

Kinetic Analysis of Mutants of HTLV-I Protease

A Dissertation
Presented to
The Academic Faculty

By

Bryan Edward Herger

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Chemistry

Georgia Institute of Technology

June, 2004

Kinetic Analysis of Mutants of HTLV-I Protease

Approved by:

Suzanne B. Shuker
Rigoberto Hernandez
Nicholas Hud
Debra Mohler
Sherry Michele Owen
Loren Williams

18 June 2004

ACKNOWLEDGEMENTS

Many thanks to everyone who has seen me through this endeavor: to my mother and father, for moral and financial support, and the rest of my family, including my brother Jim, sister Jonilee, and stepfather Tom Blinn; to Dr. Suzanne Shuker, for putting up with me for nearly six years, as well as the rest of Shuker group, including Dr. Vicki Mariani, Dr. Julie Ha, Dr. David Gaul, Dr. Kevin Caran, Dr. Beth Brewster, David Kim, Byung-hun Lee, Micah Gliedt, Josh Sasine, Suazette Reid, Bob Chen, Ta-Tanisha Favor, Mike Kulis, Kelly Joy Dennison, and the many undergraduate students who have worked in our lab; to my thesis committee, thanks for finally letting me out; to my former roommates, Dr. Cliff Lipscomb and Russell Dondero, and current roommate, Marinda Rule; to current and former fellow graduate students Ryan Price, Melissa Arredondo, Ben Ayres, Ericka Moore, Mutasem Sinnokrot, Chuck Gumienny, all the residents of IBB Wing 3A, and others I have forgotten to list here; and finally to Dr. Block and Dr. Braga, for six years of TA assignments, right up to the very end of Summer 2004! What an adventure it was! So long, and thanks for all the fish (tacos).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS AND SYMBOLS	ix
SUMMARY	xii
Chapter 1 HTLV-I: An Oncogenic Retrovirus	1
1.1 On Viruses	1
1.2 Retroviruses	2
1.3 Classification of Retroviruses	4
1.4 A Day in the Life of a Retrovirus: Structure, Function, and Replication	5
1.5 HTLV-I Specific Structure and Function	6
1.6 Pathology and Treatment	13
1.7 References	15
Chapter 2 HTLV-I Protease: A Retroviral Protease	17
2.1 Proteases	17
2.2 Classification of Proteases	17
2.3 Aspartic Acid Proteases: Structure and Mechanism	21
2.4 HTLV-I Protease: Background	24
2.5 Substrates and Inhibitors of HTLV-I Protease	27
2.6 HTLV-I Protease Expression, Purification and Assay	30
2.7 Significance of Present Work	33
2.8 References	34
Chapter 3 Development of a Theoretical Model and Function of the Ten C-terminal Residues	36
3.1 Similarity of HTLV-I Protease to Other Retroviral Proteases	36
3.2 Sequence Alignment and Generation of Theoretical Models	38
3.3 Docking a Native Substrate into the Active Site	44
3.4 The Ten C-terminal Residues	46
3.5 Activity of Full-length and Truncated HTLV-I Protease	46
3.6 Discussion	50
3.7 References	51
Chapter 4 Modifying the Substrate Specificity of HTLV-I Protease	52
4.1 Identifying the Substrate Binding Sites of HTLV-I Protease	52
4.2 Determining Residues Involved in Binding by Sequence Alignment	52
4.3 Preparation of Mutants	55

4.4 Activity of HIV-1 Substrate-specific Mutant based on Tözsér alignment	58
4.5 Activity of HIV-1 Substrate-specific Mutant based on Shuker alignment	58
4.6 Activity of HIV-1 Substrate-specific Mutant based on alternate Shuker alignment	59
4.7 Summary and Discussion	61
4.8 References	63
Chapter 5 Alanine Scan of HTLV-I Protease	64
5.1 Purpose of Alanine Scan	64
5.2 Preparation of Alanine Mutants	66
5.3 Results and Discussion	68
5.4 References	72
Chapter 6 Conclusions and Future Work	73
Chapter 7 Materials and Methods	76

LIST OF TABLES

Table 2.1	Comparison of the substrate preferences of HTLV-I and HIV-1 proteases.	26
Table 2.2	Native cleavage sites of HTLV-I protease.	28
Table 2.3	Effect of inhibitors on three aspartic acid proteases.	30
Table 3.1	Homology, or percent identical sequence, of various retroviral proteases.	37
Table 3.2	Summary of kinetic data for L40I and L40I-wo10 protease.	50
Table 4.1	Preliminary kinetic data for specificity mutants.	58
Table 4.2	Summary of kinetic data from various mutants of HTLV-I protease and HIV-1 protease on HIV-1 protease substrate.	61
Table 5.1	Summary of kinetic data of alanine mutants.	68

LIST OF FIGURES

Figure 1.1	Organization of HTLV-I virion.	7
Figure 1.2	HTLV-I genome, showing genes and open reading frame.	8
Figure 1.3	Complete translation of HTLV-I gag-pro-pol ployprotein.	9
Figure 1.4	Action of HTLV-I protease in viral maturation.	11
Figure 1.5	HTLV-I replication cycle.	12
Figure 2.1	Structures of retroviral proteases.	20
Figure 2.2	Schematic of secondary structural elements in retroviral proteases.	21
Figure 2.3	Schematic diagram of substrate positions and binding pockets.	22
Figure 2.4	Mechanism of aspartic acid proteases.	23
Figure 2.5	HTLV-I protease DNA and amino acid sequence.	25
Figure 2.6	HIV-1 protease inhibitors.	29
Figure 2.7	Fluorogenic substrate for HTLV-I protease.	32
Figure 3.1	Sequence alignment of several retroviral proteases.	39
Figure 3.2	Scripts used by MODELLER to generate the theoretical model.	40
Figure 3.3	Theoretical model of full-length HTLV-I protease.	42
Figure 3.4	Theoretical model of C-terminal truncated HTLV-I protease.	43
Figure 3.5	HTLV-I protease with docked CA/NC substrate.	45
Figure 3.6	Eadie-Hofstee plot of L40I kinetic data.	48
Figure 3.7	Eadie-Hofstee plot of L40I-wo10 kinetic data.	49
Figure 4.1	Multiple alignment of RSV protease, HTLV-I protease, and HIV-1 protease.	54
Figure 4.2	Mutation by overlap extension.	57
Figure 4.3	Mutations made based on different alignments.	59

