of viral capsids, the wild type protease was able to convert p55 to p24 in as little as ten minutes, whereas even after 6 hours capsids harboring T26S PR contained a mixture of gag precursors. The conversion of p25 to p24 appeared highly inefficient, although the ratio of p24 to p25 increased steadily as the chase was lengthened from 2 to 12 hours. This agrees with observations that the p25 to p24 processing step is the slowest cleavage in the maturation process (Steve Petit and Ronald Swanstrom, personal communication), and would therefore be expected to show the greatest decrease in processing rate with the mutant proteases. Over a 12 hour time course no processing was evident in either the D25N or A28S protease mutant capsids. Using both Western blots and p24 ELISA to follow processing over a 130 hour time course, it was apparent that the A28S protease mutant was capable of producing mature p24, but at a significantly decreased rate relative to both the wild type and T26S proteases.

The magnitude of the processing defect in virions was approximately correlated with the diminished k_{cat} values measured *in vitro*.

The infectivity of viral particles harboring the protease mutations was also measurably affected. Virions containing T26S PR were not infectious at 24 hours, and were consistently three to four-fold less infectious than wild type virions at 48 hours. The initial lag in the production of infectious T26S PR virions is presumably due to the increased time of processing that is required in the presence of the catalytically impaired protease. Infectivity of T26S PR virions remained lower than wild type virions at time points beyond 48 hours. In contrast, viruses containing A28S PR did not produce any infectious particles at 48 hours, an observation which is in keeping with our data showing only minimal p24 production at 48 this time. Although significant levels of p24 were present at later time points, the general decrease in particle infectivity after 72 hours might have contributed to the inability to detect infectious particles.

Our data suggest that there is a strong correlation between the activity of the protease, as measured by the efficiency constant (k_{cat}/K_M), and the infectivity of the virus. There is a clear effect of the mutants on both the rate of polyprotein processing, as determined by quantitation of particle associated p24 in transfected cell supernatants, and the rate at which infectious particles are produced. The T26S PR virions, in which only 25% of the total wild type protease activity is present, are still able to infect cells, despite a decreased rate of processing. Apparently this level of protease activity is sufficient to process the viral proteins within a time in which the virus is potentially viable. In the case of the A28S PR capsid samples, processing proceeds at a 50-fold slower rate, and no infectious particles are observed.

The data suggest that a four-fold reduction in proteolytic activity still permits polyprotein processing within a time in which the virus is viable. However, when processing proceeds at a 50-fold slower rate the virus does not mature within this window of viability and infectious virions are not produced. This window may be defined by other aspects of the viral aging process, such as genomic RNA dimerization (Fu, et al., 1994) or envelope stability (McKeating, et al., 1991). Our data show that infectious particle numbers drop markedly after 72 hours, despite a continual rise in total particle production, suggesting that instability of gp120 limits the lifetime of viruses in tissue culture. Pseudotyping A28S PR virions with a more stable envelope glycoprotein would allow us to better assess the ability of the A28S PR to produce infectious particles at longer time points.

Virion assembly and post-entry events.

The protease mutants studied here did not appear to affect virion assembly beyond altering the degree to which viral proteins are processed. By dot-blot analysis, wild type and mutant virions contained approximately the same amount of viral RNA when compared at the same time point. The lack of effect of protease mutations on RNA packaging has been observed with RSV and MuLV (Crawford and Goff, 1985; Stewart, et al., 1990) in which virions containing inactive protease mutants package genomic RNA as efficiently as wild type viruses. The condition of the RNA in the virions may have an effect on viral infectivity. Stewart, *et al.* reported that abnormal forms of the viral RNA were present in mutant virions (Stewart, et al., 1990), and a similar defect in RNA dimerization has recently been reported for HIV (Fu, et al., 1994). The presence of the protease mutants reported here did not appear to cause a differential

association with gp160, as both A28S PR and T26S PR mutant virions contained amounts of the envelope glycoprotein in the same proportion to total virion gag protein as the wild type virions.

Since virions appeared to assemble normally, it was possible that the loss of infectivity was due to perturbation of a post-assembly event necessary for viral replication. We surmised that uncoating of viral RNA, a crucial step for freeing the nucleoprotein complex for reverse transcription, may be inefficient in the presence of unprocessed viral proteins. To address the issue of virion stability, virions containing each of the mutant proteases were examined for susceptibility to disruption by nonionic detergent. The completely unprocessed D25N PR virions were not disrupted by 0.5% Triton X-100, even after 1 hour at 37°C, whereas wild type virions were completely disrupted. A majority of the A28S PR virion population were highly stable to disruption, while a small amount of virion associated p24, p25, and p39 was found to be soluble. It is likely that these virions represent a fraction of the virions produced early after transfection which had adequate time in which to mature. The T26S PR virions, although not completely processed, were highly susceptible to disruption. This suggests that dissociation of viral cores only requires a certain degree of polyprotein processing in order to take place.

Inefficient activation of reverse transcriptase by mutant proteases may also prevent the synthesis of viral DNA following uncoating of the viral core. Reverse transcriptase activity associated with unprocessed polyproteins has been reported at levels from six to thirty-fold lower than those found in wild type virions (Göttlinger, et al., 1989; Peng, et al., 1989; Schätzl, et al., 1991). Kaplan, *et al.* demonstrated that protease inhibitor-treated viruses containing decreased amounts of processed polyproteins and reverse transcriptase activity are not able

to direct efficient viral DNA synthesis after infection (Kaplan, et al., 1993). The level of reverse transcriptase activity in mutant virions was measured. In general, the amount of RNA-directed DNA polymerase activity associated with the wild type and T26S PR virions was equivalent when corrected for total amounts of p55 present, whereas it was slightly reduced (15 - 20%) in A28S PR virions and reduced nearly 90% in D25N virions. Although the reverse transcriptase present in mutant virions was functional in this simple polymerization assay, it may not efficiently carry out the multiple steps required to synthesize integration-competent viral DNA. Further studies on these processes in the cellular context of infection would help to elucidate the exact point at which infection is blocked when the proteolytic maturation process is incomplete.

Implications for drug resistance and anti-protease therapeutic intervention.

It is clear that the rate of the maturation process can be diminished at least 4-fold without seriously affecting the ability to form infectious particles. This suggests that a reservoir of protease activity which is not crucial for the maturation process exists in an immature virion. Such a reservoir could absorb a decrease in protease activity induced by point mutations in the enzyme. Point mutants in the protease which affect the binding of peptidomimetic inhibitors by five to ten-fold have been obtained by *in vitro* selection of viruses (El-Farrash, et al., 1994; Ho, et al., 1994; Otto, et al., 1993). It has been noted in at least one case that a virus containing an escape mutation in the protease is considerably less infectious than the parent virus (El-Farrash, et al., 1994), and that protease function is impaired. Although the presence of these mutations would aid in slowing the rate of viral spread, it is also clear that second-site mutations can

occur to restore catalytic competence while maintaining resistance to inhibitors (Ho, et al., 1994). Clearly, the development of effective protease inhibitors must now take into account the ability of the enzyme to function in the presence of point mutations that may disrupt the binding of small molecule inhibitors. The effects of such mutants on the protease can be analyzed in systems similar to the one described in this chapter to provide a detailed understanding of the activity and specificity requirements of the viral maturation process.

The use of HIV-1 protease inhibitors to prevent the spread of viral infection depends upon the ability to stop the maturation process at a point where the viruses are non-infectious, and maintain the virus in this state until it is irreversibly inactivated. It has been observed that the HIV-1 protease can reactivate following diffusion of peptidomimetic inhibitors out of virions (McQuade, et al., 1990), although it is not entirely clear whether this event is sufficient to restore infectivity (Lambert, et al., 1992). Depending upon the pharmacodynamics of inhibitors used therapeutically, it is likely that a situation will exist *in vivo* where some of the virion-associated proteases will either not be inhibited or may be reactivated. It therefore may be advantageous to develop inhibitors of the protease which would irreversibly inactivate the enzyme at essential active site amino acid residues (Meek, et al., 1989; Salto, et al., 1994). This strategy may show promise in the development of effective anti-proteolytic agents.

The observation that a 50-fold reduction in protease activity is sufficient to prevent the production of infectious virus is encouraging. It predicts that total inhibition of the protease will not be necessary to block the spread of the virus in a therapeutic situation. Consequently, methods that can reduce inhibitor sensitivity to point mutations in the protease and still inhibit the majority of

protease molecules in a virion will yield effective inhibitors for clinical therapy of HIV. An alternative approach to the discovery of inhibitors of the protease is to find macromoleculer inhibitors which associate with a large surface area of the HIV-1 protease. A larger interaction surface would render inhibitor binding less sensitive to single point mutations, such as have been observed with small molecule inhibitors. In the subsequent chapter we describe a scenario where defective monomers of the HIV-1 protease may serve this purpose, and provide a novel means of inactivating viral protease activity.

Chapter 4. Suppression of Viral Maturation by Defective Protease Monomers

Portions of this chapter are being prepared for publication. Babé, L., Rosé, J.R., and Craik, C.S. Dominant Negative Inactivation of HIV-1 Protease Prevents Viral Maturation and Infectivity.

Abstract

The dimerization of HIV-1 protease monomers is a highly specific event, the timing of which is crucial for the activation of the protease and maturation of the virus. We surmised that the formation of inactive heterodimers between wild type and catalytically defective monomers could provide an effective means of blocking viral maturation. When plasmid DNAs encoding the wild type or defective D25N protease were co-transfected into cells at ratios ranging from 1:1 to 1:6, proteolytic processing was inhibited by as much as 95%, and viral infectivity by up to 85%. A second mutation was made which caused the defective monomer to be expressed at levels approximately 10-fold higher than the wild type protease. This frameshift mutation effectively increased the dosage of the defective monomer, allowing a 95% inhibition of processing to be achieved at ratios of 1:1 mutant to wild type DNA. Viral infectivity was decreased by nearly 100-fold, an effect which appeared to be due to both defects in proteolytic processing and viral assembly in the presence of the frameshift mutant protein. When cells which stably expressed the defective protease monomer were transfected with wild type viral DNA, a significant reduction in proteolytic processing and infectious viral particles was observed. These data demonstrate that defective protease monomers can inhibit protease activation *in vivo*. A gene therapy application of this effect may provide an effective means of inhibiting the spread of viruses in an HIV-infected individual.

Macromolecular inhibition of the HIV-1 protease

Through the use of engineered protease mutants with decreased catalytic activity we have determined that a 50-fold reduction in the activity of the HIV-1 protease is sufficient to block the generation of infectious virions. We have also proposed that the development of inhibitors with a reduced sensitivity to point mutations in the protease is essential to effectively address the problem of drug-resistance. In order to achieve this goal we have used an *ex vivo* model for particle assembly and maturation to investigate the possibility of using defective monomers of the HIV-1 protease as a means of disrupting protease function.

Initial experiments in this area were performed in an *in vitro* system by Babé *et al.*, in which the dimer interface was targeted as a site at which structural perturbation could cause inactivation of the protease (Babé, et al., 1991). By mixing wild type HIV-1 PR and an inactive, truncated form of the protease in which the first five amino acids had been removed (HIV-1 Δ1-5) it was possible to demonstrate concentration-dependent inactivation of protease activity. It was assumed that this loss of activity was due to the formation of inactive heterodimers through dynamic subunit exchange. The formation of inactive heterodimers in this fashion suggested an *ex vivo* experiment, in which heterodimers are formed between wild type protease and a monomer in which the active site is mutated to remove the catalytic aspartic acid residue (D25N). This strategy would leave the dimer interface intact to provide a maximal and specific interaction surface between the wild type and mutant protease monomers.

This type of experiment is described as "dominant negative", and has been used as a method for inactivating gene products *in vivo* for genetic analysis of function (Herskowitz, 1987). The use of defective gene products which can

associate with and inactivate wild type gene products has been proposed for a number of the proteins of HIV (Feinberg and Trono, 1992). Dominant negative inactivation would be most effective at a step in the viral life cycle in which disruption of protein function has drastic effects. For this reason, an extensive effort has been made to identify mutants of the viral transcriptional activators Tat and Rev which exert dominant negative effects, as reviewed by Feinberg and Trono, 1992. For the purposes of dominant negative interference at the post-transcriptional level, the protease might be viewed as an ideal target, as it presides over a series of processing events which are required for proper viral function.

The formation of inactive heterodimers between wild type and mutant protease monomers could be an effective means of inhibiting protease activity. This heterodimerization event would occur as viral polyproteins assembled at the cell membrane. The interaction between monomers is highly specific and is of high-affinity, approximately 1-10 nM (Babé and Craik, 1992; Zhang, et al., 1991). The concentration of the protease in a virion is predicted to be approximately 300 μM (Tomasselli, et al., 1991), well above the K_d for the protease dimer. This should render the dimerization process essentially irreversible in the context of an assembled virion, making the assembly of defective heterodimers an attractive prospect for inactivation of the HIV-1 protease. The inter-monomer contacts occur across a large surface area, an interaction which is unlikely to be highly sensitive to point mutations. In addition, protease activity is highly sensitive to mutations at the dimer interface (Loeb, et al., 1989), reducing the probability that point mutants in this region would be viable.

The concentration of defective monomer needed to effectively inhibit the HIV-1 PR function can be calculated based upon two assumptions. First, that the mutant and wild type protease monomers have identical association constants, and thus selection of mutant or wild type monomers during dimerization is random. Second, that subunit exchange does not occur following dimerization. A model for what would be likely to occur during viral budding in the presence of different concentrations of defective monomers is shown in Figure 4-1A.

The probability of forming inactive protease dimers in the presence of a mixture of wild type (WT) and mutant (MUT) monomers can be calculated from the equations below, in which "x" represents the mole fraction of wild type protease monomer and "1-x" represents the mole fraction of defective protease monomer:

(Equation 4-1)

$$\text{probability of WT•WT} = [\#\ \text{WT}] \times [\#\ \text{WT}] = x^2$$

$$\text{probability of WT•MUT} = 2\{[\#\ \text{WT}] \times [\#\ \text{MUT}]\} = 2(x)(1-x)$$

$$\text{probability of MUT•MUT} = [\#\ \text{MUT}] \times [\#\ \text{MUT}] = (1-x)^2$$

By graphing these equations, predictions can be made as to the concentration of a defective monomer required to inactivate the protease (Figure 4-1B). In a situation where the wild type and mutant proteases are present in equal amounts, 75% of the dimeric species formed will be either inactive MUT•MUT or WT•MUT species. At a ratio of one wild type to six mutant monomers, nearly 98% of the protease dimers would be inactive species.

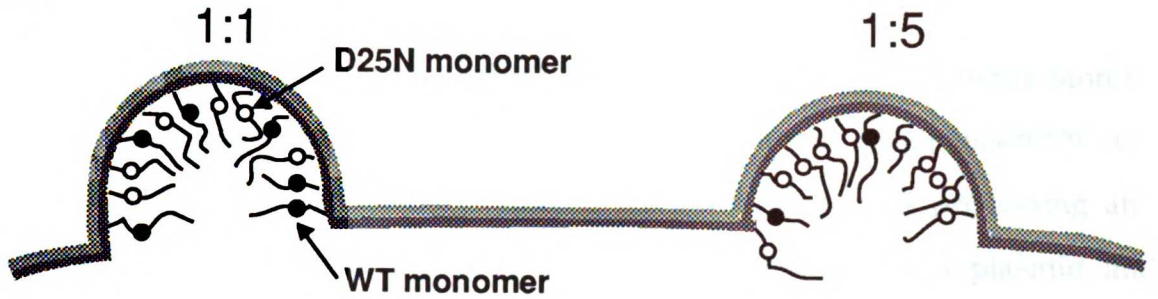
This chapter describes experiments which attempt to use HIV-1 protease heterodimer formation in cells to block viral maturation. By mixing pHIV-gpt D25N PR with pHIV-gpt in cotransfection experiments, we were able to observe

Figure 4-1. Model of wild type and mutant heterodimer formation during viral assembly.

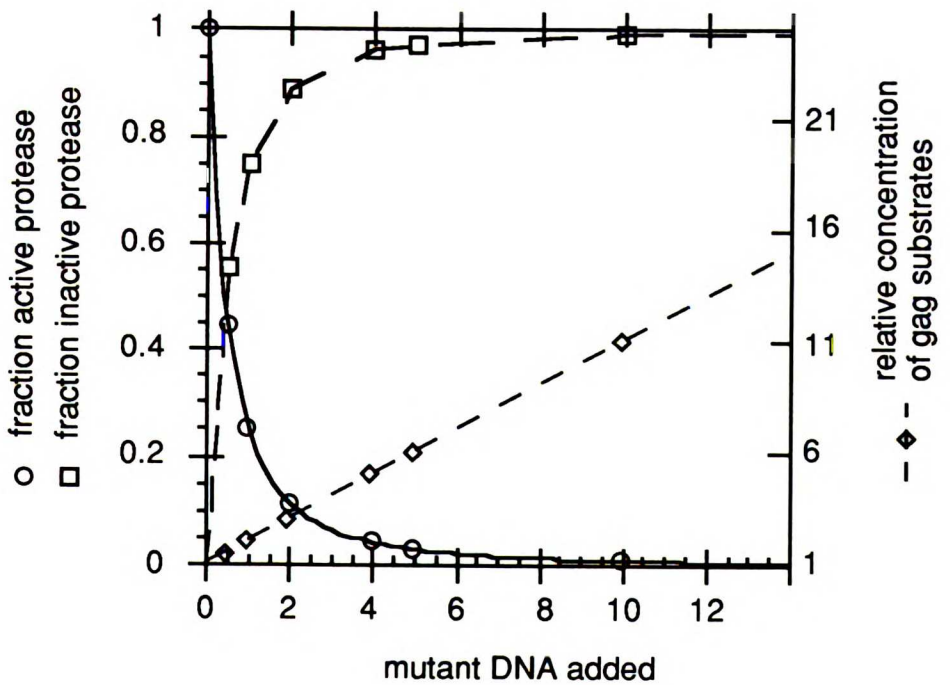
A. Schematic of budding at the cell membrane. Viral p160gag-pol polyproteins containing wild type (filled circles) or D25N mutant (open circles) proteases associate at the cell membrane during viral assembly and budding. When cells are transfected with a 1:1 mix of wild type to mutant DNA, equal amounts of wild type and mutant polyproteins are encapsidated. When a 1:5 ratio is used, more mutant protease monomers are encapsidated.

B. Graph showing concentrations of active protease (circles), inactive protease (squares) and relative concentration of gag substrates (diamonds) as predicted by the model. To generate this graph, wild type DNA concentrations were set at 1 and mutant DNA concentrations ranged from 0 to 14-fold over wild type.

A.



B.



active protease $= (WT/total)^2$

inactive protease $= 2 \cdot [(WT/total) \cdot (1 - WT/total)] + (1 - WT/total)^2$

gag substrate = total DNA

a decrease in processing efficiency at high mutant to wild type ratios. This decrease in processing was correlated with a decrease in particle infectivity. A second set of constructions was built which increased D25N PR expression by placing *gag* and *pol* in the same reading frame, such that only $p160^{gag\text{-}pol}$ was synthesized. These frameshift mutants were able to inhibit processing and dramatically lower infectivity at low ratios of mutant to wild type plasmid, and appeared to have an additional inhibitory effect on particle formation.

To examine the effect of constitutively expressed D25N PR on wild type virion formation a stable cell line was constructed in COS-7 cells using the pHIV-gpt D25N PR vector. Virions produced by transfection of this cell line with wild type pHIV-gpt contained significantly lower levels of processed viral proteins and were less infectious on a per particle basis than particles obtained from the parent COS-7 cells. Though preliminary, these results suggest that defective protease monomers can disrupt protease activation in a cellular context and may provide a means of inhibiting virus spread in a gene therapy application.