Figure 4.4	Revised alignment of HTLV-I protease and HIV-1 protease.	60
Figure 5.1	Positions of residues selected for alanine scan.	65
Figure 5.2	Substrate used in alanine scan kinetic study.	67
Figure 5.3	Stabilization of the backbone carbonyl oxygen by Thr94 prior to cleavage between Leu and Val.	70

LIST OF SYMBOLS AND ABBREVIATIONS

A, Ala	Alanine
ATL	Adult T-Cell Leukemia/lymphoma
C, Cys	Cysteine
CA	Capsid
CDC	Centers for Disease Control
D, Asp	Aspartic acid
DABCYL	4-([4-(dimethylamino)phenyl]azo)benzoic acid
E, Glu	Glutamic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDANS	5-(2-Aminoethylamino)-1-naphthalenesulfonic acid
EIAV	Equine Infectious Anemia Virus
F, Phe	Phenylalanine
FIV	Feline Immunodeficiency Virus
FRET	Fluorescence Resonance Energy Transfer
G, Gly	Glycine
H, His	Histidine
HAM	HTLV-I Associated Myelopathy
HIV	Human Immunodeficiency Virus
HTLV-I	Human T-cell Lymphotropic Virus, type I
I, Ile	Isoleucine
IN	Integrase
K, Lys	Lysine

k_{cat}	turnover number
K_M	Michaelis constant
L, Leu	Leucine
LB	Luria-Bertiani broth, Miller
LTR	Long Terminal Repeat
M	Molar concentration
M, Met	Methionine
MA	Matrix
mg	milligram
microM, μM	micromolar concentration
mL	milliliter
mM	millimolar concentration
N, Asn	asparagine
NC	Nucleocapsid
NEB	New England Biolabs
nm	nanometer
P, Pro	Proline
PR	protease
Q, Gln	Glutamine
R, Arg	Arginine
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
S, Ser	Serine

SIV	Simian Immunodeficiency Virus
T, Thr	Threonine
TSP	Tropical Spastic Paraparesis
uL, μ L	Microliter
V, Val	Valine
W, Trp	Tryptophan

SUMMARY

Human T-cell lymphotropic virus type I (HTLV-I) is a retrovirus that is the causative agent of the fatal disease adult T-cell leukemia (ATL). HTLV-I silently infects over twenty million people worldwide; up to ten percent of these will develop ATL in their lifetime. There are currently no effective treatments for this disease.

HTLV-I expresses its genome as polypeptides that must be processed in order to produce infectious virions. Like other retroviruses, HTLV-I encodes an aspartic acid protease to process these polypeptides into mature form. Because the protease is essential in the virus life cycle, it is an attractive target for the treatment of HTLV-I-induced ATL.

The present work examines the structure and function of HTLV-I protease. A theoretical structure of the protease is presented, and the function of the C-terminal extension is considered. In order to determine which residues are involved in binding substrate, two experiments were performed: first, several residues were mutated to the corresponding residues in HIV-1 protease to determine whether HTLV-I protease can be made to process an HIV-1 protease substrate; second, an alanine scan was performed to knock out individual residues to assess their importance in binding substrate. This work builds knowledge of the structure and function of HTLV-I protease. By understanding which residues play a role in binding substrate and by developing a clearer picture of the structure of the protease, it will be possible to develop specific inhibitors for HTLV-I protease.

CHAPTER 1

HTLV-I: AN ONCOGENIC RETROVIRUS

1.1 On viruses

Pathogenic diseases have plagued humankind since the beginning of recorded history. Although their cause was unknown until recent times, viral diseases have played a significant role in historical events. For example, colonization of the Americas was affected by the spread of smallpox among native populations. Also, influenza epidemics have caused countless millions of deaths. Recent epidemic scares, in fact and fiction, center around newly emerging viruses such as the coronavirus SARS, West Nile Virus, and the Ebola virus.

The cause of illness was unknown until relatively modern times. Until about 300 years ago in the Western world, diseases were considered acts of God, and virulence or contagiousness were not widely understood or accepted. Bacteria were first observed by Anton van Leeuwenhoek in 1677. The implication of bacteria in illness was fully realized in the 19th century, and in 1876, studies of anthrax, caused by *Bacillus anthracis*, led Robert Koch to develop postulates describing pathogenic microorganisms (Baron, 1996).

A number of plants are affected by a disease that causes mosaic-like patterns of normal color, light green and yellow on the leaves. This disease stunts growth and damages leaves, flowers and fruit of affected plants. Studies by biologists

revealed that certain cell-free extracts of plants with this disease could infect other plants. This work led to the discovery of the tobacco mosaic virus. Later, many other viruses were discovered and implicated in a number of plant, animal, and human diseases (Fraenkel-Conrat, 1988).

Virus research is a leading field of medicinal chemistry. While many compounds have been discovered or prepared to combat bacterial and fungal disease, few treatments have been developed for viral diseases aside from HIV, and many of these treatments are intended to alleviate symptoms rather than interrupt the virus life cycle. The principal means of fighting viral infection is preventive treatment with vaccines.

There has been debate whether viruses are either living or nonliving. In a sense, viruses are parasites; like mosquitoes, they rely on the resources of a host in order to reproduce. However, isolated virus particles are typically inert and are not living in the sense that the mosquito carries on respiration and other life processes independent of a host. However, viruses have a self-replicating genome and often produce enzymes in the host to assemble a mature virus, blurring the line between whether they are living or nonliving (Voet, 1995).

1.2 Retroviruses

Retroviruses are virus particles whose genome is RNA-based and that contain reverse transcriptase to transcribe the RNA genome into DNA. Most organisms carry a linear or circular double-stranded DNA genome, and a number of viruses also carry a single- or double-stranded DNA genome. DNA serves as

the template for cellular function through the classical central dogma of molecular biology: DNA is transcribed to RNA, which is translated to proteins that carry out cellular functions, including copying DNA and RNA. Retroviruses break this paradigm: RNA is reverse transcribed into DNA, which is inserted into the host genome. The viral genome then functions according to central dogma, but also produces RNA copies of the genome that are used to assemble infectious particles. Further discoveries, such as catalytic RNA or “ribozymes” that imitate enzyme or protein function, have also challenged the central dogma; however, the retrovirus life cycle represents the first complete system that demonstrates reversibility of the pathway (Voet, 1995).

In 1911, Peyton Rous demonstrated the existence of a pathogen present in cell-free media that could induce tumors in chickens (Rous, 1911). This pathogen is now called the Rous sarcoma virus (RSV), and Rous received the 1966 Nobel Prize in Physiology or Medicine “for his discovery of tumour-inducing viruses”. RSV is a retrovirus, but Rous did not discover this. The first retrovirus to be identified was the feline leukemia virus, found in 1973 by W.D. Hardy et al. (Hardy et al., 1973) Other retroviral oncoviruses have also been found, including the previously noted Rous sarcoma virus, equine infectious anemia virus (EIAV), and bovine leukemia virus (BLV), which causes leukemia in cattle.

The first retrovirus found in humans was HTLV-I, isolated in 1980 by Poiesz et al. (Poiesz et al., 1980) from patients with adult T-cell leukemia (ATL). This discovery was eclipsed by Gallo and others, who discovered the retrovirus HTLV-III (later called HIV) in 1983 (Gallo, 1996). HTLV-I was later found in

patients with HTLV-I associated myelopathy (HAM), tropical spastic paraparesis (TSP) (Gessain et al., 1985), and other chronic diseases including arthritis, infective dermatitis, and uveitis (Gessain, 1996). Studies have established that HTLV-I can travel horizontally through a population by sexual contact or intravenous drug use and vertically from mother to infant (Gallo, 1996). At least 20 million people are infected with HTLV-I, primarily in Japan, the Caribbean, sub-Saharan Africa, and the United States. The Centers for Disease Control have classified HTLV-I as a dangerous emerging pathogen.

1.3 Classification of retroviruses

Retroviruses are classified in the family *Retroviridae*. All retroviruses have an RNA genome, use reverse transcriptase to copy the genome, and are membrane-bound. Viral infection by certain members of this family can cause malignant tumors and leukemia in animals and humans. Other members of the retrovirus family are immunodeficiency viruses, notably HIV, that attack cells of the host immune system. The International Committee on Taxonomy of Viruses (ICTV) is responsible for standardized nomenclature and taxonomy of viruses (Murphy, 1995). The ICTV taxonomy charts are accessible online (ICTVdb, <http://www.ncbi.nlm.nih.gov/ICTVdb/>).

HTLV-I is a member of family *Retroviridae*, which includes all viruses that carry an RNA genome and utilize reverse transcriptase and integrase to insert the genome into the host. Further classification is based on several criteria, including the pathology of the virus. HTLV-I is classified in genus *Deltaretrovirus*,

the members of which are tumorigenic. Other members of genus Deltaretrovirus include bovine leukemia virus (BLV), HTLV-II, feline leukemia virus (FLV), and simian T-lymphotropic virus 1. The ICTV species name for HTLV-I is human T-lymphotropic virus type 1. HIV-1, in contrast, belongs to genus Lentivirus, which is comprised of non-tumorigenic retroviruses; this is why it is no longer named its initial classification, HTLV-III. The tumorigenic and non-tumorigenic viruses also differ structurally; tumorigenic viruses have a spherical capsid, while non-tumorigenic viruses have a conical or cylindrical capsid.

1.4A day in the life of a retrovirus: structure, function, and replication

All retroviruses have a similar genomic structure. Retroviruses express three principal polypeptides: one encoding structural proteins, one encoding DNA- and RNA-manipulating enzymes, and one encoding envelope proteins. They also encode regulatory proteins that control viral replication and define viral pathology. All retroviruses follow a similar life cycle in terms of infection and replication.

The retroviral genome expresses three principal polypeptides. The first, called group antigen or Gag, contains proteins that make up the internal structure of the mature virion: the matrix, capsid, and nucleocapsid. An aspartic acid protease is encoded at the C-terminal end of Gag. The second polypeptide, polymerase or Pol, contains reverse transcriptase, the enzyme needed to transcribe the viral RNA genome, and integrase, which splices the viral DNA into the host genome. The third polypeptide, envelope or Env, contains two envelope proteins. The envelope proteins are on the exterior of the mature virus and are

involved in recognizing receptors on the surface of target cells and beginning the infection process.

Also encoded among these viral polypeptides, usually as part of the structural proteins, is a protease, an enzyme that processes viral polypeptides into mature, active form. The envelope polyprotein is processed by a pathway that is not fully understood. It is known that Env is processed by an endopeptidase and glycosylated by a native cellular pathway (Coffin, Hughes, and Varmus, 1997).