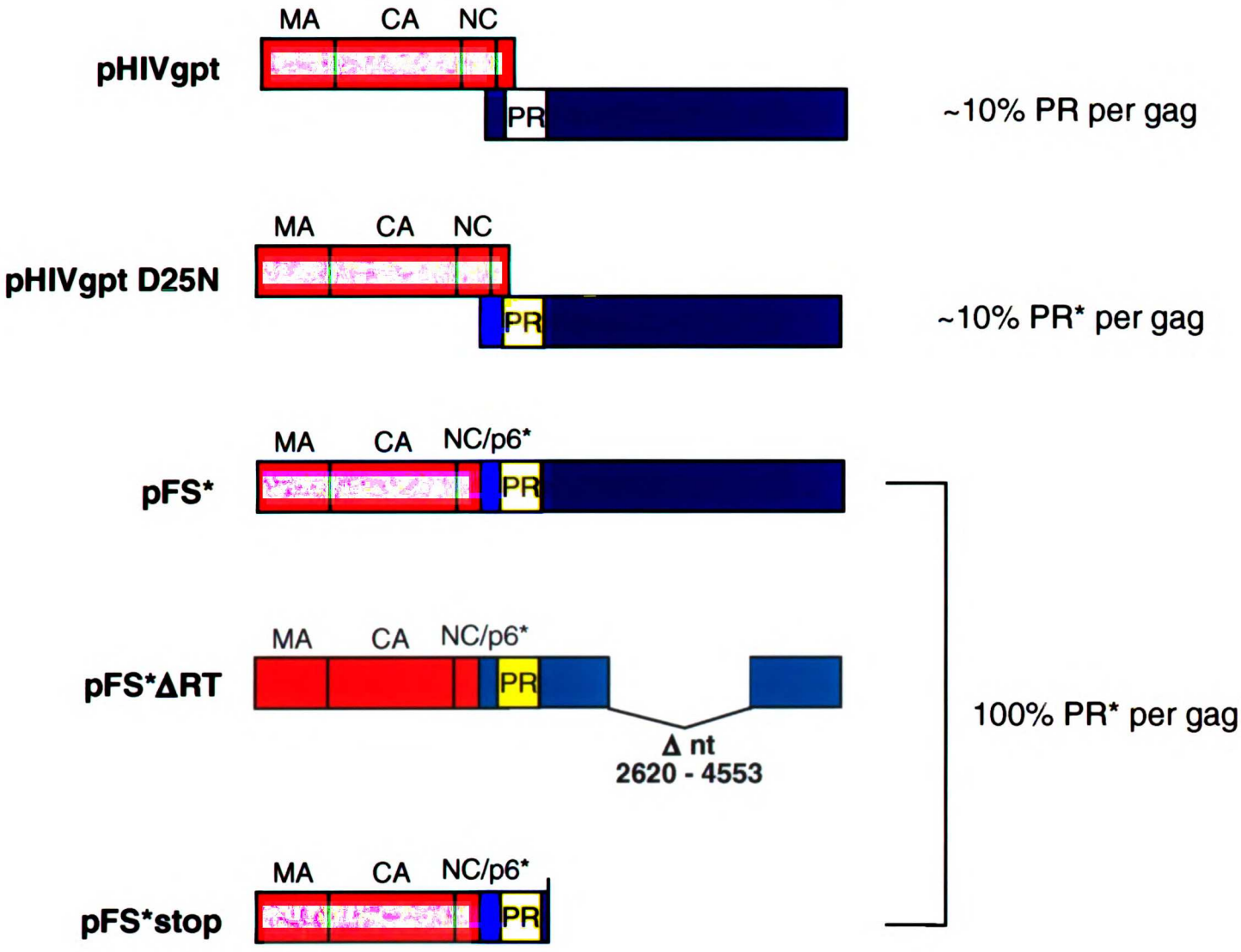
Materials and Methods

Cells, plasmids, infectivity assays, and basic tissue culture protocols were the same as those described in the previous chapter. The plasmid SV-A-MLV-env contained the coding sequences of the amphotropic murine leukemia virus envelope (nucletodies 5407 to 7846) in an SV40 derived expression vector (Page, et al., 1990).

Mutant Constructions. Frameshift constructions were designed in the HIV-gpt D25N PR vector based on a mutant reported by Karacostas, et al., 1993, and

Figure 4-2. Constructions used for dominant negative studies.

The wild type and D25N PR expressing pHIV-gpt vectors used in this study were the same as those previously described. The pFS* vector contained a frameshift mutation to place *gag* (red) and *pol* (blue) in the same reading frame, and also contained the D25N protease mutation (yellow). The pFS*ΔRT vector contained the frameshift mutation, the D25N protease mutation, and a deletion in the *pol* genes spanning nts 2319 - 4131. The pFS*stop vector contained the frameshift mutation, the D25N protease mutation, and a stop codon at nt 2108 of the *pol* gene.


MA
CA
NC
pHIVgpt
PR
~10% PR per gag
MA
CA
NC
pHIVgpt D25N
PR
~10% PR* per gag
MA
CA
NC/p6*
pFS*
PR
MA
CA
NC/p6*
pFS*ΔRT
PR
Δ nt
2620 - 4553
MA
CA
NC/p6*
pFS*stop
PR
100% PR* per gag

are shown in Figure 4-2. The oligonucleotide 5'-CAGGCTAATTTTATTAGG GAAGATCT-3' was used to insert a single "A" residue (in bold) in the phagemid vector pSPRD25N at nt 1635 of the viral genome. The mutant was confirmed by sequencing through the site of mutation and cloned back into the HIV-gpt vector to make the frameshifted, D25N-protease construction pFS*. A second construct contained a deletion in reverse transcriptase (RT) which was generated by deletion of the *Msc* I - *Msc* I (nt 2319 - 4131) fragment from pFS* to yield pFS*ΔRT. A third construct was engineered in which the oligonucleotide 5'-CATTAGCCCTTAAG CTTCTGTACCAGTA-3' was used to place a stop codon after the codon for the fourth amino acid of reverse transcriptase in pSPRD25N. Mutants were identified by screening for an *Afl* II site inserted by the same mutagenesis oligonucleotide (underlined). After confirmation by sequencing, this mutation was cloned back into pHIV-gpt to generate the construct pFS*stop.

Quantitation of p24. In order to determine particle-associated p24 levels, samples of virions purified by sucrose cushion centrifugation were diluted in PBS to volumes corresponding to dilutions of 1:10, 1:500, or 1:1000 of the initial supernatant. These samples were assayed using an enzyme-linked immunosorbent assay kit (DuPont NEN, Wilmington, DE) which is sensitive to the processed forms of the major capsid antigen (p24 and p25) but not the p55^{gag} or p160$^{gag\text{-}pol}$ precursor. For determination of relative particle numbers, 10 - 50 µl of purified capsids were diluted to 200 µl in HIV-1 protease activity buffer (50 mM NaOAc, pH 5.5, 1 M NaCl, 1 mM DTT, 5 mM EDTA) containing 0.1% Triton. The samples were digested by 100 ng of recombinant HIV-1 protease at 37°C to convert all precursor-associated p24 to its mature form. Processed samples were then assayed for p24 content by ELISA as described previously. This provided a

quantitative value for total particle number, or total particle-associated gag protein.

Stable cell lines. Stable cells were generated using techniques similar to those used to isolate cells stably expressing the wild type HIV-gpt proteins (Babé and Craik, 1994). COS-7 cells at approximately 30% confluence were transfected with 20 μg of pHIV-gpt D25N PR. After two days the transfected cells were placed in selective media consisting of DMEM containing 75 μg/ml mycophenolic acid, 250 μg/ml hypoxanthine, and 14 μg/ml xanthine in addition to 10% fetal calf serum and antibiotics. Following 20 days of selection, single colonies were isolated and transferred to 96-well plates. Thirty clones which continued to grow were expanded to 24 well plates and the culture media assayed for p55 by anti-p24 ELISA which detected polyprotein-associated p24 (Intracel, Cambridge, MA). Fourteen positive clones were expanded serially to 12-well and then 6 well plates, and finally to 10 cm dishes. Six clones which continued to express viral proteins after this process were frozen for storage at -80°C. One clone, 12A2, was chosen for further study.

RT PCR of viral capsids. To assay for the presence of the mutant protease gene, viral capsids were isolated by sucrose cushion centrifugation. Capsid samples of 5-10 μl were disrupted in PCR buffer (10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) by heating to 80°C for 10 minutes in the presence of 100 pmol of a reverse primer 5'-CCTTTTCCATCTCTGTACAAATTTC-3' (nt 2205 - 2181) in a final volume of 16 μl. Samples were cooled to 23°C and 2 μl (50 U) of MoMuLV reverse transcriptase (New England Biolabs, Beverley, MA), 1 μl (40 U) of RNAsin (Promega, Madison, WI), and 1 μL of 10 mM dNTPs were added.

Reverse transcription was carried out at 23°C for 10 minutes, followed by one hour of incubation at 42°C and denaturation of the enzyme by heating at 94°C for five minutes. Polymerase chain reaction was carried out by adding 80 µl of PCR buffer containing 50 pmol each of the reverse primer, a forward primer 5'-CAGA AATTGCAGGGCCCCTAG-3' (nt 1540 - 1561), and 2.5 units of AmpliTaq® DNA polymerase (Perkin Elmer Cetus, Norwalk, CT). Five cycles of PCR were carried out by denaturation at 94°C for one minute, annealing at 56°C for one minute, and extension at 72°C for two minutes. Thirty additional cycles were carried out by denaturation at 94°C for one minute, annealing at 65°C for one minute, and extension at 72°C for two minutes. Reaction products were purified by GeneClean (Bio101, Inc) and analyzed by electrophoresis through 1% agarose. The presence of the D25N protease mutation was confirmed by digestion with *Hpa* II, which cleaved the 665 bp PCR product into two ~330 bp pieces.

Immunofluorescence microscopy. Cells were seeded at 50% confluence on plastic 2-well chamber slides (Nunc, Napeville, OH). After 24 hours, the media was removed and the cells were washed 3 times with PBS. Cells were fixed at 25°C using a 1:1 mix of acetone/methanol for 10 minutes. Rabbit anti-p24 polyclonal antibody (Intracel, Cambridge, MA) was diluted 1:500 in PBS containing 0.5% Tween 20 and 3% bovine serum albumin (BSA) and incubated for 1 hour at 37°C. After extensive washing a fluorescein conjugated goat anti-rabbit IgG (Pierce, Rockford, IL) was used as a secondary antibody at a 1:25 dilution for 30 minutes at 37°C. Following extensive rinsing with PBS, the slides were viewed by oil immersion at 60x magnification using an Olympus BH-2 microscope equipped with a mercury 100 lamp (Chui Technical Corp.).

Results

Interference effects with pHIV-gpt D25N PR. To examine the effects of D25N PR monomers on particle maturation, wild type and D25N PR HIV-gpt vectors were mixed in ratios ranging from 1:1 to 1:6. Capsids isolated after 48 hours were analyzed by Western blot for the presence of processing intermediates. The results, shown in Figure 4-3, show that in the presence of increasing amounts of the HIV-gpt D25N PR vector there was a significant increase in the amount of precursor material present in viral capsids, as well as a decrease in the amount of fully processed p24. At 1:1 and 1:2 ratios there was an increase in the total amount of p24 and p25 detectable by ELISA (data not shown). By quantitation of the amount of processed vs unprocessed gag (Figure 4-3B) it was observed that the net amount of processing decreased as the amount of D25N PR DNA increased. At a 1:6 ratio, the total amount of processed protein fell to 5% of wild type levels, and capsid polypeptide profiles showed the presence of mostly $p55^{gag}$. In all cases, the amount of p24 and p25 present was significantly lower than what would have been expected had the mutant polyproteins been fully processed by the protease. This suggested that the presence of the D25N PR monomer was having an effect on protease activity.

Infectivity of mixed particles was studied in co-transfections with the HXB2-env plasmid. As shown in Figure 4-4, the number of infectious particles decreased as the amount of mutant DNA added to the transfection increased. Although the numbers of infectious particles were generated were very similar for wild type and low ratio mixes (1:1, 1:2), the increase in total particle number in these mixes resulted in an overall decrease in the ratio of infectious to total

ratio processed/total p55
1.2
1
0.8
0.6
0.4
0.2
0
WT
WT/1 D25N
WT/2 D25N
WT/4 D25N
WT/6 D25N
D25N

68 kDa
43kDa
29 kDa
p55
p39
p25
p24
WT
WT/1 D25N
WT/2 D25N
WT/4 D25N
WT/6 D25N
D25N

Figure 4-3. Inhibition of polyprotein processing by co-transfection of wild type and D25N PR HIV-gpt vectors.

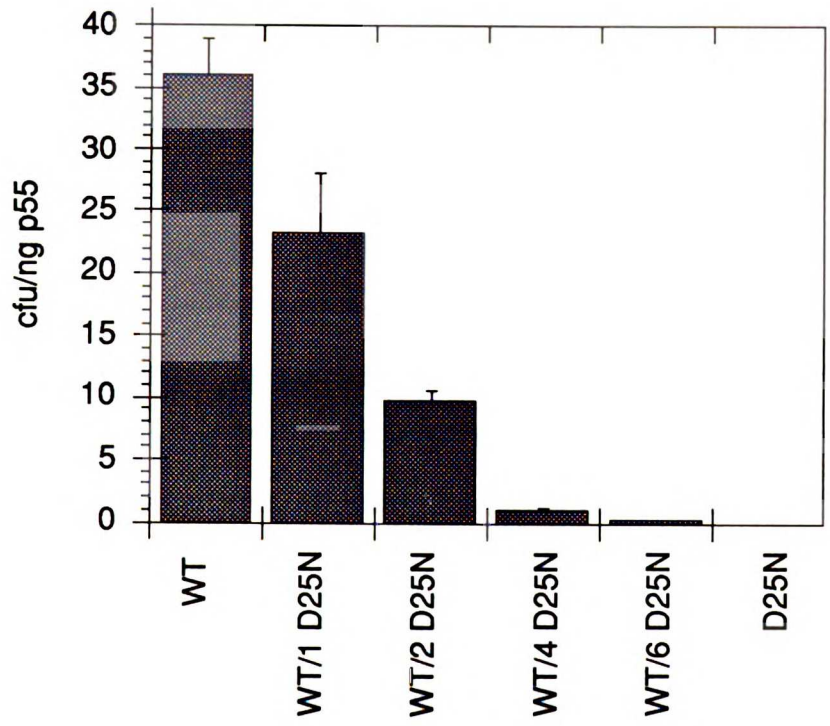
293 cells were transfected with 10 μg of wild type HIV gpt and 0, 10, 20, 40, or 60 μg of the HIV-gpt D25N PR vector. At 48 hours post transfection viral capsids were isolated and analyzed by Western blot following 10% SDS-PAGE (left panel) or by ELISA to determine processed (p24, p25) and total gag content (right panel). Degree of polyprotein processing is displayed graphically by dividing quantitated p24, p25 levels by the amount of total gag protein in purified capsids.

ratio processed/total p55
1.2
1
0.8
0.6
0.4
0.2
0
WT
WT/1 D25N
WT/2 D25N
WT/4 D25N
WT/6 D25N
D25N

68 kDa
43kDa
29 kDa
p55
p39
p25
p24
WT
WT/1 D25N
WT/2 D25N
WT/4 D25N
WT/6 D25N
D25N

Figure 4-4. Infectivity of mixed WT/D25N particles.

Viral particles were obtained by transfecting 293 cells with 60 μg of HXB2 env in addition to varying ratios of wild type and D25N PR HIV-gpt. Viruses were harvested at 48 hours post transfection and used to infect HeLaT4+ cells. Drug resistant colonies were fixed and counted following selection for 10 days. Colony numbers of WT/D25N PR mixes were divided by values for total gag protein (p55) obtained by ELISA to generate cfu/ml values.


cfu/ng p55
40
35
30
25
20
15
10
5
0
WT
WT/1 D25N
WT/2 D25N
WT/4 D25N
WT/6 D25N
D25N

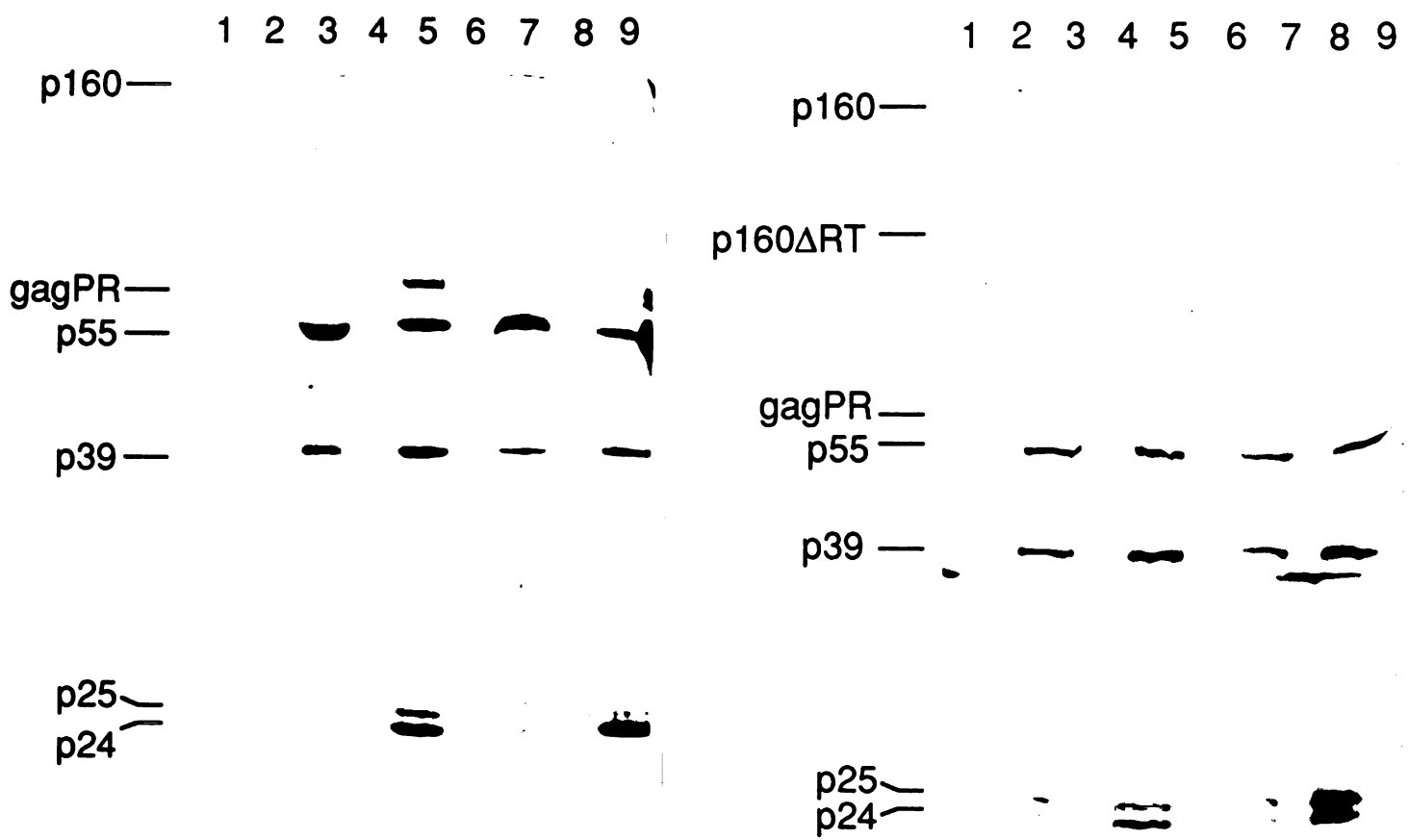
viral particles. Infectivity decreased approximately 40-fold over the range tested.

Frameshift mutations. In an attempt to increase the amount of defective protease monomer produced per microgram of transfected plasmid DNA, a mutant was made in the frameshift region of the viral genome. By inserting an additional "A" nucleotide in the *gag-pol* overlap region, the *gag* and *pol* genes were placed in the same reading frame so that 100% of the viral polyproteins were expressed as $p160^{gag\text{-}pol}$, rather than the 10% normally synthesized by the ribosomal frameshifting mechanism (Jacks, et al., 1988). Truncations were made in the *pol* region (pFS*stop, pFS*ΔRT) to facilitate incorporation of the mutant polyproteins into virions, and to reduce toxic effects which might be induced by overexpression of active RT.

Transfection of COS-7 cells with the three frameshift constructions alone, or in a 1:1 mix with wild type pHIV-gpt, yielded the results shown in Figure 4-5. In the absence of the wild type vector no gag-pol containing particles were observed in supernatants from cells transfected with pFS* or pFS*ΔRT (4-5A, lanes 2, 6, 8). A small number of particles containing an immature 66 kDa species (gag-PR) were observed in supernatants of cells transfected with pFS*stop, which agreed with results reported previously (Mergener, et al., 1992). Examination of lysates of transfected cells (Figure 4-5B) showed that the polyprotein precursors were all expressed, but apparently the $p160^{gag\text{-}pol}$ and $p160^{\Delta RT}$ polyproteins could not form particles. Viral particles could be efficiently produced when the frameshift constructs were transfected with an equal amount of wild type DNA (Figure 4-5A, lanes 3, 5, 7, 9). These particles contained p55, p39, p25, and p24, indicating that some protease activity remained. The truncated product of

Figure 4-5. Effect of frameshift constructions on viral capsid production.