1.5 HTLV-I specific structure and function

HTLV-I virus particles are spherical and membrane-enclosed and range from 80-100 nm in diameter. The structure of a typical HTLV-I particle is depicted in Figure 1.1. The membrane is derived from host cells, but it also contains virally encoded transmembrane protein (TM) and surface unit glycoprotein (SU). The core of the viral particle is composed of matrix, capsid, and nucleocapsid protein and contains two copies of the viral RNA genome and one copy each of reverse transcriptase and integrase. A copy of tRNA-Pro, also present in the capsid, initiates reverse transcription in the host (Murphy, 1995).



Figure 1.1. Organization of HTLV-I virion. Virus membrane (yellow) contains glycoproteins TM (transmembrane protein, red) and SU (surface unit, blue). Virus particle is made of matrix (MA, orange) surrounding cone- or cylinder-shaped capsid (CA, red cylinder). Capsid contains two copies of genomic RNA (silver helices) associated with nucleocapsid (NC, black spheres). Capsid also contains reverse transcriptase (RT, green cube) and integrase (IN, red diamond). A copy of tRNA-Pro (stack of blue spheres) is present and initiates reverse transcription of the viral genome.

The HTLV-I viral genome is 9032 base pairs long and encodes three genes in three open reading frames: *gag*, *pol*, and *env*. These genes are contained between long terminal repeats (LTR's). HTLV-I's genome organization is depicted in Figure 1.2. The first set of genes includes *gag*, *pro*, and *pol*. The genes *pro* and *pol* are frameshifted -1 , or one base in the 5' direction, relative to the preceding gene. The genes *env*, *rex*, and *tax* are separate from the *gag-pro-pol* complex in the 3' direction.

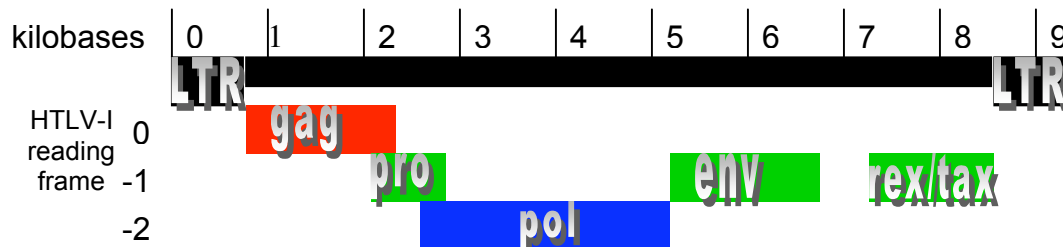


Figure 1.2. HTLV-I genome, showing genes and open reading frame.

HTLV-I expresses structural and replicative components as three polyproteins: Gag, Gag-Pro, and Gag-Pro-Pol. Gag-Pro and Gag-Pro-Pol are produced by ribosomal slippage (Nam et al., 1993; Nam et al., 1988), allowing the production of what are effectively truncated versions of the full-length polyprotein. The expressed products are cleaved by an aspartic acid protease also encoded in the viral genome. Gag (48 kDa) is processed into matrix (MA), capsid (CA), and nucleocapsid (NC) peptides. Cleavage of Pol (99 kDa) yields reverse transcriptase (RT), which transcribes viral RNA to DNA, and integrase (IN), which integrates viral RNA into the host. The protease (PR) is encoded between gag and pol in a different reading frame. The amino acid sequence of Gag-Pro-Pol is shown in Figure 1.3, colored to indicate the different protein components.

MGQIFSRASAPIPRPPRGLAAHHLNFLQAAYRLEPGPSSYDFHQKKFLK
 IALETVPWICPINYSLLASLLPKGYPGRVNEILHILIQTQAQIPSRPAPP
 PSSPTHDPDSDPQIPLPPPYVEPTAPQVLPVMHPHGAPPNHRPWQMKDLQ
 AIKQEVSAARGSPQFMQTI RLAVQQFDPTAKDLQDLLQYLCSSLVASLHH
 QQDLSLISEAETRGITGYNPLAGPLRVQANNPQQQGLRREYQQLWLAFAA
 LPGSAKDPSWASILQGLEEYPYHAFVERLNIALDNGLPEGTPKDPILRS
 LAYSNANKECQKLLQARGHTNSPLGDMLRACQWTWPKDKTKVLVVQPKKPPPNQ
 PCFRCGKAGHWSRDCTQPRPPPGPCPLCQDPTHWKRDCPRLKPTIPEPEPE
 EDALLLDLPADIPHPKNSIGGEVHSTPKKTLHRGGGLTSPPTLQQVFLNQD
 PASILPVIPLDPARRPVIKAQVDTQTSHPKTIEALLDTGADMTVLP
 IALFSNTPLKNTSVLGAGGQTQDHFKLTSPLVILIRLPFRTTPIVLT
 SCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVILPIQAPAVL
 GLEHLPRPPEISQFPLNGKRAACNLANTGASRPWTRTPPKAPRNQPV
 PFKPERLQALQHLVRKALEAGHIESYTGPGNNPVFPVKKANGTWRFI
 HDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTI DLKDAFFQI
 PLPKQFQPYFAFTVPQQCNYPGTRYAWKVL PQGFKNSPTLFEMQ
 LAHILQPIRQAFPOCTILQYMD DILLASPSHEDLLSEATMASLISH
 GLPVSENKTQQTPGTIKFLGQIISP NHLTYDAVPTVPIRSRWALPEL
 QALLGEIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQ
 SLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPL
 VWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIH
 HNISTQTENQFIQTS DHPSVPILLHSHR FKNLGAQTGELWNTFLK
 TAAPLAPVKALMPVFTLS PVIINTAPCLFSDGSTSRAAYILWDK
 HILSQRSFPLPPPHKSAQRAELGLLHGLSSARSWRCLNIFLDSKY
 LYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTN
 LPDPI SRLNALTDALLITPVLQLSPAELHSFT HCGQTALTQGAT
 TTEASNILRSCHACRKNNPQHQM PRGHIRRGLLPNHIWQGDITHF
 KYKNTLYRLHVWVDTFSGAISATQKRKETSSEAISSLLQAIAYL
 GKPSYINTDNGPAYISQDFLNMCTSLAIRHTTHVPYNPTSSGLVER
 SNGILKTLLYKYFTDKPDLPM DNALSIALWTINHLNVLTNCHKTRW
 QLHHSPRLQPI PETHSLSNKQTHWYYFKLPGLNSRQWKGPQEALQEA
 AGAALIPVSASSAQWIPWRLKRAACPRPVGGPADPKEKDHQHHG

Figure 1.3. Complete translation of HTLV-I gag-pro-pol polyprotein: **matrix**, **capsid**, **nucleocapsid**, **protease**, **reverse transcriptase**, **integrase**. Uncolored sequences are linker regions with no known function.

Env is expressed independently from the Gag-Pro-Pol complex. Cleavage of Env yields two membrane (**en**velope) proteins, the transmembrane protein (TM) and the surface glycoprotein (SU). The regulatory proteins rex and tax are encoded in the same reading frame as Env.

One mode of replication of HTLV-I is similar to other retroviruses, including HIV (Morrow, Park, and Wakefield, 1994). Virus infection occurs through binding of a virus surface protein to a T-cell receptor. The virus particle enters the cell and the RNA genome is copied to double-stranded DNA by reverse transcriptase. The viral DNA is integrated into the host DNA. The 5' LTR, upon activation by rex and tax regulatory proteins, promotes transcription of the viral genome. Translated polyproteins are processed into mature form by the viral protease. The function of the protease is depicted in Figure 1.4.

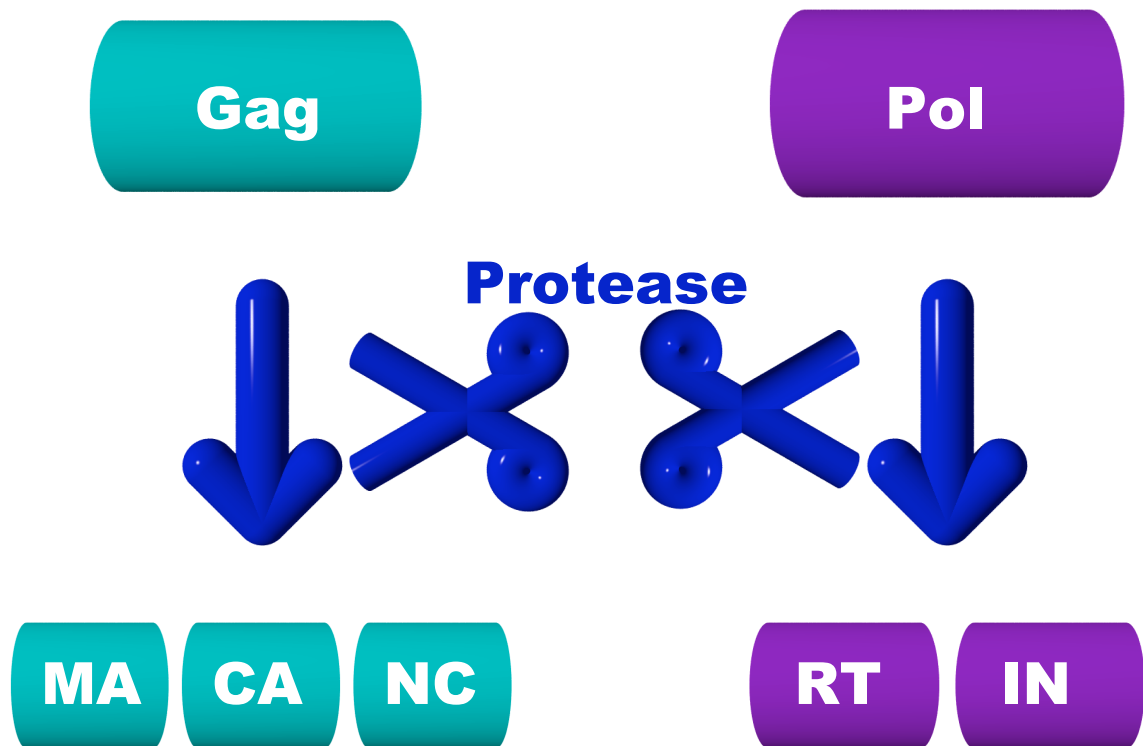


Figure 1.4. Action of HTLV-I protease in viral maturation. Gag polyprotein is cleaved into structural proteins matrix (MA), capsid (CA), and nucleocapsid (NC), while Pol polyprotein is cleaved into genome-processing proteins reverse transcriptase (RT) and integrase (IN).

Mature viral proteins and the transcribed viral RNA genome assemble near the cell membrane. The virus core acquires a lipid bilayer containing virally expressed surface proteins from the host cell membrane as it buds into a separate virus particle. The entire virus life cycle is summarized in Figure 1.5.