Transfections were performed using either frameshift constructions alone (lanes 2, 4, 6, 8) or in 1:1 mixes with wild type pHIV-gpt DNA (lanes 3, 5, 7, 9). At 48 hours post-transfection, samples of virus from cell supernatants (A) were purified by sucrose cushion as described previously. Intracellular viral proteins (B) were obtained by lysing cells and separating supernatant and pellet fractions. Proteins were separated by 10% SDS-PAGE and analyzed by Western blot with anti-p24 antibodies. Lane 1, no DNA; lanes 2, 3, 6, and 7, pFS*; lanes 4 and 5, pFS*stop; lanes 8 and 9, pFS*ΔRT.


1 2 3 4 5 6 7 8 9
p160
gagPR
p55
p39
p25
p24
1 2 3 4 5 6 7 8 9
p160
p160ΔRT
gagPR
p55
p39
p25
p24

pFS*stop (gag-PR) could be detected in particles in the presence of the wild type proteins, whereas no significant amounts of $p160^{\Delta RT}$ were observed in mixed particles.

In terms of diminishing p24 and p25 production, the FS* construct appeared most effective, followed by the FS*stop clone and the FS*ΔRT clone, which was the least effective. In all cases, however, the effect of mixing the mutant and frameshift constructs equally was considerably greater than that observed when mixing the wild type and D25N constructs at a 1:1 ratio.

Polyprotein processing of capsids formed from 1:1 mixes between WT/FS* and WT/FS*stop vectors was analyzed over a time course from 24 to 72 hours to assess the irreversiblity of protease inactivation (Figure 4-6). Neither mixed-capsid population showed significant processing at 24 hours. At 48 and 72 hours a significant amount of p24, p25 and p39 were observed in the WT/FS*stop mixed virions, while WT/FS* mixes did not show significant levels of processing even after 72 hours. Both polyproteins appeared to inhibit protease activity in an irreversible manner, although the FS* gag-pol polyprotein appeared to be more efficient than the FS*stop polyprotein.

A titration was performed with the FS* construct using amounts of DNA that were 0.25, 0.5, and 1 relative to wild type. The effect of these ratios on p24, p25 levels, and capsid polyprotein profiles is shown in Figure 4-7. At a ratio of 1:0.25, processed precursor levels were reduced nearly 2-fold, and nearly 4-fold and 10-fold at ratios of 1:0.5 and 1:1, respectively. When compared to the titration results obtained from mixing wild type and D25N PR vectors (Figure 4-3), the FS* construct displayed nearly a 6-fold increase in efficacy at reducing polyprotein processing.

Figure 4-6. Time course of polyprotein processing in WT/frameshift mixed particles.

293 cells were transfected with wild type HIV-gpt DNA (lane 1) or 1:1 mixes between wild type HIV-gpt and pFS* (lane 2) or pFS*stop (lane 3). Capsids were harvested from cell supernatants at 24, 48, or 72 hours post transfection. Capsid contents were analyzed by Western blot following separation by 10% SDS-PAGE.

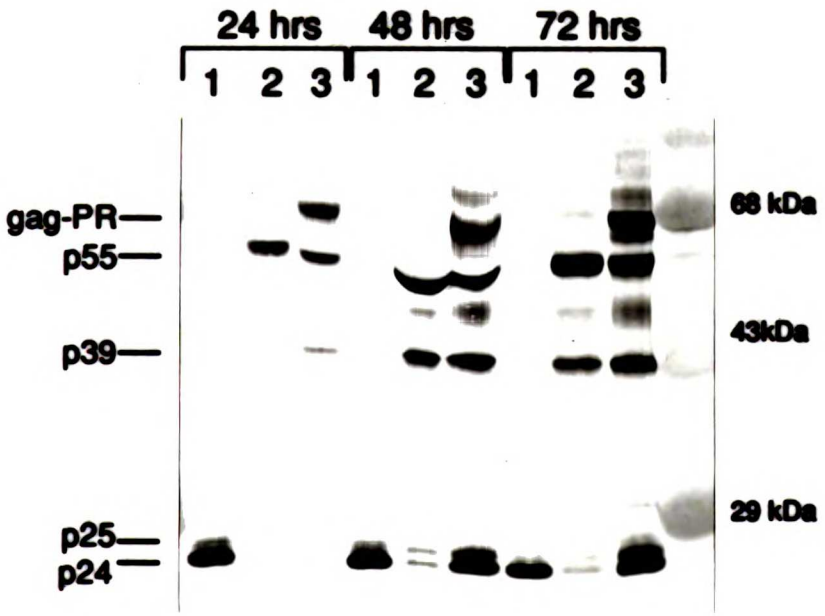

24 hrs
48 hrs
72 hrs
1 2 3 1 2 3 1 2 3
gag-PR
p55
p39
p25
p24
68 kDa
43kDa
29 kDa

Figure 4-7. Titration of pHIV-gpt with pFS*.

293 cells were transfected with 30 μg of wild type HIV-gpt and 0, 7.5, 15, or 30 μg of pFS* DNA. At 48 hours post transfection viral capsids were isolated and analyzed by Western blot following separation by 10% SDS-PAGE (left panel) or ELISA (right panel) to determine processed and total gag content (right panel). As described previously, Processed gag (p24, p25) and total gag protein was quantitated by ELISA as described in Materials and Methods.

ratio processed/total p55
0
0.2
0.4
0.6
0.8
1
1.2
WT
WT/0.25 FS*
WT/0.5 FS*
WT/1 FS*
FS*

68 kDa
43kDa
29 kDa
p55
p39
p25
p24
WT
WT/0.25 FS*
WT/0.5 FS*
WT/1 FS*
FS*

Infectivity of frameshift mutant mixes. The infectivity of particles generated by co-transfection of equal amounts of the wild type plasmid and frameshift mutants was examined at 48 hours. The amphotropic murine leukemia virus (A-MLV) envelope was used in order to achieve higher titers, as no infectious particles could be obtained in wild type/frameshift mixes using the HIV HXB2 envelope (data not shown). The results are shown in Figure 4-8. Relative infectivities corresponded approximately to the degree of maturation of the viral particles. Virions formed from 1:1 co-transfections with wild type HIV-gpt DNA and the full length FS* construct were most severely compromised, showing nearly a 100-fold reduction in per-particle titer than wild type virions. Particles formed from mixes with the *pol*-truncated constructs were approximately 20-50-fold less infectious on a per particle basis than fully wild type particles. Although these results were reproducible, the number of infectious particles which were detected were very low, even in the presence of the A-MLV envelope which ordinarily yielded very high titers (Steve Fabian, personal communication).

Reverse transcriptase activity in WT/*pol*-truncation mixes. The presence of active reverse transcriptase was analyzed in mixes betwen the wild type HIV-gpt vector and the FS*ΔRT and FS*stop constructs. Since gag-pol precursors supplied by the mutants contained no functional RT, any activity present would be due to wild type $p160^{gag\text{-}pol}$ incorporated into particles. The results are shown in Figure 4-9. It had been anticipated that particles formed by mixes between wild type and truncated gag-pol polyproteins would have considerably less RT activity than wild type particles, since the only source of active RT would be the wild type polyproteins. Significant RT activity was found in these

Figure 4-8. Infectivity of mixed WT/frameshift particles

293 cells were transfected with mixes between 15 μg of wild type DNA and 15 μg of pFS*, pFS*stop, or pFS*ΔRT, in addition to 30 μg of pSV-A-MLV. Viruses were harvested at 48 hours post-transfection and used to infect HeLaT4+ cells. Drug resistant colonies were fixed and counted following selection for 10 days. Total gag protein was quantitated as described previously and infectious particle numbers reported as colony forming units (cfu) per ng of p55. Numerical values for frameshift mixes are also shown.

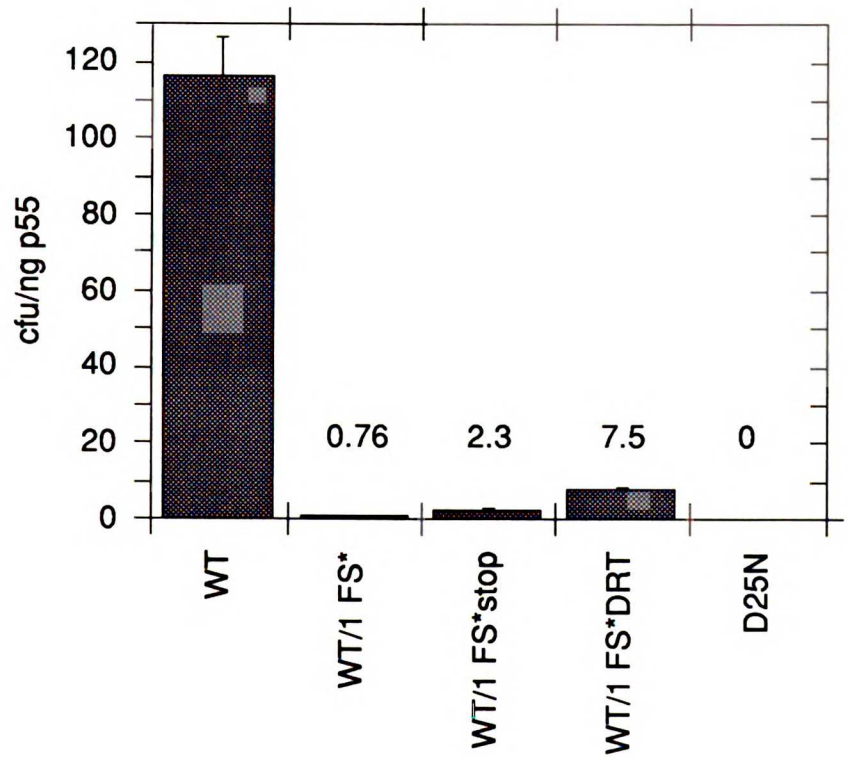
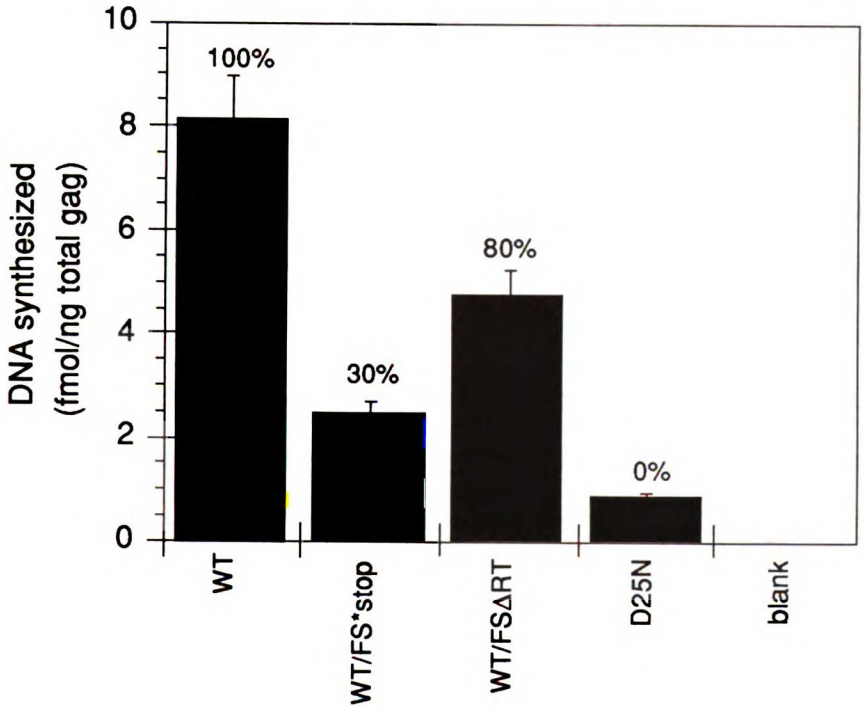

cfu/ng p55
120
100
80
60
40
20
0
0.76
2.3
7.5
0
WT
WT/1 FS*
WT/1 FS*stop
WT/1 FS*DRT
D25N

Figure 4-9. Reverse transcriptase activity in capsids formed from wild type/*pol*-truncation mixes.

Virions were purified from transfected cell supernatants at 48 hours post-transfection as previously described. Endogenous RNA-directed DNA polymerase activity in viral lysates was assayed and normalized for the amount of total gag protein present. Mixes between wild type DNA and the pFS*stop and pFS*ΔRT constructs were made at 1:1 ratios. The percentage values above the data columns indicate the degree to which viral proteins were processed in the virus samples, as defined by dividing the amount of p24 and p25 divided by total gag protein levels.


DNA synthesized
(fmol/ng total gag)
10
8
6
4
2
0
100%
30%
80%
0%
WT
WT/FS*stop
WT/FSΔRT
D25N
blank

particles, however, suggesting that full length wild type gag-pol was efficiently encapsidated. This raised the possibility that a preferential association occurred between full length and truncated gag-pol products, a possibility which is consistent with the decreased ability of the truncated constructs to block polyprotein processing.

Stable expression of D25N PR in mammalian cells. A stable cell line expressing the D25N PR mutant was isolated in order to test the effects of constitutive defective-monomer expression on wild type virus production. The simple D25N PR mutation was chosen over the FS* construct for this experiment since it was not clear what effect stable expression of high levels of active RT and IN would have on cell viability. COS-7 cells transfected with pHIV-gpt D25N PR were grown in the presence of high concentrations of mycophenolic acid (75 μg/ml) to select for stable integration of the *gpt* gene encoded in the vector. Six clones were isolated which expressed levels of $p55^{gag}$ from 20 - 120 pg/ml/hr per confluent 10 cm dish. A single clone, 12A2, was found to express viral proteins at approximately 60 pg/ml/hr per confluent 10 cm dish over a period of two months, and was expanded for further study.

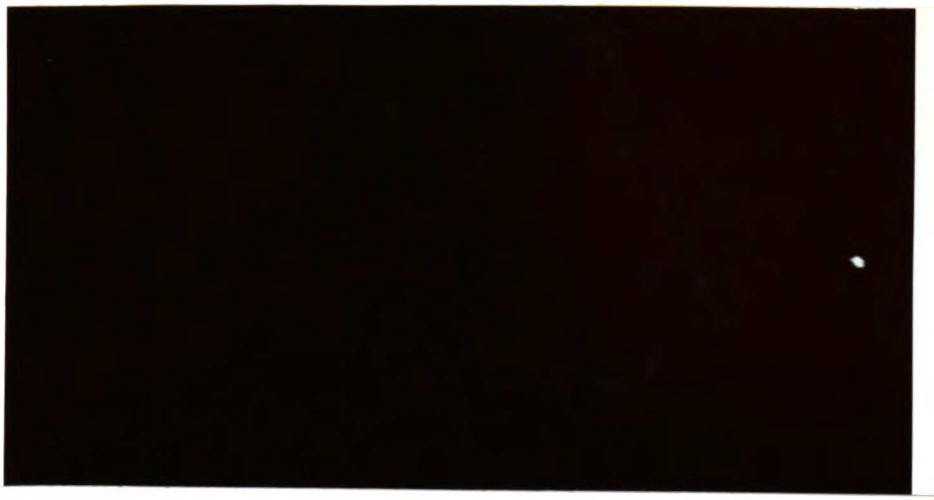
In order to characterize the HIV-1-specific protein expression of the 12A2 cell line, parental COS-7 cells and 12A2 cells at approximately 50% confluence were fixed and stained with a polyclonal antibody to p24. Proteins stained by the anti-p24 antibody were visualized by immunofluorescence microscopy using a fluorescein conjugated secondary antibody. As shown in Figure 4-10, the parental cells (A) showed little or no staining with this antibody. Nearly all of the 12A2 cells (B) stained extensively with the p24 antibody, although expression levels varied from one cell to the next. The staining pattern observed is similar to

Figure 4-10. Characterization of HIV-gpt D25N cell line.

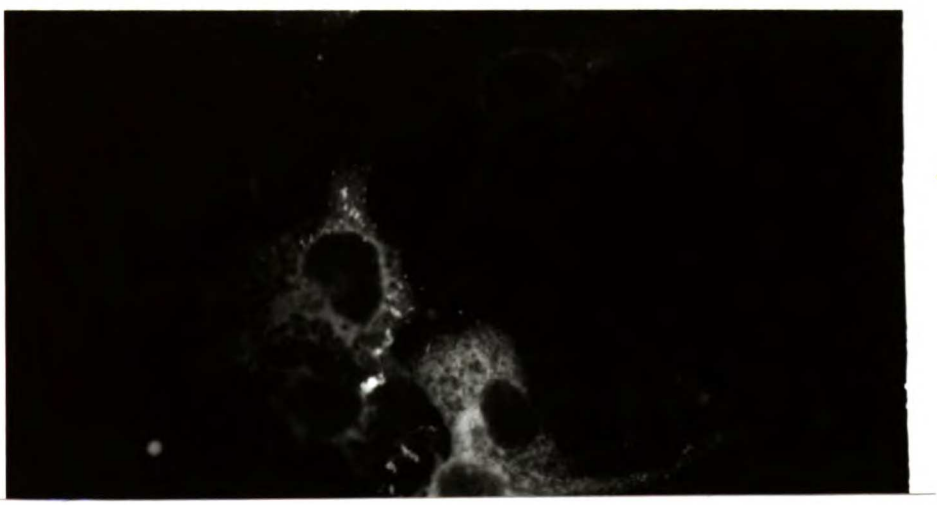
Immunofluorescence microscopy with anti-p24 antibody of (A) COS-7 cells at approximately 15% confluence, and (B) 12A2 cells at approximately 10% confluence. The 12A2 cells show extensive staining with the antibody when compared with the parental COS-7 line. The staining pattern observed is similar to that obtained for a cell line stably expressing the wild type HIV-gpt proteins in COS-7 cells (Babé and Craik, 1994).

C. Capsid analysis. RT-PCR was performed on viral capsids harvested from from cells transfected with pHIV-gpt (lanes 1, 2), pHIV-gpt D25N (lanes 5, 6), or isolated from the stable cell line 12A2 (lanes 3, 4). Lanes 2, 4, and 6 represent RT-PCR products digested with *Hpa* II. The 660 and ~900 bp PCR products represent full length products spanning the *PR* gene. Cleavage at the *Hpa* II site in genes harboring the D25N mutation produces 330 and ~600 bp species from these products.

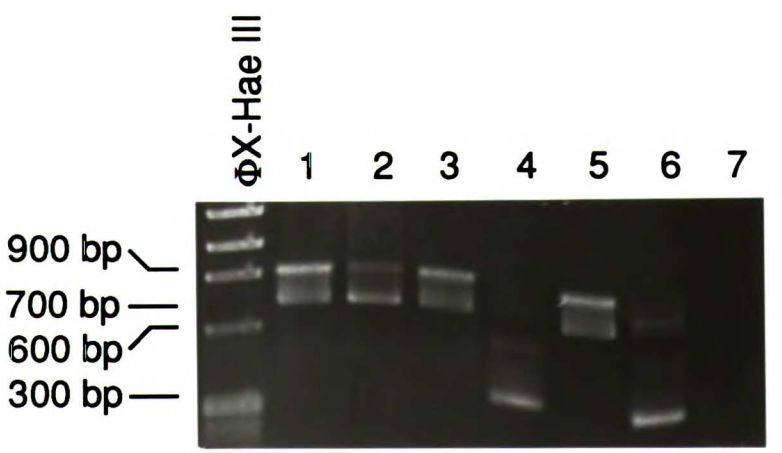
A.



B.



C.