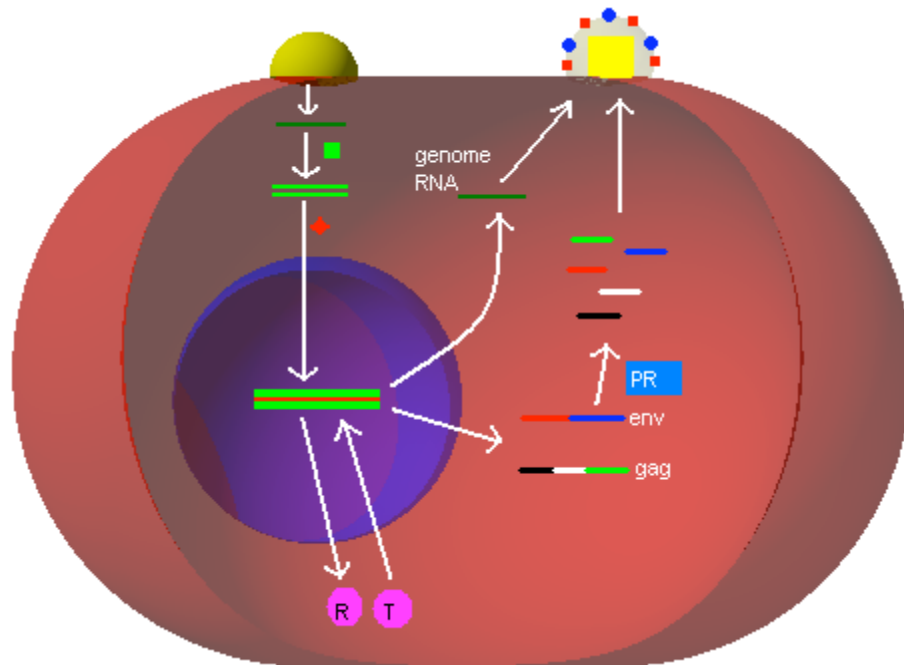


Figure 1.5. HTLV-I replication cycle. The virus particle (yellow sphere, top left) binds receptors of the cell surface and its membrane fuses with the cell membrane. The genomic RNA (dark green single strand) and reverse transcriptase (RT, green square) are inserted into the cell. RT transcribes the RNA genome to double-stranded DNA (light green) and, with integrase (red diamond), enters the nucleus (blue sphere, left). The virus genome is inserted into the host DNA by integrase, and regulatory proteins rex and tax are expressed. Tax re-enters the nucleus and binds the 5' LTR of the viral genome, triggering transcription of the genome, structural proteins, and replicative enzymes. The gag-pro-pol complex and env are expressed (right side); the protease (PR, blue square) processes gag-pro-pol into active form, and the virus proteins and genome assemble with env products at the cell membrane (top right). The completed virus buds from the cell, taking part of the cell membrane as part of its shell.

HTLV-I propagates by another method that occurs more frequently *in vivo* compared to the described process of particle formation. David Derse et al. (Derse et al., 2001) demonstrated that, while it is possible to generate virions *in vitro*, very few (as low as 1 in 10,000) are infectious. Indeed, cell-free transfusions seem to pose little risk. It is therefore more likely that HTLV-I spreads by cell-to-cell contact. This is supported by the observations of Igakura et al. (Igakura et al., 2003), who showed that HTLV-I can polarize the cell membrane of lymphocytes. Lymphocytes have a polarization mechanism that facilitates presentation of antigens. Virus-induced polarization leverages this system to assemble viral proteins and form a synapse between cells by which the virus is transferred. Other work has shown that virions need not be present to induce tumor formation. Transformation with the tax regulatory protein alone causes tumorigenesis (Grossman et al., 1995), though a means of introducing the tax gene into the cellular genome is still required. However, the protease is still necessary for viral maturation leading to the propagation of cell infection.

1.6 Pathology and treatment

HTLV-I has been shown to be the causative agent of ATL. On the cellular level, HTLV-I infection causes overexpression of IL-2 receptors, which bind IL-2, the T-cell specific growth factor (Wong-Staal and Gallo, 1985). Clinically, ATL manifests in several forms, linked by serologic analysis confirming presence of HTLV-I genes or gene products. The principal form of ATL manifests with hypercalcemia, lymphadenopathy, lymphomatous meningitis, involvement of the

liver or spleen, dermal involvement, and lytic bone lesions (Hjelle, 1991). This form is aggressive, with life expectancy of 6-18 months following emergence of symptoms. Patients with HTLV-I infection have also been characterized with non-Hodgkin's lymphoma, "smoldering" ATL with normal blood cell count but prominent skin lesions, and isolated T-cell cutaneous lymphoma.

Two other closely related conditions, HAM and TSP, are associated with HTLV-I infection. These neurological disorders are characterized by HTLV-I-specific T cell clones in the blood and spinal fluid. These disorders are progressive, with symptoms including sensory disturbance, muscle weakness in the extremities, urinary dysfunction, muscle spasticity, and hyper-reflexia (Hoger et al., 1997).

No specific treatments for HTLV-I have been approved. There have been several reports on the use of HIV-1 reverse transcriptase inhibitors to impair viral replication, but these are not an effective treatment option. There are literature reports of tests of HTLV-I protease inhibitors *in vitro* by Ding *et al.*, Akaji *et al.*, and others (Akaji, Teruya, and Aimoto, 2003; Ding, Rich, and Ikeda, 1998), but no cell-based assays or clinical trials have yet been reported.