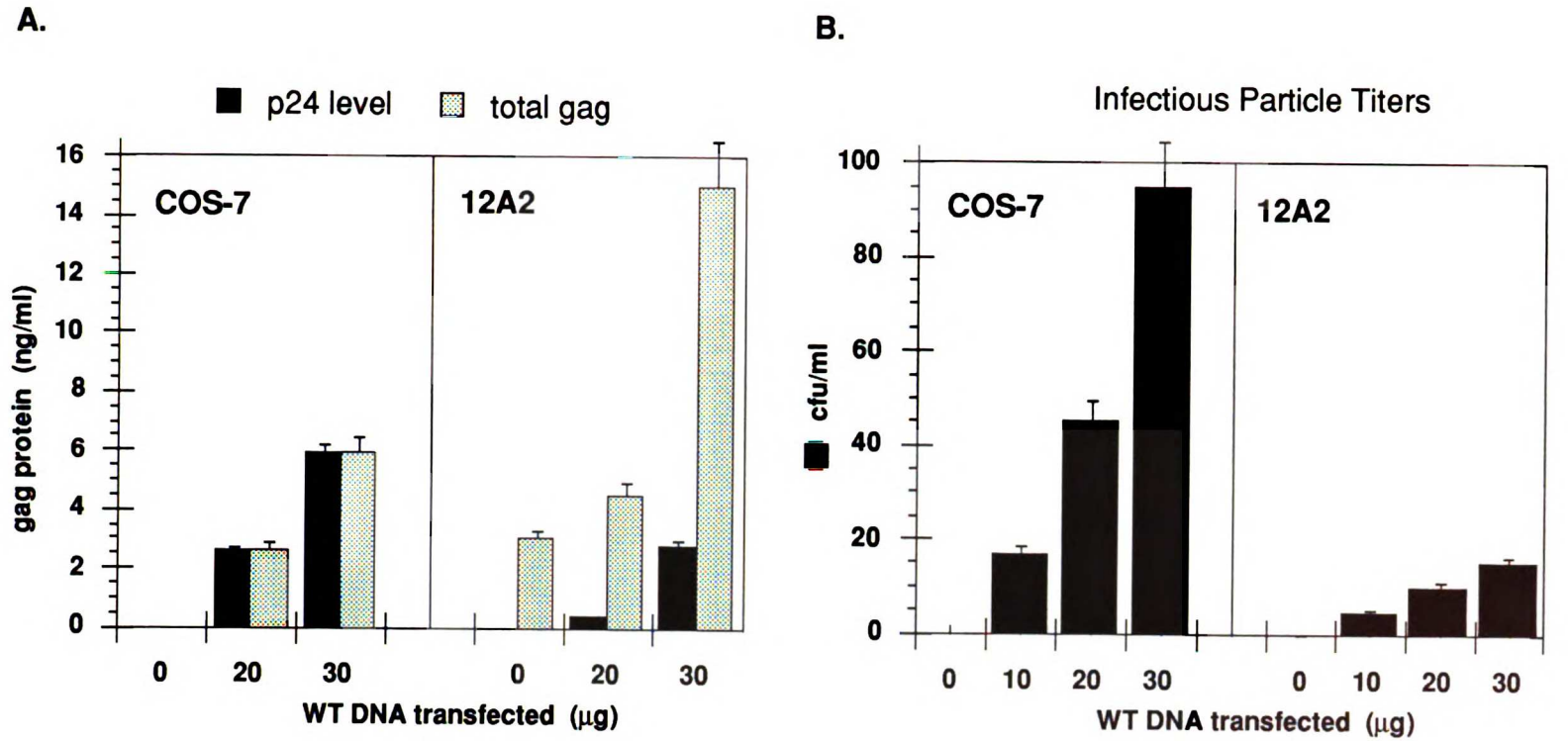
that previously observed for expression of wild type HIV-1 proteins in COS-7 cells (Babé and Craik, 1994).

The contents of viral capsids produced by 12A2 cells were analyzed for the presence of viral RNA by RT-PCR of pelleted material in sucrose cushions. As shown in Figure 4-10C, RT-PCR of 12A2 cell supernatants gave amplified products of the same sizes as those obtained from COS-7 cells transfected with either wild type or pHIV-gpt D25N PR. The doublet observed at ~650 and ~900 bp originated from the reverse primer priming at a second region of the genome which was highly homologous to the nucleotide sequence at 2180-2205. This could not be eliminated without significantly lowering the efficiency of the PCR reaction (data not shown). The PCR products were digested with *Hpa* II, which was diagnostic for the D25N mutation. Products derived from 12A2 and D25N-transfected capsids were cleaved, whereas those from wild-type transfected cells were not. This confirmed that the 12A2 capsids contained viral RNA, and that the RNA encoded the D25N protease mutation.

Interference with infectious virus production in 12A2 cells. Transfections were carried out to examine the ability of constitutively expressed HIV-1 D25N PR polyproteins to disrupt particle maturation. Parental COS-7 and D25N PR-expressing 12A2 cells were transfected with 60 μg of HXB2 env and amounts of HIV-gpt DNA ranging from 0 to 30 μg (Figure 4-11). Supernatants harvested at 48 hours were assayed for polyprotein content (A) and infectivity (B). In COS-7 cells, particles were completely processed and significant numbers of infectious particles were detected. In transfected 12A2 cells, however, a majority of the particle associated gag protein was present in immature precursors. The number of infectious particles produced by this cell line was reduced 5 - 7 fold relative to

Figure 4-11. Effect of constitutively expressed D25N PR on HIV particle maturation and infectivity

COS-7 and 12A2 cells were transfected with 60 μg of HXB2 env and amounts of wild type HIV-gpt DNA ranging from 0 to 30 μg. Particles were isolated at 48 hours and assayed for gag protein content (A). Quantitative values for gag content were not available for cells transfected with 10 μg of wild type DNA. Infectious particle titers (B) were determined and are reported here as colony forming units per ml of cell supernatant (cfu/ml).


A.
p24 level
total gag
COS-7
12A2
gag protein (ng/ml)
16
14
12
10
8
6
4
2
0
0
20
30
0
20
30
WT DNA transfected (μg)
B.
Infectious Particle Titers
COS-7
12A2
cfu/ml
100
80
60
40
20
0
0
10
20
30
0
10
20
30
WT DNA transfected (μg)

numbers observed in COS-7 cells. This demonstrated that polyproteins containing a defective protease monomer could inhibit protease function and infectious virus production when constitutively expressed.

Discussion

Reduction of HIV-1 protease activity by at least 50-fold is sufficient to prevent the maturation process and render viral particles noninfectious (Rosé, et al., 1995). Although this level of inhibition can be achieved by many small molecule inhibitors, these inhibitors have proven to be ineffective in the presence of point mutations in the protease that lead to a loss of inhibitor sensitivity. In addition, diffusion of small molecules out of virions might allow inhibited proteases to be reactivated. In order to circumvent these difficulties, we proposed to use an inhibitory approach which could exploit extensive and highly specific macromolecular interactions with the HIV-1 protease. A strategy using a defective monomer of the HIV-1 protease which lacked the catalytic aspartate provided us with both a highly specific and high affinity interaction, and one which should be irreversible in the context of the virion.

Simple mixing experiments, in which D25N PR and wild type HIV-gpt vectors were co-transfected, indicated that polyprotein processing could be inhibited in the presence of excess mutant polyproteins. Interpretation of the results was complicated by the large increase in total gag protein expressed from the D25N PR genome. In co-transfections at ratios ranging from 1:1 to 1:4 wild type to mutant, absolute p24 and p25 levels actually increased. This was most likely due to significant amounts of residual protease activity (predicted to range from 25% to 4%) in the presence of large excesses of gag substrates expressed from the D25N PR vector. In all cases, however, the amount of p24 and p25

observed fell significantly below that expected should all available gag and gag-pol proteins from the wild type and mutant constructions have been processed. The decrease in the percentage of total gag proteins which were processed in each co-transfection reflects this effect.

It is interesting to note that a dramatic reduction in gag processing appears to occur between the 1:2 and 1:4 ratios of wild type to HIV-gpt D25N PR DNA. Based on the original model, a 1:4 ratio of wild type to mutant DNA represents a scenario where protease activity would be approximately 5%, a level similar to that observed with the A28S protease mutant described in the previous study. By using mixes of wild type and D25N PR DNA it may be possible to generate viral particles with protease activity levels that range between the T26S and A28S protease mutants, thereby allowing for a more stringent identification of the amount of protease activity required for efficient virion maturation.

The decrease in total gag processing was also reflected in a decrease in particle infectivity as more HIV-gpt D25N PR DNA was added. At a 6:1 excess of mutant to wild type DNA, where capsid processing fell to only five percent, particle infectivity was reduced approximately 40-fold. It is notable that infectivity dropped consistently as the amount of mutant DNA increased, despite the observation that polyprotein processing was not severly compromised until the mutant DNA was present in large excess over wild type DNA.

The data obtained here allow for an evaluation of the accuracy of the model mentioned in the beginning of this chapter. Drawing from this model, the predicted protease activity in 1:1 mixed virions is 25% of wild type. The amount of processed vs unprocessed gag protein in these samples is on the order of 80-90%, as judged by ELISA. Both the predicted level of activity and the observed p24/total gag ratio are similar to what was observed in the previous chapter for

the T26S mutant protease, where a known reduction of activity to 25% of wild type levels led to about 90% of gag being processed. As mentioned previously, where activity is predicted to be only 4% (1:4 ratio) the level of residual polyprotein processing is very similar to what was observed for the A28S mutant protease These data suggest that calculations for the reduction of protease activity based on the model for mixing wild type and mutant DNAs are valid ones.

Technical difficulties in the design of the experiment limited the usefulness of this model in evaluating the ability of mutant DNAs to completely suppress infectious particle formation. Since there is always a statistical possibility that a cell will take up only wild type DNA, a complete block to infectious particle formation is not assured. This could explain why infectious particle titer in the simple WT:D25N mixing experiments did not reach zero, even at 1:6 ratios of wild type to mutant DNA. Furthermore, the amount of DNA required to generate large drops in protease activity (five to eight-fold over wild type) results in a total amount of DNA during transfection which affects cell viability. The total amount of HIV-gpt DNA that was used reached as much as 70 μg at high wild type to mutant ratios. We also used 60 μg of HXB2 env plasmid in all transfections to try and maximize envelope expression for infectivity studies. Using more DNA than this resulted in inefficient or failed transfections and considerable cell toxicity. Therefore we felt it would be more useful to find a way to get away from this problem in the experimental design.

Genetic strategies to increase potency of the D25N monomer.

A method was sought to improve the potency of the defective monomer so that less DNA could be used. To this end, three constructions were built which

expressed only the gag-pol fusion protein. By functionally increasing the efficiency of the ribosomal frameshift at the gag-pol junction, the D25N protease could be synthesized at roughly 10 times the level of the wild type protease per molecule of mRNA present in the cell. Deletions and truncations were made in *pol* to reduce the toxicity which might be associated with overexpression of reverse transcriptase in the target cell. In addition, it was thought that truncated gag-pol proteins would be more easily incorporated into budding particles (Karacostas, et al., 1993).

The frameshift construction and two derivatives lacking part or all of *pol* were highly effective at suppressing polyprotein processing, showing significant reductions in particle maturation when mixed at 1:1 ratios with the wild type HIV-gpt vector. The truncated polyproteins, however, were less effective than the full length $p160^{gag\text{-}pol}$, which suggested that association between wild type $p160^{gag\text{-}pol}$ and the truncated gag-pol proteins was not efficient. Reverse transcriptase activity in mixed wild type/pol-truncated particles was assayed to test this hypothesis. If the truncated and wild type polyproteins were associating equally, it would be expected that the maximum RT activity present in these particles would be 50% that of fully wild type particles. Although processing was not complete in the mixed particles, RT levels in WT:FS*ΔRT particles were nearly 60% of wild type, and levels in WT:FS* particles were 30% of wild type. These data suggest that interactions between the pol proteins in $p160^{gag\text{-}pol}$ play a role in polyprotein dimerization. Thus, the full length frameshift appeared to be the most effective construct for maximizing association of D25N PR and wild type monomers.

The frameshift constructions were tested for other effects on particle maturation. In a 72 hour time course with WT/FS* and WT/FS*stop mixes, it

was observed that polyprotein processing remained effectively blocked, suggesting that no subunit exchange occurred between inactive heterodimers to regenerate wild type protease. Titration of the wild type vector with the FS* vector showed that polyprotein processing could be inhibited up to 90% at a 1:1 ratio. This was an approximately six-fold increase in potency over the original pHIV-gpt D25N PR vector. This increase in potency was expected, based on the prediction that frameshift efficiency, and therefore defective monomer synthesis, is increased by as much as 10-fold in the frameshifted constructions relative to the HIV-gpt D25N PR vector.

It was also noted that the frameshift constructions used in this study had an effect on particle formation as well as an effect on viral polyprotein processing. The frameshifted proteins alone were incapable of forming viral capsids, as reported in other systems (Karacostas, et al., 1993; Mergener, et al., 1992; Park and Morrow, 1992), but accumulated intracellularly. Mixing these in 1:1 ratios with the wild type HIV-gpt vector rescued particle formation, although the number of particles released was roughly equivalent to that observed with the wild type construction alone. In contrast, 1:1 mixes with the WT and pHIV-gpt D25N constructions led to the production of approximately twice the number of viral capsids. Thus it was possible that the reduction in infectious particle numbers with the frameshift constructs is the result of a bipartite effect of these constructs on both polyprotein processing and virus assembly. The reason for this block of particle formation is not understood. It may simply be due to the presence of relatively small amounts of $p55^{gag}$, which appear to be the major proteins involved in driving particle formation. It has also been proposed that steric hindrance or the absence of the p6 protein at the C-terminus of gag may inhibit particle formation (Craven, et al., 1991; Mergener, et al., 1992). This

feature of the frameshift constructions may be a beneficial one in a therapeutic setting where a decrease in antigenic particle production as well as particle maturation and infectivity may be desirable.

Constitutive expression of defective monomers.

The goal of a using a gene-therapy application of the dominant negative effect to reduce protease activity would be based on expressing the defective protease monomers in a target cell to interfere with viral polyprotein processing after infection with HIV. To model this situation we constructed a cell line (12A2) which stably expressed the proteins encoded by the HIV-gpt D25N PR genome. Expression of the mutant viral proteins was confirmed to be homogeneous throughout the cell population by immunofluorescence microscopy. Immature viral proteins were detected in a pelletable fraction of media harvested from these cells, and RT-PCR of the isolated pellets indicated that viral RNA encoding the D25N protease mutation was present in these particles.

It was necessary to model infection of the 12A2 cell line by HIV-1 by transfection with wild type HIV-gpt DNA, since the 12A2 line did not express the CD4 receptor required for HIV entry. Relative to the parental COS-7 cells, the 12A2 cell line produced mostly immature particles after transfection with wild type DNA. The amount of processing which was observed increased as the amount of wild type DNA transfected increased, reaching approximately 20% of wild type levels when 30 μg of HIV-gpt DNA was used. The production of infectious particles was also diminished relative to COS-7 transfection results, reaching only 15 - 25% of wild type levels.

This situation, in which the entire population of cells expresses the defective monomer, is preferable to the previous mixing experiments for gauging the

effects of defective monomers on protease activity. In previous experiments it was not possible to assure that all cells received D25N PR as well as wild type DNAs. From fluorescence microscopy studies it was clear that all of the cells in the 12A2 clonal population expressed the defective protease, although expression levels did vary slightly throughout the population. It is not clear what effect this variation in expression will have on suppressing infectious particle formation. This is a relevant issue to the use of defective monomers as therapeutic agents, since expression levels will be likely to vary throughout a population of cells transfected with constructs expressing the defective monomer.

Although preliminary, these results suggest that defective protease monomers may inhibit particle maturation when expressed in cells prior to the introduction of wild type HIV sequences. A stable cell line which expresses the CD4 receptor and the D25N PR genome has recently been derived from HeLaT4 cells (unpublished results), and may be used to address more realistic infection scenarios involving wild type HIV or pseudotyped virions produced by the HIV-gpt vector.

Further considerations in optimizing defective monomer presentation.

Genetic strategies which result in overexpression of full length $p160^{gag\text{-}pol}$ precursors containing a defective protease monomer were the most effective at inhibiting protease activity. In a gene therapy approach, the presence of active reverse transcriptase and integrase in a host cell might interfere with normal cellular function. Rather than constructing *pol* truncations, however, RT and IN activity could be eliminated by making point mutations at the active sites of these two enzymes which would inactivate them, while leaving the structures of the proteins intact. It might even be possible to increase the efficacy of a therapeutic

construction by identifying mutants in RT and IN which have dominant negative effects and using them in tandem with mutants in the protease to interrupt more than one enzymatic pathway.

A number of issues need to be addressed in order to generate more effective defective-monomer delivery systems. First, there is a possibility that small molecule protease inhibitors might be used in combination with the dominant-negative gene therapy approach. It would be ideal if defective protease monomers were designed in such a way as to prevent binding of protease inhibitors to inactive homo- or heterodimers. This would allow the selection of fully active wild type proteases as the primary targets for small molecule inhibitors and improve the synergistic action of the two therapeutic strategies. Following from this idea is the thought that additional protein engineering approaches may be employed to affect the association between mutant and wild type monomers. The forces which drive this interaction are presumed to be equal in both the wild type and D25N mutant protease. As described in equation 4-1, this implies that some of the defective monomers will be absorbed in mutant:mutant interactions rather than wild type:mutant interactions. In order to maximize the efficiency of wild type:mutant heterodimer formation it would be ideal to identify additional mutants in the protease which would increase the efficiency of mutant:wild type interactions and/or decrease the efficiency of mutant:mutant interactions. Mutants such as D25K, which is predicted to exploit charge-charge interactions in selecting dimerization partners, have recently been designed and tested in the laboratory, and shown to be more effective in inhibiting polyprotein processing than the single D25N mutation (Dr. Fiona McPhee, personal communication).

Chapter 5. Future Directions

The work described in this thesis has been aimed at using current techniques in protein chemistry and molecular biology to study the structure, function, and biology of an important chemotherapeutic target of the human immunodeficiency virus. Through site-directed mutagenesis, it is possible to introduce mutants into proteins to specifically alter their functions. We have used this strategy to engineer mutants into the HIV protease for the purpose of probing the sensitivity of the viral maturation process to variations in protease activity. We have been able to use recombinantly produced protease and tissue culture models of viral assembly to characterize mutants by both classical biophysical methods, and by observing the effects of mutants on viral maturation and infectivity.

The data derived from site-directed protease mutants have identified two interesting features associated with the protease activity in a virion. First, a reservoir of activity exists such that 75-80% of the activity in a virion can be eliminated without significant effects on the maturation process. Second, a 50-fold reduction in protease activity is sufficient to block the formation of infectious particles. The dramatic reduction in proteolytic processing of virions containing 50-fold less protease activity relative to those containing 5-fold less activity suggest that a threshold of activity exitsts, and that processing of polyproteins is not linearly related to protease activity. Mixing wild type and mutant DNAs in co-transfections allowed us to study levels of protease activity between those of site-directed mutants. This procedure has defined the threshold of activity as being between 11 and 4% of the wild type level (Figure 4-3). Further mixing studies may help to define this threshold with greater accuracy. These observations will help to define how effective inhibitors of the protease must be in order to block the spread of the virus.

The catalytic mutants of the protease have also raised an interesting question regarding the precise reason for the loss of infectivity with diminished processing. Since viruses harboring the mutant proteases assemble normally, it is possible that they may be having an effect on the early events in the viral life cycle. The protease has been implicated as playing an active role in uncoating the viral RNA (Baboonian, et al., 1991; Nagy, et al., 1994), although conflicting reports have been published on this subject (Jacobsen, et al., 1992). A number of early events in the viral life cycle may be affected by insufficient protease activity and polyprotein processing: RNA uncoating may be prevented, polyprotein associated RT and IN may be inefficient, or some subtle defect may exist in the RNA packaged by immature virions. The mutant virions generated in our studies could provide useful tools for studying the early steps in viral infection.

Our work to date has been aimed at studying the effect of different levels of HIV-1 protease activity on viral maturation. The system we have described could also be used to study mutants with altered substrate specificities. In particular, point mutations emerging from clinical trials of protease inhibitors may alter protease substrate specificity and could be rapidly characterized using the bacterial expression system we have described. Analysis of viable drug-resistant mutants emerging from clinical trials may offer us new information on the importance of particular subsites in determining substrate specificity in the viral maturation process. Ideally, the process of designing protease inhibitors could incorporate knowledge about subsites in the binding pocket which are prone to mutation to yield second-generation inhibitors whose binding to the enzyme is less susceptible to mutations at these subsites.