1.7 References

- Akaji, K., Teruya, K., and Aimoto, S. (2003). Solid-phase synthesis of HTLV-1 protease inhibitors containing hydroxyethylamine dipeptide isostere. *J Org Chem* **68**(12), 4755-63.
- Baron, S. (1996). "Medical Microbiology." 4 ed. University of Texas Medical Branch at Galveston, Galveston.
- Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). "Retroviruses." Cold Spring Harbor Laboratory Press, Woodbury, NY.
- Derse, D., Hill, S. A., Lloyd, P. A., Chung, H., and Morse, B. A. (2001). Examining human T-lymphotropic virus type 1 infection and replication by cell-free infection with recombinant virus vectors. *J Virol* **75**(18), 8461-8.
- Ding, Y. S., Rich, D. H., and Ikeda, R. A. (1998). Substrates and inhibitors of human T-cell leukemia virus type I protease. *Biochemistry* **37**(50), 17514-8.
- Fraenkel-Conrat, H., Kimball, P. C., Levy, J. A. (1988). "Virology." 2 ed. Prentice Hall, Englewood Cliffs, NJ.
- Gallo, R. C., Thomson, M. M. (1996). Introduction. In "Human T-cell Lymphotropic Virus Type I" (P. Hollsberg, Hafler, D. A., Ed.), pp. 325. Wiley and Sons, Chichester.
- Gessain, A. (1996). Epidemiology of HTLV-I and associated diseases. In "Human T-Cell Lymphotropic Virus Type I" (P. Hollsberg, and D. A. Hafler, Eds.), pp. 31-64. Wiley, Chichester, UK.
- Gessain, A., Barin, F., Vernant, J. C., Gout, O., Maurs, L., Calender, A., and de The, G. (1985). Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* **2**(8452), 407-10.
- Grossman, W. J., Kimata, J. T., Wong, F. H., Zutter, M., Ley, T. J., and Ratner, L. (1995). Development of leukemia in mice transgenic for the tax gene of human T-cell leukemia virus type I. *Proc Natl Acad Sci U S A* **92**(4), 1057-61.
- Hardy, W. D., Jr., Old, L. J., Hess, P. W., Essex, M., and Cotter, S. (1973). Horizontal transmission of feline leukaemia virus. *Nature* **244**(5414), 266-9.
- Hjelle, B. (1991). Human T-cell leukemia/lymphoma viruses. Life cycle, pathogenicity, epidemiology, and diagnosis. *Arch Pathol Lab Med* **115**(5), 440-50.
- Hoger, T. A., Jacobson, S., Kawanishi, T., Kato, T., Nishioka, K., and Yamamoto, K. (1997). Accumulation of human T lymphotropic virus (HTLV)-I-specific T cell clones in HTLV-I-associated myelopathy/tropical spastic paraparesis patients. *J Immunol* **159**(4), 2042-8.
- Igakura, T., Stinchcombe, J. C., Goon, P. K., Taylor, G. P., Weber, J. N., Griffiths, G. M., Tanaka, Y., Osame, M., and Bangham, C. R. (2003). Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science* **299**(5613), 1713-6.

- Morrow, C. D., Park, J., and Wakefield, J. K. (1994). Viral gene products and replication of the human immunodeficiency type 1 virus. *Am J Physiol* **266**(5 Pt 1), C1135-56.
- Murphy, F. A. (1995). "Virus taxonomy: classification and nomenclature of viruses: sixth report of the International Committee on Taxonomy of Viruses." Archives of virology. Springer-Verlag, New York City.
- Nam, S. H., Copeland, T. D., Hatanaka, M., and Oroszlan, S. (1993). Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *J Virol* **67**(1), 196-203.
- Nam, S. H., Kidokoro, M., Shida, H., and Hatanaka, M. (1988). Processing of gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. *J Virol* **62**(10), 3718-28.
- 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 **77**, 7415-7419.
- Rous, P. (1911). Transmission of a malignant new growth by means of a cell-free filtrate. *J. Am. Med. Assoc.* **56**, 198.
- Voet, D., Voet, J. G. (1995). "Biochemistry." 2 ed. John Wiley and Sons, New York City.
- Wong-Staal, F., and Gallo, R. C. (1985). Human T-lymphotropic retroviruses. *Nature* **317**(6036), 395-403.

CHAPTER 2

HTLV-I PROTEASE: A RETROVIRAL PROTEASE

2.1 Proteases

Enzymes are proteins that are able to catalyze biologically important reactions. Proteases are enzymes that are able to catalyze the cutting of other proteins. Proteases serve a number of functions in organisms, including the breakdown of proteins as part of digestion, for example, trypsin and pepsin. Other proteases cleave inactive precursor proteins to form the active protein. For example, as part of the blood clotting mechanism, thrombin cleaves a specific peptide bond in fibrinogen, forming fibrin (Voet, 1995).

2.2 Classification of proteases

The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology classifies proteases in subgroup 4 of the hydrolases group of enzymes. Proteases in the subclass are further divided into families based on the type of reaction that is catalyzed, active site residues, and evolutionary relation in terms of sequence and structure (Rao et al., 1998).

Proteases are collected in an online database called MEROPS (Rawlings, O'Brien, and Barrett, 2002). In this system, proteases are classified by active site residue, clan, and family. Top-level classification of proteases is based on the catalytic group. Most known proteases utilize either an aspartic acid, serine,

cysteine, or bound metal ion or metalloprotease; however, there is a fifth group, proteases of mixed or unknown active site. Clan grouping of proteases is based on apparent evolutionary origin. In the case of HTLV-I protease, the protease sequence is thought to be derived from a pepsin-like aspartic acid protease ancestor. Therefore, it is placed in clan AA with pepsin, retropepsin, and pararetrovirus proteases. However, while these three representative proteases have evolutionary relation, they vary significantly in sequence and structure and are classified into separate families. Pepsin is classified in family A1, a family whose members have aspartic acid as the catalytic residue and which are monomers with two lobes, each with a catalytic aspartic acid. Retroviral proteases are members of the A2 family. Members of this family are structurally similar to pepsin with two lobes, but they are homodimers. The two monomers, each with one active aspartic acid, dimerize to form the active bilobed structure (Rawlings and Barrett, 1995). A significant difference between A1 and A2 families is that A2 family proteases are sequentially identical dimers, while the monomers of A1 family proteases have topologically similar lobes that do not have identical sequence.