A second area related to the issue of substrate specificity is presented by the existence of closely related proteolytic enzymes, such as those from the human

type-2 and simian immunodeficiency viruses. These proteases have been characterized kinetically and found to be distinct from HIV-1 protease (Grant, et al., 1991; Tomasselli, et al., 1990), but have also been shown *in vitro* to process HIV-1 $p55^{gag}$ protein into its normal cleavage products (Grant, et al., 1991; Pichuantes, et al., 1990). Recent mutagenesis studies (Sardana, et al., 1994), and ongoing work in the Craik lab have been aimed at identification of the residues which are differential specificity determinants between the HIV-1 and HIV-2 enzymes. The ability of the HIV-2 and SIV enzymes to substitute for the HIV-1 protease in the viral assembly process might provide us with valuable clues as to the flexibility of substrate recognition in the viral maturation process.

The bacterial expression systems and stabilized recombinant enzymes should assist ongoing studies at UCSF aimed at development of non-peptide inhibitors of the HIV-1 protease (Rutenber, et al., 1993). Site directed mutants have already proven useful in identifying the modes of action of various irreversible inhibitors derived from initial small molecule leads (Salto, et al., 1994). Current efforts aimed at obtaining new crystal structures of the HIV-1 and SIVmac proteases complexed with lead compounds have been facilitated by the availability of the Q7K and S4H variants, respectively. The bacterial expression systems may also be useful in identifying inactive mutants of the HIV-1 protease which will serve as dominant negative inhibitors of protease function. If expression of mutant protease monomers in bacteria also harboring a wild type expression construct is sufficient to inactivate the protease, it is possible that the toxicity of the protease will be prevented, thus giving a screen for mutants which have dominant negative effects. This rapid screen would also be useful in testing whether mutants of the wild type protease can be generated which will be resistant to dominant-negative inhibition. These mutants could be identified by saturation

mutagenesis of the wild type protease, followed by screening in cells stably expressing a dominant negative protease mutant to look for cells in which protease toxicity is restored.

The optimization of genetically-administered defective protease monomers as inhibitors of the HIV-1 protease remains a priority. Although our preliminary data suggest that such an approach could be highly effective, several issues remain to be addressed. Alternative promoters or additional genetic manipulations may be able to enhance the expression level of defective monomers. Further mutations within the defective protease monomers can be envisioned which would increase their ability to interact preferentially with wild type monomers. These mutations could exploit charge-charge interactions at the active site, or provide hydrophobic residues so placed as to interact favorably with the substrate binding pockets in a wild type monomer. Possible synergistic mutations in other viral enzymatic functions (RT, RNAse H, IN) may also be useful in combination with protease mutants.

The development of the effects we have observed in tissue culture into a useful therapeutic application will require that a number of important issues be addressed. Further studies on the ability of cell lines expressing dominant negative mutants to block the late steps of the viral life cycle in an actual infection scenario (i.e. viral infection of CD4+ cells vs. transfection) may provide us with more information on the level of defective monomer which would be adequate in a therapeutic application. The availability of an animal model, such as the SCID-hu mouse, will allow for a test of this strategy in a truly *in-vivo* setting. This will be important in addressing safety issues regarding the immunogenicity of the defective monomer and the best method for delivering therapeutic genes to cells which are the targets of HIV infection.

Bibliography

1. Anderwont, H. B. and Bryan, W. R. (1944) Properties of the mouse mammary-tumor agent. *J. Natl. Caner Inst.* **5** 143 - 149.

2. Babé, L. M. and Craik, C. S. (1992) Synthetic "interface" peptides alter dimeric assembly of the HIV1 and 2 proteases. *Protein Science* **1** 1244 - 1253.

3. Babé, L. M. and Craik, C. S. (1994) Constitutive production of nonenveloped human immunodeficiency virus type 1 particles by a mammalian cell line and effects of a protease inhibitor on particle maturation. *Antimicrob. Agents Chemother.* **38** (10), 2430 - 2439.

4. Babé, L. M., Pichuantes, S., Barr, P. J., Bathurst, I. C., Masiarz, F. R. and Craik, C. S. (1990) HIV1 protease: bacterial expression, purification and characterization. In *Proteins and Pharmaceutical Engineering.* Craik, C. S., Fletterick, R., Matthews, C.R., and Wells, J., eds. Vol. 110, Wiley-Liss Press, New York, 71-88.

5. Babé, L. M., Pichuantes, S. and Craik, C. S. (1991) Inhibition of HIV protease by heterodimer formation. *Biochem.* **30** 106-111.

6. Baboonian, C., Dalgleish, A., Bountiff, L., Gross, J., Oroszlan, S., Rickett, G., Smith-Burchnell, C., Troke, P. and Merson, J. (1991) HIV-1 proteinase is required for synthesis of pro-viral DNA. *Biochem. Biophys. Res. Commun.* **179** (1), 17 - 24.

7. Baim, S. B., Pietras, D. F., Eustice, D. C. and Sherman, F. (1985) A Mutation Allowing an mRNA Secondary Structure Diminishes Translation of *Saccharomyces cerevisiae* Iso-1-Cytochrome C. *Mol. Cell Biol.* **5** (8), 1839 - 1846.

8. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226** 1209 - 1211.

9. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J. G., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220** 868 - 871.

10. Baum, E. Z., Bebernitz, G. A. and Gluzman, Y. (1990) Isolation of mutants of human immunodeficiency virus protease based on the toxicity of the enzyme in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **87** 5573 - 5577.

11. Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A. and Chang, S. (1987) Processing of the Initiation Methionine from Proteins: Properties of the *Escherichia coli* Methionine Aminopeptidase and Its Gene Structure. *J. Bacteriol.* **169** (2), 751-757.

12. Berkhout, B., Schmidt, B. F., van Strien, A., van Boom, J., van Westernen, J. and van Duin, J. (1987) Lysis Gene of Bacteriophage MS2 is Activated by

Translation Termination at the Overlapping Coat Gene. *J. Mol. Biol.* **195** 517 - 524.

13. Bittner, J. J. (1936) Some possible effects of nursing on the mammary gland tumor incidence in mice. *Science* **84** 162.

14. Blundell, T. and Pearl, L. (1989) A second front against AIDS. *Nature* **337** 596 - 597.

15. Braxton, S. and Wells, J. A. (1992) Incorporation of a stabilizing Ca^{2+}-binding loop into subtilisin BPN'. *Biochem.* **31** 7796 - 7801.

16. Browner, M. F., Rasor, P., Tugendreich, S. and Fletterick, R. J. (1991) Temperature-Sensitive Production of Rabbit Muscle Glycogen Phosphorylase in *Escherichia coli*. *Protein Engineering* **4** (3), 351-357.

17. Button, L. L., Reiner, N. E. and McMaster, R. W. (1991) Modification of GP63 Genes From Diverse Species of *Leishmania* for Expression of Recombinant Protein at High Levels in *Escherichia coli*. *Mol. Biochem. Parasitol.* **44** 213-224.

18. Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M. A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L. and Montagnier, L. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233** 343-346.

19. Co, E., Koelsch, G., Lin, Y., Ido, E., Hartsuck, J. A. and Tang, J. (1994) Proteolytic processing mechanisms of a miniprecursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry* **33** (5), 1248 -1254.

20. Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. and Weiss, R. (1986) Human Immunodeficiency Viruses. *Science* **232** 697.

21. Craven, R. C., Bennett, R. P. and Wills, J. W. (1991) Role of the avian retroviral protease in the activation of reverse transcriptase during virion assembly. *J. Virol.* **65** (11), 6205 - 6217.

22. Crawford, S. and Goff, S. P. (1985) A deletion mutation in the 5' part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polyproteins. *J. Virol.* **53** 899 - 907.

23. Daniel, M. D., Letvin, N. L., King, N. W., Kannagi, M., Sehgal, P. K., Hunt, R. D., Kanki, P. J., Essex, M. and Desrosiers, R. C. (1985) Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* **228** 1201-1204.

24. Darke, P. L., Leu, C. T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Holl, W. S., Dixon, R. A. and Sigal, I. S. (1989) Human immunodeficiency virus protease: Bacterial expression and characterization of the purified aspartic protease. *J. Biol. Chem.* **264** 2307-2312.

25. Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C.-T., Lumma, P. K., Freidinger, R. M., Veber, D. F. and Sigal, I. S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. *Biochem. Biophys. Res. Commun.* **156** (1), 297 - 303.

26. deBoer, H. A., Comstock, L. J. and Vasser, M. (1983) The *tac* Promoter: A Functional Hybrid Derived From the *trp* and *lac* Promoters. *Proc. Natl. Acad. Sci. USA* **80** 21-25.

27. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W. and Rosenberg, M. (1987) Human immunodeficiency virus protease expressed in *Escherischia coli* exhibits autoprocessing and specific maturation of the gag precursor. *Proc. Natl. Acad. Sci. U.S.A.* **84** 8903-8906.

28. Desrosiers, R. C., Daniel, M. D. and Li, Y. (1989) HIV-related lentiviruses of nonhuman primates. *Aids Res. Hum. Retrovir.* **5** (5), 465 - 473.

29. Devlin, P. E., Drummond, R. J., Toy, P., Mark, D. F., Watt, K. W. K. and Devlin, J. J. (1988) Alteration of amino-terminal codons of human granulocyte-colony-stimulating factor increases expression levels and allows efficient processing by methionine aminopeptidase in *Escherichia coli*. *Gene* **65** 13-22.

30. Dittmar, K. J. and Moelling, K. (1978) Biochemical properties of p15-associated protease in an avian RNA tumor virus. *J. Virol.* **28** (1), 106 - 188.

31. Doolittle, R. F., Feng, D. F., McClure, M. A. and Johnson, M. S. (1990) Retrovirus phylogeny and evolution. In *Current Topics in Microbiology and Immunology*. eds. Vol. 157, Springer-Verlag, Heidelberg, 1 - 18.

32. Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H. and Craik, C. S. (1992) The sequence, organization, and expression of the major cysteine protease (Cruzain) from *Trypanosoma cruzi*. *J. Biol. Chem.* **267** (11), 7411 - 7420.

33. El-Farrash, M., Kuroda, M., Kitazaki, T., Masuda, T., Kato, K., Hatanaka, M. and Harada, S. (1994) Generation and characterization of a human immunodeficiency virus type 1 (HIV-1) mutant resistant to an HIV-1 protease inhibitor. *J. Virol.* **68** (1), 233 - 239.

34. Erickson-Viitanen, S., Manfredi, J., Viitanen, P., Tribe, D. E., Tritch, R., Hutchinson, C. A., Loeb, D. and Swanstrom, R. (1989) Cleavage of HIV-1 *gag* polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res. Hum. Retrovir.* **5** (6), 577 - 591.

35. Feinberg, M. B. and Trono, D. (1992) Intracellular Immunization: *Trans*-Dominant Mutants of HIV Gene Products as Tools for the Study and Interruption of Viral Replication. *Aids. Res. Hum. Retrovir.* **8** (6), 1013 - 1022.

36. Frorath, B., Abney, C., Berthold, H., Scanarini, M. and Northemann, W. (1992) Production of Recombinant Rat Interleukin-6 in *Escherichia coli* Using a Novel Highly Efficient Expression Vector pGEX-3T. *BioTechniques* **12** (4), 558 - 563.

37. Fu, W., Gorelinck, R. J. and Rein, A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.* **68** (8), 5013 - 5018.

38. Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., Mann, D., Sidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J. and Popovic, M. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220** 865 - 867.

39. Giam, C. Z. and Boros, I. (1988) *In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *Escherichia coli*. *J. Biol. Chem.* **263** 14617-14620.

40. Goff, S. A. and Goldberg, A. L. (1985) Production of Abnormal Proteins in *E. coli* Stimulates Transcription of *lon* and Other Heat Shock Genes. *Cell* **41** 587-595.

41. Göttlinger, H. G., Sodroski, J. G. and Haseltine, W. A. (1989) Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **86** 5781 - 5785.

42. Gowda, S. J., Stein, B. S. and Engleman, E. G. (1989) Identification of protein intermediates in the processing of the p55 HIV-1 gag precursor in cells infected with recombinant vaccinia virus. *J. Biol. Chem.* **264** (15), 8459 - 8462.

43. Graham, F. L. and van der Eb, J. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. *J. Virol.* **63** 2527 - 2533.

44. Grant, S. K., Deckman, I. C., Minnich, M. D., Culp, J., Franklin, S., Dreyer, G. B., Tomaszek T.A., J., Debouck, C. and Meek, T. D. (1991) Purification and biochemical characterization of recombinant simian immunodeficiency virus protease and comparison to human immunodeficiency virus type 1 protease. *Biochemistry* **30** (34), 8424 - 8434.

45. Graves, M., Lim, J., Heimer, E. and Kramer, R. (1988) An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. *Proc. Natl. Acad. Sci. USA* **85** 2449 - 2453.

46. Grinde, B., Cameron, C. E., Leis, J., Weber, I. T., Wlodawer, A., Burstein, H., Bizub, D. and Skalka, A. M. (1992) Mutations that alter the activity of the rous sarcoma virus protease. *J. Biol. Chem.* **267** (14), 9481 - 9490.

47. Gross, L. (1951) "Spontaneous" leukemia developing in C3H mice following inoculation, in infancy, with AK-leukemic extracts, or AK-embryos. *Proc. Soc. Exp. Biol. Med.* **76** 27 - 32.

48. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montaigner, L. and Alizon, M. (1987) Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* **326** 662-669.

49. Guzman, C. A., Walker, M. J., Rohde, M. and Timmis, K. N. (1992) Expression of Bordetella pertussis filamentous hemagglutinin in *Escherichia coli* using a two cistron system. *Microbial Pathogenesis* **12** (5), 383 - 389.

50. Hall, M. N., Gabay, J., Débarbouillé and Schwartz, M. (1982) A Role for mRNA Secondary Structure in the Control of Translation Initiation. *Nature* **285** 616-618.

51. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Bio.* **166** 557-580.

52. Hawley, D. K. and McClure, W. R. (1983) Compilation and Analysis of *Escherichia coli* Promoter DNA Sequences. *Nucleic Acids Res.* **11** (8), 2237-2255.

53. Herskowitz, I. (1987) Functional inactivation of genes by dominant negative mutations. *Nature* **329** 219-222.

54. Higaki, J. N., Haymore, B. L., Chen, S., Fletterick, R. J. and Craik, C. S. (1990) Regulation of Serine Protease Activity by an Engineered Metal Switch. *Biochemistry* **29** 8582 - 8586.

55. Ho, D. D., Toyoshima, T., Mo, H., Kempf, D. J., Norbeck, D., Chen, C.-M., Wideburg, N. E., Burt, S. K., Erickson, J. W. and Singh, M. K. (1994) Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C_2-symmetric protease inhibitor. *J. Virol.* **68** (3), 2016 - 2020.

56. Hunkapillar, M. W., Hewich, R. M., Dreyer, W. J. and Hood, L. E. (1983) *Methods Enzymol.* **91** 399 - 413.

57. Hyland, L. J., Tomaszek Jr, T. A. and Meek, T. D. (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanisms. *Biochemistry* **30** 8454 - 8463.

58. Ido, E., Han, H.-P., Kezdy, F. J. and Tang, J. (1991) Kinetic Studies of Human Immunodeficiency Virus Type 1 Protease and Its Active-site Hydrogen Bond Mutant A28S. *J. Biol. Chem.* **266** 24359 - 24366.

59. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* **331** 280 - 283.

60. Jacks, T., Townsley, K., Varmus, H. E. and Majors, J. (1987) Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. *Proc. Natl. Acad. Sci. U.S.A.* **84** (12), 4298 - 4302.

61. Jacks, T. and Varmus, H. E. (1985) Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science* **230** 1237 - 1242.

62. Jacobsen, H., Ahlborn-Laake, L., Gugel, R. and Mous, J. (1992) Progression of early steps of human immunodeficiency virus type 1 replication in e presence of an inhibitor of viral protease. *J. Virol.* **66** (8), 5087 - 5091.

63. Jarrett, W. F. H., Crawford, E. M., Martin, W. B. and Davie, F. (1964) A virus-like particle associated with leukaemia (lymphosarcoma). *Nature* **202** 567 - 568.

64. Jupp, R. A., Phylip, L. H., Mills, J. S., Le Grice, S. F. J. and Kay, J. (1991) Mutating P2 and P1 residues at cleavage junctions in the HIV-1 *pol* polyprotein. *FEBS* **283** (2), 180 - 184.

65. Kanki, P. J., McLane, M. F., King, N. W., Letvin, N., Hunt, R. D., Sehgal, P., Daniel, M. D., Desrosiers, R. C. and Essex, M. (1985) Serologic identification and characterization of a macaque T-lymphotropic retrovirus closely related to HTLV-III. *Science* **228** 1199-1201.

66. Kaplan, A. and Swanstrom, R. (1991) Human immunodeficiency virus type 1 gag proteins are processed in two cellular compartments. *Proc. Natl. Acad. Sci. USA* **88** 4528 - 4532.

67. Kaplan, A. H., Zack, J. A., Knigge, M., Paul, D. A., Kempf, D. J., Norbeck, D. W. and Swanstrom, R. (1993) Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *J. Virol.* **67** (7), 4050 - 4055.

68. Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. and Moss, B. (1993) Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193** 661 - 671.

69. Karlström, A. R., Shames, B. D. and Levine, R. L. (1993) Reactivity of cysteine residues in the protease from human immunodeficiency virus: identification of a surface-exposed region which affects enzyme function. *Arch. Biochem. Biophys.* **304** 163-169.

70. Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T. and Oroszlan, S. (1985) Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145** 280 - 292.

71. Kennell, D. and Riezman, H. (1977) Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. *J. Mol. Biol.* **114** 1-21.

72. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. and Sigal, I. S. (1988) Active human

immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **85** 4686-4690.

73. Konigsberg, W. and Godson, G. N. (1983) Evidence For Use of Rare Codons in the *dnaG* Gene and Other Regulatory Genes of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **80** 687-691.

74. Korant, B. D. and Rizzo, C. J. (1991) An E. coli expression system which detoxifies the HIV protease. *Biomed. Biochim. Acta*. **50 (4-6)**, 643 - 646.

75. Kotula, L. and Curtis, P. J. (1991) Evaluation of Foreign Codon Optimization in Yeast: Expression of a Mouse Ig Kappa Chain. *Bio/Technology* **9** 1386-1389.

76. Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F. and Reddy, E. P. (1986) HTLV-III *gag* protein is processed in yeast cells by the virus *pol*-protease. *Science* **231** 1580 - 1584.

77. Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154** 367 - 382.

78. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **277** 680 - 682.

79. Lambert, D. M., Petteway, S. R., McDanal, C. E., Hart, T. K., Leary, J. J., Dreyer, G. B., Meek, T. D., Bugelski, P. J., Bolognesi, D. P., Metcalf, B. W. and

Matthews, T. J. (1992) Human immunodeficiency virus type 1 protease inhibitors irreversibly block infectivity of purified virions from chronically infected cells. *Antimicrob. Agents Chemother.* **36** (5), 982 - 988.

80. Larder, B. A. (1993) Inhibitors of HIV reverse transcriptase as antiviral agents and drug resistance. In *Reverse Transcriptases*. Skalka, A. M. and Goff, S. P., eds. Vol. Cold Spring Harbor Laboratory, New York, 205 - 222.

81. Letvin, N. L., Daniel, M. D., Sehgal, P. K., Desrosiers, R. C., Hunt, R. D., Waldron, L. M., MacKey, J. J., Schmidt, D. K., Chalifoux, L. V. and King, N. W. (1985) Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III. *Science* **230** 71 - 73.

82. Levin, J. G., Hu, S. C., Rein, A., Messer, L. I. and Gerwin, B. I. (1984) Murine leukemia virus mutant with a frameshift in the reverse transcriptase coding region: implications for *pol* gene structure. *J. Virol.* **51** (2), 470 - 478.

83. Liebhaber, S. A., Cash, F. and Eshleman, S. S. (1992) Translation Inhibition by an mRNA Coding Region Secondary Structure is Determined by Its Proximity to the AUG Initiation Codon. *J. Mol. Biol.* **226** 609 - 621.

84. Lillehoj, E. P., Salazar, F. H. R., Mervis, R. J., Raum, M. G., Chan, H. W., Ahmad, N. and Venkatesan, S. (1988) Purification and structural characterization of the putative gag-pol protease of human immunodeficiency virus. *J. Virol.* **62** 3053-3058.

85. Loeb, D. D., Hutchison, C. A., Edgell, M. H., Farmerie, W. G. and Swanstrom, R. (1989) Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *J. Virol.* **63** 111-121.

86. Louis, J. M., Smith, C. A. D., Wondrak, E. M., Mora, P. T. and Oroszlan, S. (1989) Substitution mutations of the highly conserved arginine 87 of HIV-1 protease result in loss of proteolytic activity. *Biochem. Biophys. Res. Commun.* **164** (1), 30 - 38.

87. Makoff, A. J. and Smallwood, A. E. (1990) The use of two-cistron constructions in improving the expression of a heterologous gene in *E. coli*. *Nucleic Acids Res.* **18** (7), 1711 - 1718.

88. Margolin, N., Heath, W., Osborne, E., Lai, M. and Vlahos, C. (1990) Substitutions at the P2' site of *gag* p17-p24 affect cleavage efficiency by HIV-1 protease. *Biochem. Biophys. Res. Comm.* **167** (2), 554 - 560.

89. Martin, L. N., Soike, K. F., Murphey-Corb, M., Bohm, R. P., Roberts, E. D., Kakuk, T. J., Thaisrivongs, S., Vidmar, T. J., Ruwart, M. J., Davio, S. R. and Tarpley, W. G. (1994) Effects of U-75875, a peptidomimetic inhibitor of retroviral proteases, on simian immunodeficiency virus infection in rhesus monkeys. *Antimicrob. Agents Chemother.* **38** (6), 1277 - 1283.

90. Matsumura, P., Rydel, J. J., Linzmeier, R. and Vacante, D. (1984) Overexpression and sequence of the *Escherichia coli cheY* gene and biochemical activities of the CheY protein. *J. Bacteriol.* **160** (1), 36 - 41.

91. McKeating, J. A., McKnight, A. and Moore, J. P. (1991) Differential loss of envelope glycoprotein gp120 from virions of human immunodeficiency virus type 1 isolates: effects on infectivity and neutralization. *J. Virol.* **65** (2), 852 - 860.

92. McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrikson, R. L. and Tarpley, W. G. (1990) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* **247** 454-456.

93. Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M. L., Magaard, V. W. and Debouck, C. (1989) Human immunodeficiency virus 1 protease expressed in *Escherichia coli* behaves as a dimeric aspartic protease. *Proc. Natl. Acad. Sci. U.S.A.* **86** 1841-1845.

94. Mergener, K., Fäcke, M., Welker, R., Brinkmann, V., Gelderblom, H. R. and Kräusslich, H.-G. (1992) Analysis of HIV particle formation using transient expression of subviral constructs in mammalian cells. *Virology* **186** 25 - 39.

95. Mermer, B., Malamy, M. and Coffin, J. (1983) Rous sarcoma virus contains sequences which permit expression of the *gag* gene in *Escherichia coli. Mol. Cell. Biol.* **3** (10), 1746 - 1758.

96. Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J., Tomich, C.-S. C., Smith, C. W., Heinrikson, R. L. and Tomasselli, A. G. (1994) The HIV-1 protease as enzyme and substrate - mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry* **33** (32), 9405 - 9413.

97. Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J. and Wlodawer, A. (1989) Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337** 576-579.

98. Mous, J., Heimer, E. P. and LeGrice, S. F. J. (1988) Processing protease and reverse transcriptase from human immunodeficiency virus type 1 polyprotein in *Escherichia coli. J. Virol.* **62** 1433 - 1440.

99. Nagy, K., Young, M., Baboonian, C., Merson, J., Whittle, P. and Oroszlan, S. (1994) Antiviral activity of human immunodeficiency virus type 1 protease inhbitors in a single cycle of infection: evidence for a role of protease in the early phase. *J. Virol.* **68** (2), 757 - 765.

100. Nassal, M., Mogi, T., Karnik, S. S. and Khorana, H. G. (1987) Structure-Function Studies on Bacteriorhodopsin. III. Total Synthesis of a Gene for

Bacterio-opsin and its Expression in *Escherichia coli*. *J. Biol. Chem.* **262** (19), 9264-9270.

101. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. and Springer, J. P. (1989) Three-dimensional structure of the aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337** 615-620.

102. Nilsson, G., Belasco, J. G., Cohen, S. N. and von Gabain, A. (1984) Growth-rate Dependent Regulation of mRNA Stability in *Escherichia coli*. *Nature* **312** 75-77.

103. North, M. J. (1989) Prevention of Unwanted Proteolysis. In *Proteolytic enzymes: a practical approach*. Reynon, R. J. and Bond, J. S., eds. Vol. IRL Press, Oxford, 105 - 124.

104. Otto, M. J., Garber, S., Winslow, D. L., Reid, C. D., Aldrich, P., Jadhav, P. K., Patterson, C. E., Hodge, C. N. and Cheng, Y.-S. E. (1993) *In Vitro* isolation and identification of human immunodeficiency virus (HIV) variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proc. Natl. Acad. Sci. USA* **90** 7543 - 7547.

105. Page, K. A., Landau, R. N. and Littman, D. R. (1990) Construction and use of an immunodeficiency virus vector for analysis of virus infectivity. *J. Virol.* **64** 5270-5376.

106. Park, J. and Morrow, C. D. (1992) The nonmyristylated Pr160$^{gag\text{-}pol}$ polyprotein of human immunodeficiency virus type 1 interacts with Pr55^{gag} and is incorporated into viruslike particles. *J. Virol.* **66** (11), 6304 - 6313.

107. Park, J. and Morrow, C. D. (1993) Mutations in the protease gene of human immunodeficiency virus type 1 affect release and stability of virus particles. *Virology* **194** (2), 843 - 850.

108. Partin, K., Kräusslich, H. G., Ehrlich, L., Wimmer, E. and Carter, C. (1990) Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase. *J. Virol.* **64** (8), 3938 - 3947.

109. Pearl, L. H. and Taylor, W. R. (1987) A structural model for the retroviral proteases. *Nature* **329** 351-354.

110. Peng, C., Ho, B. K., Chang, T. W. and Chang, N. T. (1989) Role of human immunodeficiency virus type 1 specific protease in core protein maturation and viral infectivity. *J. Virol.* **63** 2550-2556.

111. Pichuantes, S., Babé, L., Barr, P. and Craik, C. S. (1989) Expression of authentic and enzymatically active HIV1 PR in bacteria and yeast. In *Current Communications in Molecular Biology: Viral Proteinases as Targets for Chemotherapy*. eds. Vol. E. Wimmer, Cold Spring Harbor, 215-222.

112. Pichuantes, S., Babé, L. M., Barr, P. J. and Craik, C. S. (1989) Recombinant HIV1 protease secreted by Saccharomyces cerevisiae correctly processes myristylated gag polyprotein. *Proteins: Struct., Funct., Genet.* **6** 324-337.

113. Pichuantes, S., Babé, L. M., Barr, P. J., DeCamp, D. L. and Craik, C. S. (1990) Recombinant HIV2 protease processes HIV1Pr53gag and analogous junction peptides in vitro. *J. Biol. Chem.* **265** 13890-13898.

114. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. and Gallo, R. C. (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Aca. Sci. U.S.A.* **77** 7415-7419.

115. Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L. and Kézdy, F. J. (1991) A cumulative specificity model for proteases from immunodeficiency virus types 1 and 2 inferred from statistical analysis of an extended substrate data base. *J. Biol. Chem.* **266** 14554-14561.

116. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway Jr., S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. and Wong-Staal, F. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313** 277-284.

117. Reznikoff, W. and Gold, L., eds (1986), "Maximizing Gene Expression" Butterworths, Boston

118. Rosé, J. R., Babe, L. M. and Craik, C. S. (1995) Defining the level of protease activity required for HIV-1 particle maturation and infectivity. *J. Virol., in press.*

119. Rosé, J. R., Salto, R. and Craik., C. S. (1993) Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J. Biol. Chem.* **268** (16), 11939 - 11945.

120. Rose, R., Rosé, J., Salto, R., Craik, C. and Stroud, R. (1993) Structure of the protease from simian immunodeficiency virus: complex with an irreversible nonpeptide inhibitor. *Biochemistry* **32** 12498 - 12507.

121. Rous, P. (1910) A transmissible avian neoplasm (sarcoma of the common fowl). *J. Exp. Med.* **12** 696 - 705.

122. Rutenber, E., Fauman, E., Keenan, R., Fong, S., Furth, P., Ortiz de Montellano, P., Meng, E., Kuntz, I., DeCamp, D., Salto, R., Rosé, J., Craik, C. and Stroud, R. (1993) Structure of a non-peptide inhibitor complexed with the HIV-1 protease. *J. Biol. Chem.* **268** (21), 15343 - 15346.

123. Salto, R., Babé, L. M., Li, J., Rosé, J. R., Yu, Z., Burlingame, A., De Voss, J. J., Ortiz de Montellano, P. and Craik, C. S. (1994) *In vitro* characterization of nonpeptide irreversible inhibitors of HIV proteases. *J. Biol. Chem.* **269** 10691 - 10698.

124. Sardana, V. V., Schlabach, A. J., Graham, P., Bush, B. L., Condra, J. H., Culberson, J. C., Gotlib, L., Graham, D. J., Kohl, N. E., LaFemina, R. L., Schneider, C. L., Wolanski, B. S., Woldfgang, J. A. and Emini, E. A. (1994) Human immunodeficiency virus type 1 protease inhibitors: evaluation of resistance engendered by amino acid substitutions in the enzyme's substrate binding site. *Biochemistry* **33** 2004 - 2010.

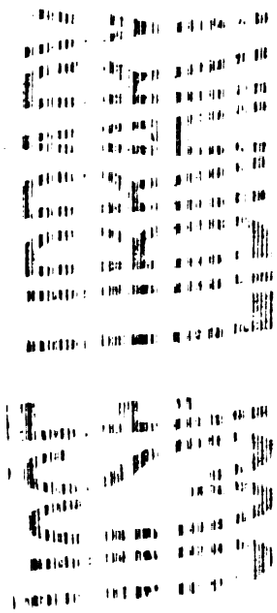
125. Schätzl, H., Gelderblom, H. R., Nitschko, H. and von der Helm, K. (1991) Analysis of non-infectious HIV particles produced in presence of HIV proteinase inhibitor. *Arch. Virol.* **120** (1-2), 71 - 81.

126. Schechter, I. and Berger, A. (1967) On the size of the active site in protease. I. Papain. *Biochem. Biophys, Res. Commun.* **27** 157.

127. Schoner, B. E., Belagaje, R. M. and Schoner, R. G. (1986) Translation of a Synthetic Two-Cistron mRNA in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **83** 8506-8510.

128. Schoner, B. E., Hsiung, H. M., Belagaje, R. M., Mayne, N. G. and Schoner, R. G. (1984) Role of mRNA translational efficiency in bovine growth hormone expression in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81** 5403 - 5407.

129. Schüpbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G. and Gallo, R. C. (1984) Serlogical analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224** 503 - 505.



130. Schwartz, D. E., Tizard, R. and Gilbert, W. (1983) Nucleotide sequence of Rous sarcoma virus. *Cell* **32** 853 - 869.

131. Sen, G. C., Haspel, H. C. and Sarkar, N. H. (1980) Presence of a proteolytic activity in murine mammary tumor virus. *J. Biol. Chem.* **255** (15), 7098 - 7101.

132. Sharma, S. K. and Hopkins, T. R. (1981) Recent Developments in the Activation Process of Bovine Chymotrypsinogen A. *Bioorg. Chem.* **10** 357 - 374.

133. Shepard, H. M., Yelverton, E. and Goeddel, D. V. (1982) Increased Synthesis in *E. coli* of Fibroblast and Leukocyte Interferons Through Alterations in Ribosome Binding Sites. *DNA* **1** (2), 125 - 131.

134. Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71** 1342 - 1346.

135. Sigal, E., Grunberger, D., Highland, E., Gross, C., Dixon, R. A. and Craik, C. S. (1990) Expression of cloned human reticulocyte 15-lipoxygenase and immunological evidence that 15-lipoxygenases of different cell types are related. *J. Biol. Chem.* **265** 5113-5120.

136. Smith, R. L. and Shaw, E. (1969) Pseudotrypsin: A Modified Bovine Trypsin Produced by Limited Autodigestion. *J. Biol. Chem.* **244** (17), 4704 - 4712.

137. Spanjaard, R. A., van Dijk, M. C. M., Turion, A. J. and van Duin, J. (1989) Expression of the rat interferon-α_1 gene in *Escherichia coli* controlled by the secondary structure of the translation-initiation region. *Gene* **80** 345 - 351.

138. Steitz, J. A. (1979) In *Biological Regulation and Development.* Goldberger, R. F., eds. Vol. 1, Plenum Press, New York, 349.

139. Stewart, L., Schatz, G. and Vogt, V. M. (1990) Properties of avian retrovirus particles defective in viral protease. *J. Virol.* **64** (10), 5076 - 5092.

140. Stormo, G. D., Schneider, T. D. and Gold, L. M. (1982) Characterization of Translational Initiation Sites in *E.coli. Nucleic Acids Res.* **10** (9), 2971-2996.

141. Straney, S. B. and Crothers, D. M. (1987) Kinetics of the stages of transcription initiation at the Escherichia coli *lac UV5* promoter. *Biochemistry* **26** (16), 5063-5070.

142. Strickler, J. E., Gorniak, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, J. and Debouck, C. (1989) Characterization and autoprocessing of precursor and mature forms of human immunodeficiency virus type 1 (HIV 1) protease purified from *Escherichia coli. Proteins* **6** 139 - 154.

143. Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Kent, S. B. H. and Wlodawer, A. (1990) X-ray crystallographic structure of a complex between

HIV-1 protease and a hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. USA* **87** 8805-8809.

144. Tang, J. (1971) Specific and irreversible inactivation of pepsin by substrate-like epoxides. *J. Biol. Chem.* **246** 4510-4517.

145. Taniguchi, T. and Weissmann, C. (1978) Site-Directed Mutations in the Initiator Region of the Bacteriophage Qβ Coat Cistron and Their Effect on Ribosome Binding. *J. Mol. Biol.* **118** 533-565.

146. Temin, H. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226** 1211 - 1213.

147. Temin, H. M. (1964) Nature of the provirus of Rous sarcoma virus. *Natl. Cancer Inst., Monogr.* **17** 557 - 570.

148. Toh, H., Ono, M., Saigo, K. and Miyata, T. (1985) Retroviral protease-like sequence in the yeast transposon Ty*1*. *Nature* **315** 691 - 692.

149. Tomasselli, A., Hui, J. O., Sawyer, T. K., Staples, D. J., Bannow, C., Reardon, I. M., Howe, W. J., DeCamp, D. L., Craik, C. S. and Heinrikson, R. L. (1990) Specificity and inhibition of proteases from human immunodefiency viruses 1 and 2. *J. Biol. Chem.* **265** 14675-14683.

150. Tomasselli, A. G., Howe, W. J., Sawyer, T. K., Wlodawer, A. and Heinrikson, R. L. (1991) The complexities of AIDS: an assessment of the HIV protease as a therapeutic target. *Chimica Oggi* **May** 6-274.

151. Tomasselli, A. G., Olsen, M. K., Hui, J. O., Staples, D. J., Sawyer, T. K., Heinrikson, R. L. and Tomich, C.-S. (1990) Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease. *Biochem.* **29** 264-269.

152. Tomich, C.-S. C., Olson, E. R., Olsen, M. K., Kaytes, P. S., Rockenback, S. K. and Hatzenbuhler, N. T. (1989) Effect of Nucleotide Sequences Directly Downstream From the AUG on the Expression of Bovine Somatotropin in *E.coli*. *Nucleic Acids Res.* **17** (8), 3179-3197.

153. Vogt, V. M., Wight, A. and Eisenman, R. (1979) *In vitro* cleavage of avian retrovirus *gag* proteins by viral protease p15. *Virology* **98** 154 - 167.

154. von der Helm, K. (1977) Cleavage of Rous sarcoma viral polypeptide precursor into internal structural protein *in vitro* involves viral protein p15. *Proc. Natl. Acad. Sci. USA* **74** (3), 911 - 915.

155. Williams, D. P., Regier, D., Akiyoshi, D., Genbauffe, F. and Murphy, J. R. (1988) Design, Synthesis and Expression of a Human Interleukin-2 Gene Incorporating the Codon Usage Bias Found in Highly Expressed *Escherichia coli* Genes. *Nucleic Acids. Res.* **16** (22), 10453-10467.

[illegible]

[illegible]

[illegible]

156. Wlodawer, A. and Erickson, J. (1993) Structure-Based Inhibitors of HIV-1 Protease. In *Annual Review of Biochemistry*. Richardson, C. C., Abelson, J. N., Meister, A. and Walsh, C. T., eds. Vol. 62, Annual Reviews, Inc., Palo Alto, CA, 543 - 585.

157. Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. and Kent, S. B. H. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245** 616-621.

158. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33** 103 - 119.

159. Yoshinaka, Y., Katoh, I., Copeland, T. D. and Oroszlan, S. (1985) Murine leukemia virus protease in encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. USA* **82** 1618 - 1622.

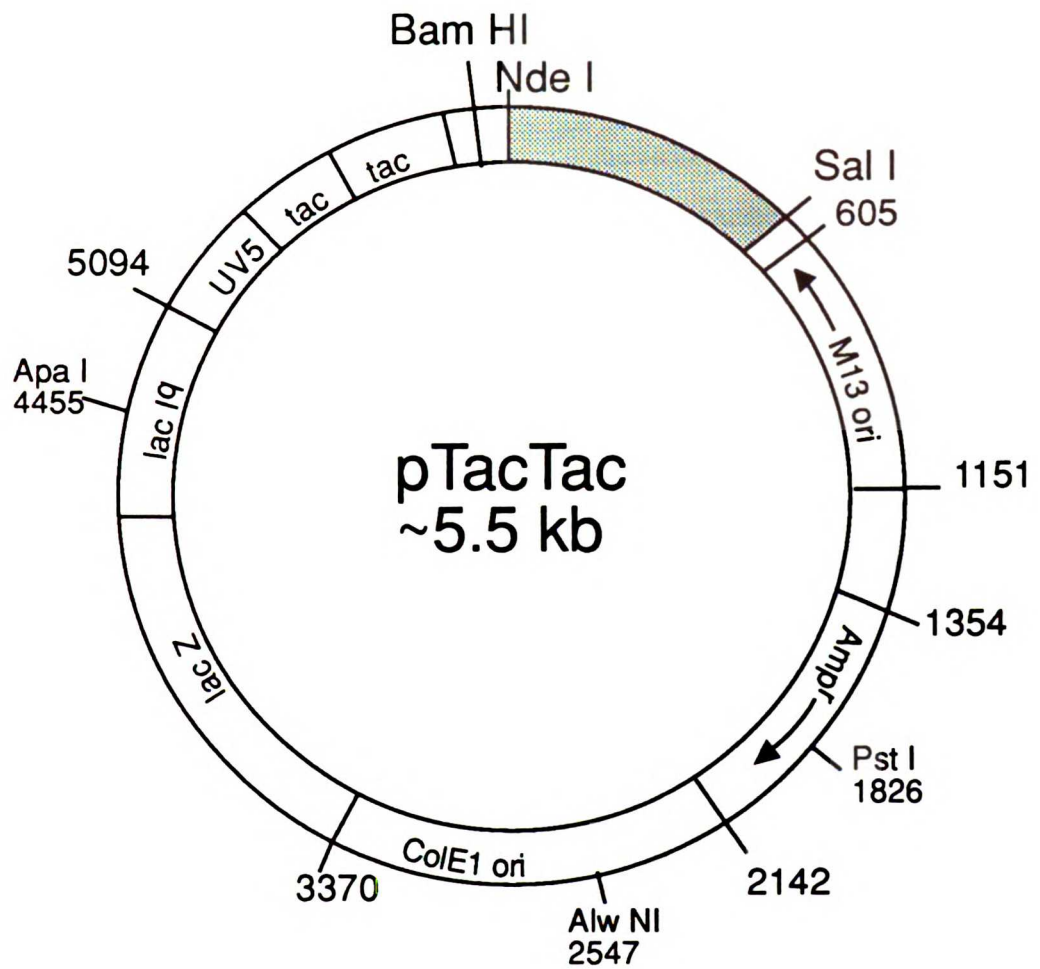
160. Yoshinaka, Y., Katoh, I., Copeland, T. D., Smythers, G. W. and Oroszlan, S. (1986) Bovine leukemia virus protease: purification, chemical analysis, and in vitro processing of *gag* precursor polyproteins. *J. Virol.* **57** (3), 826 - 832.

161. Yoshinaka, Y. and Luftig, R. B. (1977) Properties of a p70 proteolytic factor of murine leukemia viruses. *Cell* **12** 709 - 719.

162. Zhang, A.-Y., Reardon, I. M., Hui, J. O., O'Connell, K. L., Poorman, R. A., Tomasselli, A. G. and Heinrikson, R. L. (1991) Zinc inhibition of renin and the protease from human immunodeficiency virus type 1. *Biochemistry* **30** (36), 8717 - 8721.

163. Zhang, Z. Y., Poorman, R. A., Maggiora, L. L., Heinrikson, R. L. and Kézdy, F. J. (1991) Dissociative inhibition of dimeric emzymes. *J. Biol. Chem.* **266** (24), 15591-15594.

Appendix A. Nucleotide Sequences and Maps of Expression Plasmids and Genes



nt 1- 508: native gene (HIV genes inserted from nt 1 - 514)

nt 508: Xba I; nt 514: Sal I; nt 529: Hind III

nt 569 - 596: *trpa* transcription terminator

nt 605 - 1152: M13 origin

nt 1152 - 3369: pBR322 backbone (4286 - 2066)

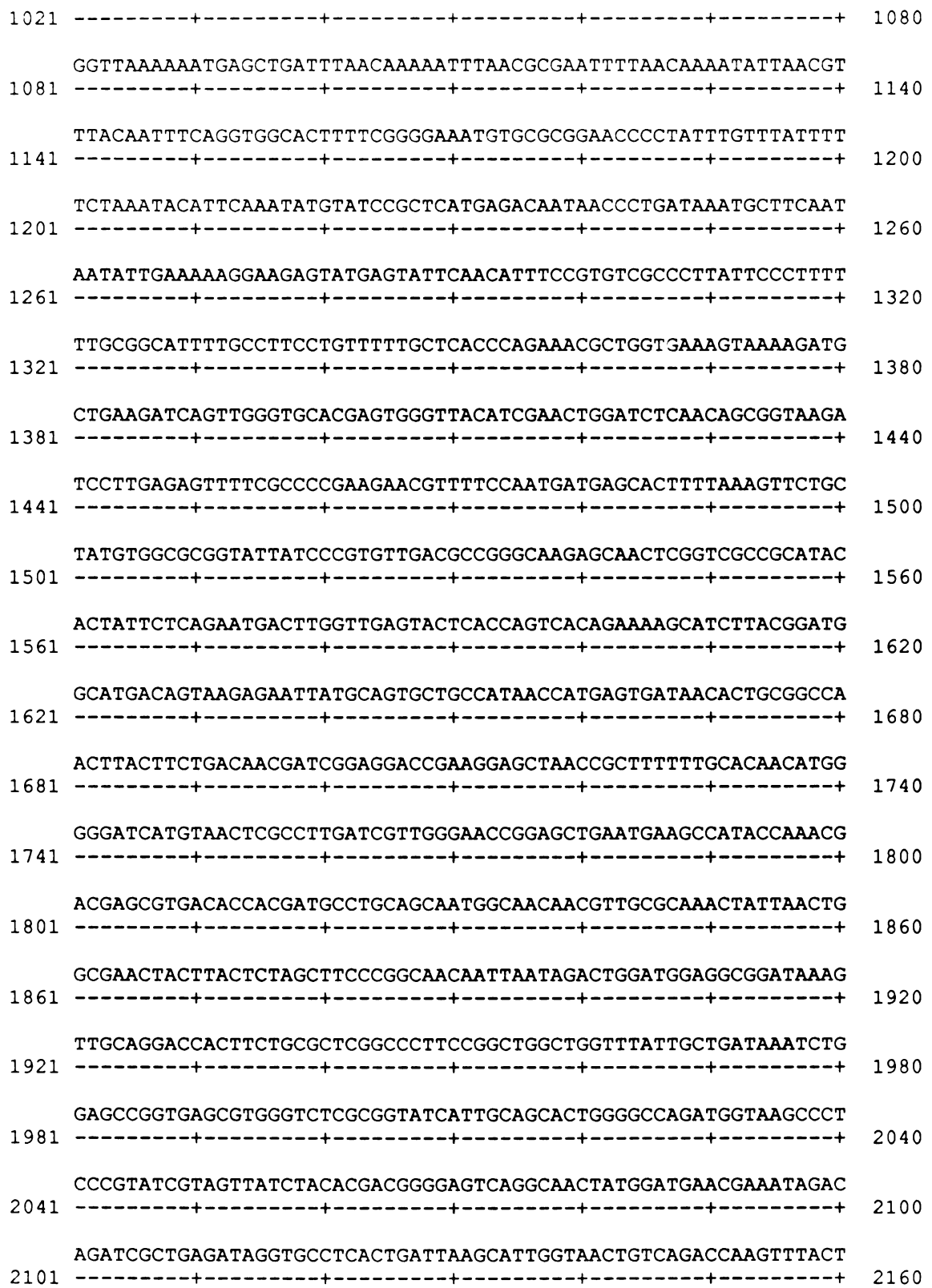
nt 3370 - 5094: lac Z, lacIq fragment

nt 5095 - 5453: promoter cassette (lacUV5, tac, tac)

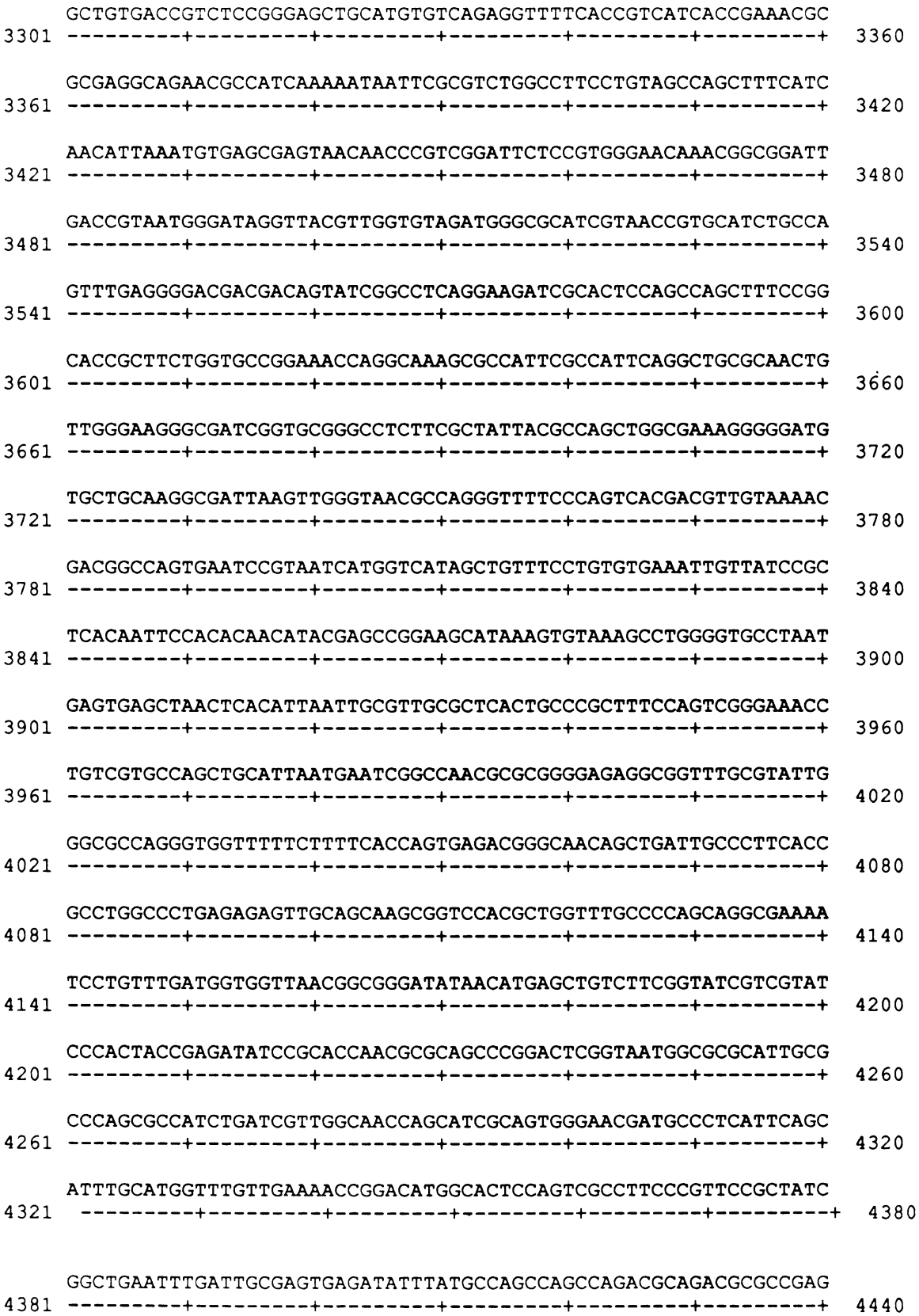
nt 5454 - Bam HI

Nucleotide Sequence of native pTacTac plasmid

```
      ATGACCGGTATGACGAATGTAACAAAGCTGGCCAGCGAGCCGTCAGGCCAGGAATTTCTG
    1 ---------+---------+---------+---------+---------+---------+    60

      GTATTTACCCTTGGTGATGAAGAGTACGGTATTGATATCCTGAAAGTGCAGGAGATCCGT
   61 ---------+---------+---------+---------+---------+---------+   120

      GGCTACGATCAGGTAACACGGATTGCGAACACGCCAGCGTTTATCAAAGGCGTCACGAAT
  121 ---------+---------+---------+---------+---------+---------+   180

      CTGCGCGGCGTTATTGTGCCGATTGTTGACTTACGAATTAAGTTCAGCCAGGTGGATGTG
  181 ---------+---------+---------+---------+---------+---------+   240

      GACTATAACGACAACACGGTAGTTATCGTCCTGAATCTCGGACAGCGGGTGGTCGGCATC
  241 ---------+---------+---------+---------+---------+---------+   300

      GTGGTTGACGGCGTCTCAGACGTGCTTTCATTGACGGCGGAGCAAATTCGTCCGGCACCG
  301 ---------+---------+---------+---------+---------+---------+   360

      GAATTTGCCGTGACGCTTTCAACAGAATATCTCACTGGACTGGGCGCACTGGGCGACCGG
  361 ---------+---------+---------+---------+---------+---------+   420

      ATGTTGATTCTGGTGAACATCGAAAAACTGCTGAACAGCGAAGAGATGGCGCTGTTAGAT
  421 ---------+---------+---------+---------+---------+---------+   480
                                  Xba I Sal I                 Hind III
      AGCGCGGCGTCAGAAGTGGCGTAATTCTCTAGAGTCGACCTGCAGCCCAAGCTTATCGAT
  481 ---------+---------+---------+---------+---------+---------+   540

      GATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGCGGGCTTTTTTTTCAGA
  541 ---------+---------+---------+---------+---------+---------+   600

      TCTGCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATA
  601 ---------+---------+---------+---------+---------+---------+   660

      ATAATGGTTTCTTAGCGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCG
  661 ---------+---------+---------+---------+---------+---------+   720

      CGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCG
  721 ---------+---------+---------+---------+---------+---------+   780

      CTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTC
  781 ---------+---------+---------+---------+---------+---------+   840

      TAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAA
  841 ---------+---------+---------+---------+---------+---------+   900

      AACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
  901 ---------+---------+---------+---------+---------+---------+   960

      CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACAC
  961 ---------+---------+---------+---------+---------+---------+  1020


      TCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
```



```
     CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGA
2161 ---------+---------+---------+---------+---------+---------+  2220

     TCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
2221 ---------+---------+---------+---------+---------+---------+  2280

     CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT
2281 ---------+---------+---------+---------+---------+---------+  2340

     GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGC
2341 ---------+---------+---------+---------+---------+---------+  2400

     TACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCC
2401 ---------+---------+---------+---------+---------+---------+  2460

     TTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACC
2461 ---------+---------+---------+---------+---------+---------+  2520

     TCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCG
2521 ---------+---------+---------+---------+---------+---------+  2580

     GGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTT
2581 ---------+---------+---------+---------+---------+---------+  2640

     CGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
2641 ---------+---------+---------+---------+---------+---------+  2700

     AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG
2701 ---------+---------+---------+---------+---------+---------+  2760

     GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTT
2761 ---------+---------+---------+---------+---------+---------+  2820

     ATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
2821 ---------+---------+---------+---------+---------+---------+  2880

     GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
2881 ---------+---------+---------+---------+---------+---------+  2940

     GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTA
2941 ---------+---------+---------+---------+---------+---------+  3000

     TTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
3001 ---------+---------+---------+---------+---------+---------+  3060

     CAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCG
3061 ---------+---------+---------+---------+---------+---------+  3120

     GTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT
3121 ---------+---------+---------+---------+---------+---------+  3180

     AAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCG
3181 ---------+---------+---------+---------+---------+---------+  3240

     CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAA
3241 ---------+---------+---------+---------+---------+---------+  3300
```


GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGC
3301 ---------+---------+---------+---------+---------+---------+ 3360
GCGAGGCAGAACGCCATCAAAAATAATTCGCGTCTGGCCTTCCTGTAGCCAGCTTTCATC
3361 ---------+---------+---------+---------+---------+---------+ 3420
AACATTAAATGTGAGCGAGTAACAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATT
3421 ---------+---------+---------+---------+---------+---------+ 3480
GACCGTAATGGGATAGGTTACGTTGGTGTAGATGGGCGCATCGTAACCGTGCATCTGCCA
3481 ---------+---------+---------+---------+---------+---------+ 3540
GTTTGAGGGGACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCTTTCCGG
3541 ---------+---------+---------+---------+---------+---------+ 3600
CACCGCTTCTGGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGCTGCGCAACTG
3601 ---------+---------+---------+---------+---------+---------+ 3660
TTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG
3661 ---------+---------+---------+---------+---------+---------+ 3720
TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAC
3721 ---------+---------+---------+---------+---------+---------+ 3780
GACGGCCAGTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
3781 ---------+---------+---------+---------+---------+---------+ 3840
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAAT
3841 ---------+---------+---------+---------+---------+---------+ 3900
GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
3901 ---------+---------+---------+---------+---------+---------+ 3960
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
3961 ---------+---------+---------+---------+---------+---------+ 4020
GGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACC
4021 ---------+---------+---------+---------+---------+---------+ 4080
GCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAA
4081 ---------+---------+---------+---------+---------+---------+ 4140
TCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT
4141 ---------+---------+---------+---------+---------+---------+ 4200
CCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCG
4201 ---------+---------+---------+---------+---------+---------+ 4260
CCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGC
4261 ---------+---------+---------+---------+---------+---------+ 4320
ATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATC
4321 ---------+---------+---------+---------+---------+---------+ 4380
GGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAG
4381 ---------+---------+---------+---------+---------+---------+ 4440

```
     ACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGC
4441 ---------+---------+---------+---------+---------+---------+  4500

     TCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGG
4501 ---------+---------+---------+---------+---------+---------+  4560

     TCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCA
4561 ---------+---------+---------+---------+---------+---------+  4620

     TCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
4621 ---------+---------+---------+---------+---------+---------+  4680

     TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTG
4681 ---------+---------+---------+---------+---------+---------+  4740

     GCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGG
4741 ---------+---------+---------+---------+---------+---------+  4800

     GCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCC
4801 ---------+---------+---------+---------+---------+---------+  4860

     ACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC
4861 ---------+---------+---------+---------+---------+---------+  4920

     GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCA
4921 ---------+---------+---------+---------+---------+---------+  4980

     TACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCT
4981 ---------+---------+---------+---------+---------+---------+  5040

     TCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCTGGCAC
5041 ---------+---------+---------+---------+---------+---------+  5100

     GACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC
5101 ---------+---------+---------+---------+---------+---------+  5160

     ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATT
5161 ---------+---------+---------+---------+---------+---------+  5220

     GTGAGCGGATAACAATTTCACACAGGAAACAGGATCGATCCATCGATGAGCTTACTCCCC
5221 ---------+---------+---------+---------+---------+---------+  5280

     ATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC
5281 ---------+---------+---------+---------+---------+---------+  5340

     AATTTCACACAGGAAACAGGATCAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATC
5341 ---------+---------+---------+---------+---------+---------+  5400

     GGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGGATCCA
5401 ---------+---------+---------+---------+---------+---------+  5460

     TCGATGCTTAGGAGGTCAT
5461 ---------+---------  5479
```

Limited Restriction Site Map of pTacTac

Enzyme	Sequence	#sites	Location of Sites
AccI	GTCGAC, GTATAC	2:	514, 3187
AflIII	ACPuPyGT	2:	2956, 4660
AhaII	GPuCGPyC	6:	169, 310, 486, 1527, 4021, 4704
AhaIII	TTTAAA	3:	1488, 2180, 2199
AlwI	GGATC	8:	1421, 1742, 2206, 2304, 2390, 5252 5359, 5454
AlwNI	CAG---CTG	1:	2542
ApaI	GGGCCC	1:	4453
ApaLI	GTGCAC	4:	1396, 2642, 3142, 4680
AseI	ATTAAT	6:	1892, 3917, 3976, 5143, 5296, 5391
AvaI	CTCGGG	1:	1031
AvaII	GGTCC, GGACC	3:	1704, 1926, 4109
BalI	TGGCCA	1:	29
BamHI	GGATCC	1:	5454
BanI	GGPyPuCC	9:	354, 883, 2115, 3599, 3611, 3891 4021, 4740, 5169
BanII	GpuGCPyC	3:	568, 849, 4453
BbvI	GCTGC, GCAGC	19:	522, 742, 1647, 1824, 2013, 2341, 2547, 2550, 2615, 3034, 3052, 3223 3320, 3649, 3722, 3971, 4100, 4229 4600
BclI	TGATCA	1:	4646
BglI	GCC-----GGC	2:	1944, 3640
BglII	AGATCT	2:	562, 598
BinI	GGATC, GATCC	14:	114, 1421, 1439, 1742, 2206, 2219, 2304, 2316, 2390, 5252, 5257, 5359 5454, 5455
BspHI	TCATGA	3:	653 1228 2236
BspMI	ACCTGC	1:	518
BssHII	GCGCGC	1:	4249
BstEII	GGT-ACC	1:	4478

BstXI	CCA------TGG	3:	4607, 4730, 4859
BstYI	PuGATCPy	10:	113, 562, 598, 1421, 1438, 2206, 2218, 2304, 2315, 5454
Bsu36I	CCT-AGG	1:	3567
ClaI	ATCGAT	4:	535, 5254 5262 5460
DraIII	CAC---GTG	1:	922
EaeI	PyGGCCPu	4:	29, 1675, 3783, 3986
EcorII	CCTGG, CCAGG	14:	48, 228, 2796, 2809, 2930, 3622, 3749, 3886, 4025, 4082, 4622, 4937, 5093, 5175
Eco0109-2	AGG-CCT	1:	617
EcoRV	GATATC	2:	94, 4212
FokI	GGATG, CATCC	11:	234, 419, 1616, 1903, 2084, 3284, 3716, 4619, 4628, 5280, 5375
FspI	TGCGCA	2:	1843, 3651
HaeI	A/T GGCC A/T	7:	29, 45, 2480, 2932, 2943, 3395, 4934
HaeII	PuGCGCPy	9:	468, 765, 773, 2712, 3082, 3630, 4021, 4264, 5045
HgaI	GACGC, GCGTC	14:	170, 311, 372, 487, 675, 1527, 2277, 2855, 3253, 3390, 4424, 4430 4659, 4704
HgiAI	GAGCAC, GTGCAC	5:	1396, 1481, 2642, 3142, 4680
HgiEII	ACC------GGT	3:	2369, 3602, 5061
HincII	GTPyPuAC	7:	205, 304, 514, 1524, 4156, 5289, 5384
HindIII	AAGCTT	1:	529
HpaI	GTTAAC	1:	4156
HphI	GGTGA, TCACC	15:	73, 432, 912, 1349, 1364, 1590, 1986, 2213, 3340, 3349, 4043, 4076 4478, 4943, 5018
MboII	GAAGA, TCTTC	16:	79, 460, 608, 793, 1274, 1383, 1461, 2216, 2307, 3078, 3573, 3686 4183, 4522, 4673, 5038
MluI	ACGCGT	1:	4660

MstII	CCT-AGG	1:	3567
NaeI	GCCGGC	1:	819
NarI	GGCGCC	1:	4021
NciI	CCCGG, CCGGG	7:	1531, 1882, 2578, 3279, 3314, 4233, 5042
Nsp7524I	ACATGT, GCATGT	2:	2956, 3323
OxaNI	CCT-AGG	1:	3567
PflMI	CCA-----TGG	1:	5079
PleI	GAGTC, GACTC	8:	512, 972, 994, 2069, 2586, 3057, 4237, 5033
PstI	CTGCAG	2:	520, 1822
PvuI	CGATCG	2:	1696, 3671
PvuII	CAGCTG	3:	3701, 3969, 4062
SalI	GTCGAC	1:	514
ScaI	AGTACT	1:	1585
SfaNI	GATGC, GCATC	16:	296, 1377, 1607, 1817, 2869, 3089, 3110, 3167, 3283, 3518, 3531, 4289 4305, 4496, 4618, 5463
SnaI	GTATAC	1:	3187
SspI	AATATT	2:	1130, 1261
StyI-3	CCTTGG	1:	69
TthIII1	GAC---GTC	2:	307, 3211
TthIII2-1	CAAACA	2:	551, 2349
TthIII2-3	TGTTTG	3:	2382, 4144, 4837
TthIII2-4	TGCTTG	2:	603, 2343
XbaI	TCTAGA	1:	508
XmnI	GAA----TTC	1:	1464

The following enzymes did not cleave pTacTac:

NAME	SEQUENCE	NAME	SEQUENCE	NAME	SEQUENCE
AatII	GACGTC	AccI-1	GTAGAC	AccI-2	GTCTAC
AflII	CTTAAG	AflIII-2	ACGTGT	AflIII-3	ACACGT
AhaII-3	GACGTC	AsuII	TTCGAA	AvaI-1	CTCGAG
AvaI-2	CCCGAG	AvaI-4	CCCGGG	AvrII	CCTAGG
BanI-1	GGTACC	BanII-1	GAGCTC	BsmI-1	GAATGC
BsmI-2	GCATTC	Bsp1286-3	GTGCCC	Bsp1286-5	GAGCTC
Bsp1286-6	GTGCTC	Bsp1286-7	GGGCAC	BspMI-2	GCAGGT
BspMII	TCCGGA	BstBI	TTCGAA	EaeI-1	TGGCCG
EaeI-3	CGGCCG	EagI	CGGCCG	EcorI	GAATTC
EcoNI	CCT-----AGG	Eco0109-1	GGG-CCT	Eco0109-3	GGG-CCC
Eco0109-4	AGG-CCC	FspII	TTCGAA	HaeI-3	AGGCCT
HaeII-1	AGCGCT	HgiAI-2	GTGCTC	HgiAI-4	GAGCTC
HincII-2	GTCAAC	KpnI	GGTACC	NcoI	CCATGG
NdeI	CATATG	NheI	GCTAGC	NotI	GCGGCCGC
NruI	TCGCGA	NsiI	ATGCAT	NspBII-4	CCGCGG
Nsp7524I-2	ACATGC	Nsp7524I-4	GCATGC	PpuMI-1	GGGACCT
PpuMI-2	AGGACCT	PpuMI-3	GGGACCC	PpuMI-4	AGGACCC
PpuMI-5	GGGTCCT	PpuMI-6	AGGTCCT	PpuMI-7	GGGTCCC
PpuMI-8	AGGTCCC	RsrII-1	CGGACCG	RsrII-2	CGGTCCG
SacI	GAGCTC	SacII	CCGCGG	SfiI	GGCC-----GGCC
SmaI	CCCGGG	SnaBI	TACGTA	SpeI	ACTAGT
SphI	GCATGC	StuI	AGGCCT	StyI-1	CCAAGG
StyI-2	CCATGG	StyI-4	CCTAGG	TthIII2-2	CAAGCA
XhoI	CTCGAG				

Map of hSOD gene in bicistronic construction

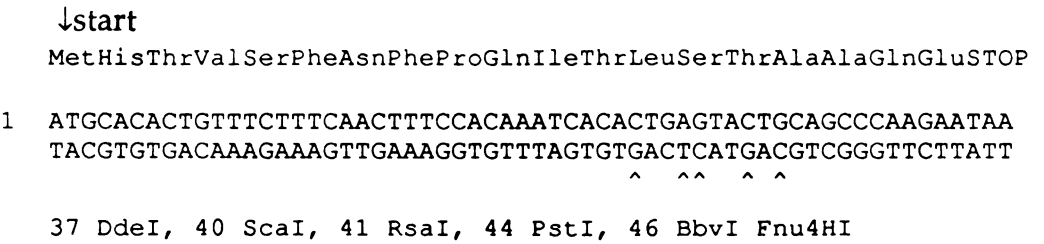
```
     ↓start
     MetAlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsn

   1 ATGGCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAAT
     TACCGCTGCTTCCGGCACACGCACGACTTCCCGCTGCCGGGTCACGTCCCGTAGTAGTTA
               ^                         ^           ^

     12 HaeIII, 37 HaeIII Sau96I, 50 SfaNI


     PheGluGlnLysGluSerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThr

  61 TTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACT
     AAGCTCGTCTTCCTTTCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGA
      ^                    ^ ^                        ^

     62 TaqI, 82 AvaII Sau96I, 84 HgiEII, 107 MseI


     GluGlyLeuHisGlyPheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSer

 121 GAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGT
     CTTCCGGACGTACCTAAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCA
       ^^     ^   ^    ^    ^^                            ^ ^

     123 HaeI StuI, 124 HaeIII, 130 NlaIII, 134 HinfI, 139 NlaIII,
     144 BspHI, 145 NlaIII, 173 RsaI, 175 HgiEII


     AlaGlyProHisPheAsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArg

 181 GCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGG
     CGTCCAGGAGTGAAATTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCC
     ^ ^^  ^      ^    ^                    ^^      ^   ^    ^  ^

     181 BspMI, 183 Eco0109 PpuMI, 184 AvaII Sau96I, 187 MnlI, 194
     MseI, 199 MnlI, 220 Sau96I, 221 HaeIII, 228 FokI, 232 MboI
     237 MnlI, 240 Nsp7524I,


     HisValGlyAspLeuGlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIle

 241 CATGTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATT
     GTACAACCTCTGAACCCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAA
     ^                   ^                      ^^

     241 NlaIII, 262 MaeIII, 284 EaeI, 285 HaeIII,
```

```
     GluAspSerValIleSerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValVal

301  GAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTC
     CTTCTAAGACACTAGAGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAG
     ^  ^       ^        ^              ^      ^^^            ^

     301 MboII, 304 HinfI, 312 MboI, 321 DdeI, 335 SfaNI, 342 Eae I

     343 HaeIII, 344 Fnu4HI, 357 AvaII Sau96I,


     HisGlnLysAlaAspAspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsn

361  CATCAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAAC
     GTAGTTTTTCGTCTACTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTG
                                           ^      ^

     397 MboII, 404 RsaI,


     AlaGlySerArgLeuAlaCysGlyValIleGlyIle

421  GCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATC | start of bicistronic
     CGACCTTCAGCAAACCGAACACCACATTAACCCTAG |      cassette
                                     ^^

     452 AlwI BinI, 453 MboI,
```

Map of CheY gene in bicistronic constructions

```
      ↓start
      MetAlaAspLysGluLeuLysPheLeuValValAspAspPheSerThrMetArgArgIle

  1   ATGGCGGATAAAGAACTTAAATTTTTGGTTGTGGATGACTTTTCCACCATGCGACGCATA
      TACCGCCTATTTCTTGAATTTAAAAACCAACACCTACTGAAAAGGTGGTACGCTGCGTAT
                      ^               ^               ^    ^

      17 MseI, 33 FokI, 48 NlaIII, 53 HgaI


      ValArgAsnLeuLeuLysGluLeuGlyPheAsnAsnValGluGluAlaGluAspGlyVal

 61   GTGCGTAACCTGCTGAAAGAGCTGGGATTCAATAATGTTGAGGAAGCGGAAGATGGCGTC
      CACGCATTGGACGACTTTCTCGACCCTAAGTTATTACAACTCCTTCGCCTTCTACCGCAG
          ^  ^           ^     ^              ^        ^     ^^ ^^

      65 MaeIII, 68 BspMI, 80 AluI, 86 HinfI, 100 MnlI, 109 MboII,
      115 AhaII, 116 HgaI, 118 AccI HincII SalI, 119 TaqI,


      AspAlaLeuAsnLysLeuGlnAlaGlyGlyTyrGlyPheValIleSerAspTrpAsnMet

121   GACGCTCTCAATAAGTTGCAGGCAGGCGGTTATGGATTTGTTATCTCCGACTGGAACATG
      CTGCGAGAGTTATTCAACGTCCGTCCGCCAATACCTAAACAATAGAGGCTGACCTTGTAC
      ^                                                       ^^

      121 HgaI, 176 Nsp7524I, 177 NlaIII,


      ProAsnMetAspGlyLeuGluLeuLeuLysThrIleArgAlaAspGlyAlaMetSerAla

181   CCCAATATGGATGGCCTGGAATTGCTGAAAACAATTCGTGCGGATGGCGCGATGTCGGCA
      GGGTTATACCTACCGGACCTTAACGACTTTTGTTAAGCACGCCTACCGCGCTACAGCCGT
            ^  ^^ ^                              ^    ^^

      189 FokI, 192 HaeI, 193 HaeIII, 195 EcorII ScrFI, 222 FokI,
      227 HhaI, 228 BstUI FnuDII ThaI,


      LeuProValLeuMetValThrAlaAlaGlnSTOP

241   TTGCCAGTGTTAATGGTGACTGCAGCCCAATAA
      AACGGTCACAATTACCACTGACGTCGGGTTATT
              ^    ^^   ^ ^

      250 MseI, 255 HphI, 256 MaeIII, 260 PstI, 262 BbvI Fnu4HI
```

Map of synthetic operon in bicistronic constructions



Map of HIV-1 protease gene in bicistronic constructions

```
     ProGlnIleThrLeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLys

   1 CCACAAATCACCTTGTGGCAAAGACCACTAGTTACCATCAGAATCGGTGGACAATTGAAA
     GGTGTTTAGTGGAACACCGTTTCTGGTGATCAATGGTAGTCTTAGCCACCTGTTAACTTT
            ^^                  ^^  ^         ^

     8 HphI, 9 DraIII, 27 SpeI, 28 MaeI, 31 MaeIII, 41 HinfI,


     GluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGly

  61 GAAGCTTTGTTGGACACCGGTGCTGACGACACCGTTTTGGAAGAAATGAACTTGCCAGGA
     CTTCGAAACAACCTGTGGCCACGACTGCTGTGGCAAAACCTTCTTTACTTGAACGGTCCT
      ^^             ^                    ^              ^

     62 HindIII, 63 AluI, 77 HpaII, 100 MboII, 115 BstXI EcorII ScrFI


     LysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAsp

 121 AAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGAT
     TTTACCTTTGGTTTTTACTATCCCCCTTAACCTCCAAAATAGTTTCATTCTGTCATGCTA
                                  ^                    ^   ^

     152 MnlI, 174 RsaI, 178 MboI,


     GlnIleProValGluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThr

 181 CAGATACCTGTAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACA
     GTCTATGGACATCTTTAGACACCTGTATTTCGATATCCATGTCATAATCATCCTGGATGT
                               ^       ^           ^^

     210 AluI, 218 RsaI, 231 EcoO109 PpuMI, 232 AvaII Sau96I,


     ProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheSTOP

 241 CCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCTAG
     GGACAGTTGTATTAACCTTCTTTAGACAACTGAGTCTAACCAACATGAAATTTAAAGATC
       ^           ^         ^  ^ ^          ^   ^^       ^  ^

     244 HincII, 257 MboII, 267 HincII, 270 HinfI PleI, 272 DdeI,
     284 RsaI, 288 AhaIII, 289 MseI, 297 MaeI, 300 AccI Hinc II SalI
```

Map of HIV-2 gene in bicistronic constructions

```
     ProGlnPheSerLeuTrpLysArgProValValThrAlaTyrIleGluGlyGlnProVal

   1 CCTCAATTCTCTCTTTGGAAAAGACCAGTGGTCACCGCATACATTGAGGGTCAGCCAGTA
     GGAGTTAAGAGAGAAACCTTTTCTGGTCACCAGTGGCGTATGTAACTCCCAGTCGGTCAT
     ^                               ^^^             ^

     1 MnlI, 30 BstEII, 31 MaeIII, 32 HphI, 46 MnlI


     GluValLeuLeuAspThrGlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsn

  61 GAAGTCTTGTTAGACACAGGGGCTGACGACAGCATAGTAGCAGGAATAGAGTTAGGGAAC
     CTTCAGAACAATCTGTGTCCCCGACTGCTGTCGTATCATCGTCCTTATCTCAATCCCTTG
                    ^

     77 AlwNI


     AsnTyrSerProLysIleValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLys

 121 AATTATAGCCCAAAAATAGTAGGGGGAATAGGGGGATTCATAAATACCAAGGAATATAAA
     TTAATATCGGGTTTTTATCATCCCCCTTATCCCCCTAAGTATTTATGGTTCCTTATATTT
                                         ^           ^

     155 HinfI, 167 StyI


     AsnValGluIleGluValLeuAsnLysLysValArgAlaThrIleMetThrGlyAspThr

 181 AATGTAGAAATAGAAGTTCTAAATAAAAAGGTCCGGGCCACCATAATGACAGGCGACACC
     TTACATCTTTATCTTCAAGATTTATTTTTCCAGGCCCGGTGGTATTACTGTCCGCTGTGG
                                   ^  ^ ^^

     210 AvaII Sau96I, 213 HpaII NciI ScrFI, 215 Sau96I, 216 HaeIII


     ProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGlyMetSerLeuAsnLeuSTOP

 241 CCAATCAATATTTTTGGCAGAAATATTCTGACAGCCTTAGGCATGTCATTAAATCTATAG
     GGTTAGTTATAAAAACCGTCTTTATAAGACTGTCGGAATCCGTACAGTAATTTAGATATC
           ^               ^             ^^    ^^      ^

     247 SspI, 262 SspI, 275 Bsu36I MstII OxaNI, 276 DdeI, 281
     Nsp7524I, 282 NlaIII, 289 MseI
```

Map of SIVmac$_{239}$ gene in bicistronic constructions

```
    ProGlnPheSerLeuTrpArgArgProValValThrAlaHisIleGluGlyGlnProVal

  1 CCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTA
    GGAGTTAAGAGAGAAACCTCCTCTGGTCATCAGTGACGAGTATAACTTCCTGTCGGACAT
    ^               ^             ^

    1 MnlI, 18 MnlI, 31 MaeIII


    GluValLeuLeuAspThrGlyAlaAspAspSerIleValThrGlyIleGluLeuGlyPro

 61 GAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCA
    CTTCATAATGACCTATGTCCCCGACTACTAAGATAACATTGTCCTTATCTCAATCCAGGT
                    ^           ^         ^                   ^

    77 AlwNI, 88 HinfI, 97 MaeIII, 115 AvaII Sau96I


    HisTyrThrProLysIleValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLys

121 CATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAA
    GTAATATGGGGTTTTTATCATCCTCCTTATCCTCCAAAATAATTATGATTTCTTATGTTT
                            ^         ^        ^^

    143 MnlI, 152 MnlI, 160 AseI, 161 MseI


    AsnValGluIleGluValLeuGlyLysArgIleLysGlyThrIleMetThrGlyAspThr

181 AATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACC
    TTACATCTTTATCTTCAAAATCCGTTTTCCTAATTTCCCTGTTAGTACTGTCCCCTGTGG
                                      ^            ^^

    212 MseI, 224 BspHI, 225 NlaIII


    ProIleAsnIlePheGlyArgAsnLeuLeuThrAlaLeuGlyMetSerLeuAsnPheSTOP

241 CCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATTTTTAG
    GGCTAATTGTAAAAACCATCTTTAAACGATTGTCGAGACCCCTACAGAGATTTAAAAATC
      ^                                 ^        ^

    245 MseI, 273 AluI, 281 FokI
```

Appendix B. Amino Acid Probability Tables for HIV and SIV Substrates

Table B-1. Statistical analysis of residue preferences in HIV-1 substrates.

Numbers represent the ratio of residue abundance at a given site of a substrate to abundance in globular proteins (Poorman, et al., 1991).

residue	P4	P3	P2	P1	P1'	P2'	P3'	P4'
Ala	2.01	0.57	1.15	0.57	1.72	0.57	1.15	0.57
Cys	0.50	0.50	1.56	0.50	0.50	0.50	1.56	1.56
Asp	1.75	0.88	0.11	0.11	0.11	0.11	0.88	0.75
Glu	0.78	3.12	1.17	0.39	1.17	7.81	1.95	0.39
Phe	0.64	1.92	0.16	7.69	3.21	0.16	2.56	2.56
Gly	1.28	0.64	0.08	0.96	0.32	0.08	0.64	0.96
His	1.14	0.50	1.14	0.50	0.50	0.50	0.50	0.50
Ile	0.12	0.96	3.37	0.12	0.48	1.44	1.44	0.12
Lys	0.74	0.37	0.09	0.09	0.09	0.09	0.09	1.10
Leu	0.3	1.52	0.91	3.66	0.91	1.22	0.08	0.91
Met	0.50	0.50	0.50	3.57	2.38	0.50	2.38	2.38
Asn	0.14	0.14	3.41	2.27	0.14	0.14	0.57	1.14
Pro	2.22	0.56	0.14	0.14	2.78	0.14	0.14	2.22
Gln	0.64	3.85	0.16	0.16	0.16	3.85	0.16	0.64
Arg	1.56	1.56	0.13	0.13	1.56	0.13	2.08	0.52
Ser	1.89	0.38	0.76	0.09	0.38	0.09	1.52	1.89
Thr	0.86	1.72	1.72	0.11	0.86	0.43	2.16	1.29
Val	0.36	0.09	3.21	0.09	1.07	1.43	0.71	0.09
Trp	0.50	0.50	0.50	0.50	2.08	0.50	0.50	0.50
Tyr	0.76	0.19	0.19	2.27	3.03	0.19	0.76	0.19

Table B-2. Statistical analysis of residue preferences in HIV-2 substrates.

Numbers represent the ratio of residue abundance at a given site of a substrate to abundance in globular proteins (Poorman, et al., 1991).

residue	P4	P3	P2	P1	P1'	P2'	P3'	P4'
Ala	1.04	0.52	2.09	1.57	3.13	0.13	1.57	1.04
Cys	0.50	0.50	0.50	0.50	0.50	0.50	2.84	0.50
Asp	0.80	0.80	0.20	0.20	0.20	0.20	0.80	0.20
Glu	0.71	2.13	2.13	0.71	0.18	7.10	2.13	0.18
Phe	0.50	2.33	0.50	2.33	3.50	0.50	1.17	4.66
Gly	2.33	1.17	0.15	1.75	0.15	0.15	0.58	1.17
His	0.50	2.07	0.50	0.50	0.50	0.50	0.50	2.07
Ile	0.22	0.87	2.62	0.22	0.22	1.75	0.87	0.22
Lys	0.17	0.17	0.17	0.17	0.17	0.17	0.17	1.34
Leu	0.14	1.11	0.55	3.88	0.55	1.11	1.66	0.55
Met	0.50	0.50	0.50	6.49	4.33	0.50	2.16	2.16
Asn	1.03	0.50	6.20	1.03	0.50	1.03	0.50	1.03
Pro	3.03	1.01	0.50	0.50	5.05	0.50	0.50	0.50
Gln	0.50	5.83	0.50	0.50	0.50	4.66	2.33	0.50
Arg	2.84	0.95	0.24	0.24	0.24	0.24	1.89	1.89
Ser	2.07	0.17	0.69	0.17	0.17	0.17	0.69	0.69
Thr	0.78	1.57	0.78	0.20	2.35	0.20	0.78	0.20
Val	0.16	0.16	1.95	0.16	0.65	1.95	0.65	0.16
Trp	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Tyr	2.75	0.50	0.50	2.75	1.38	0.50	0.50	0.50


UC
San Francisco
LIBRARY