The structures of several A2-family proteases have been determined experimentally. The first retroviral protease to be crystallized was RSV protease, in 1988 (Miller, Leis, and Wlodawer, 1988), while HIV-1 protease is perhaps the most widely studied, with some 200 structures of various mutants and inhibitor complexes reported to date (HIVdb, <http://mcl1.ncifcrf.gov/hivdb/index.html>). The structures of HIV-1 protease (PDB:1D4L) (Tyndall et al., 2002), HIV-2 protease

(PDB:1IDA) (Tong et al., 1993), simian immunodeficiency virus (SIV) protease (PDB:1YTH) (Rose et al., 1996), feline immunodeficiency virus (FIV) protease (PDB:4FIV) (Kervinen et al., 1998), Rous sarcoma virus (RSV) protease (PDB:1BAI) (Wu et al., 1998), avian myeloblastosis associated virus (MAV) protease (PDB:1MVP) (Ohlendorf et al., 1992), and equine infectious anemia virus (EIAV) protease (PDB:1FMB) (Gustchina et al., 1996) have been determined experimentally. Representative structures of these retroviral proteases, as well as a structure of pepsin (PDB:1PSO) (Sielecki et al., 1990), are shown in Figure 2.1. The structures of HTLV-I protease and related BLV protease have not been experimentally determined. Theoretical, computer-generated models of HTLV-I protease have been reported ((Tözsér et al., 2000), Shuker SB et al., PDB file 1O0J).

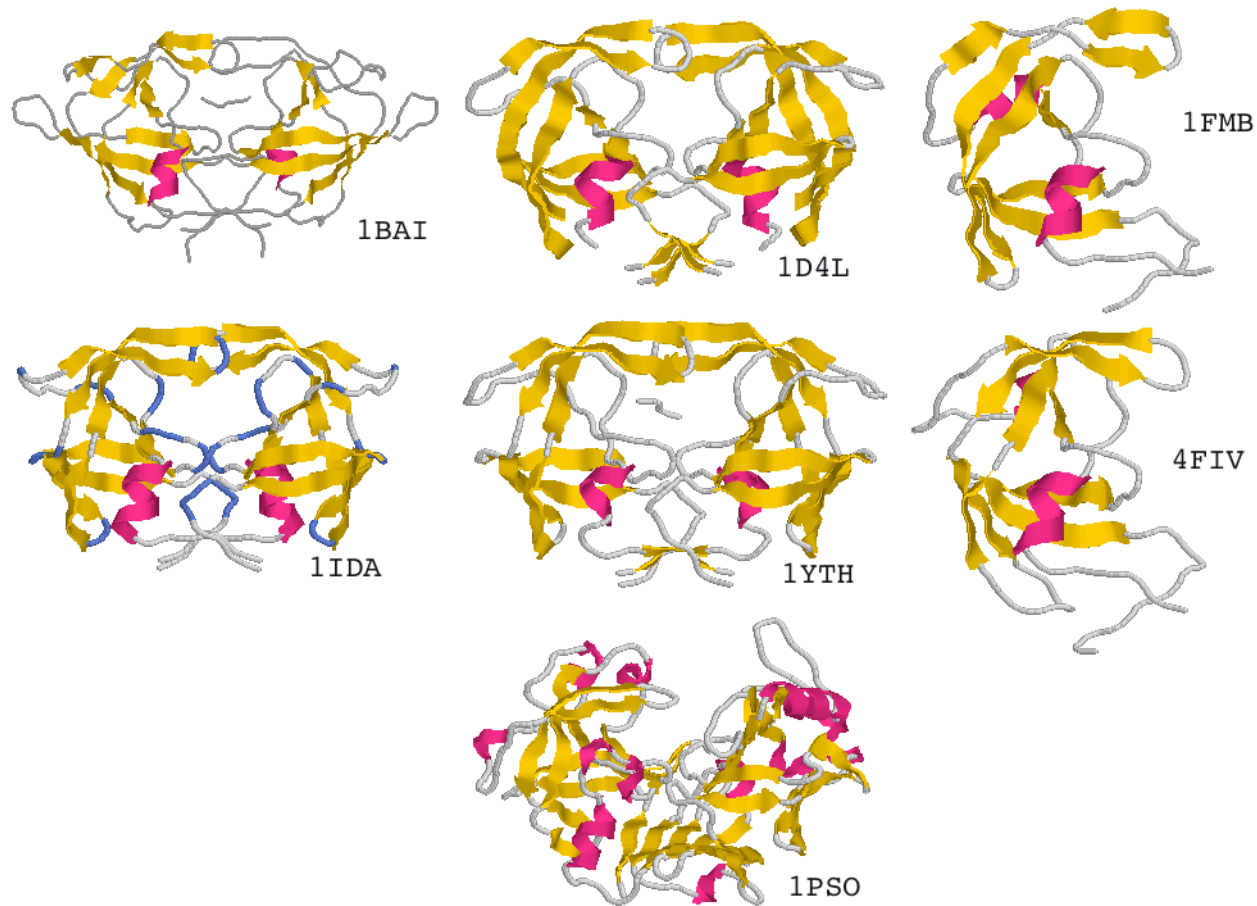


Figure 2.1. Structures of retroviral proteases. 1BAI, RSV protease, 1D4L, HIV-1 protease; 1FMB, EIAV protease; 1IDA, HIV-2 protease; 1YTH, SIV protease; 4FIV, FIV protease; 1PSO, pepsin. In 1FMB and 4FIV, only one chain of the homodimer was stored in the PDB. 1BAI and 1YTH are shown with bound inhibitor. (References in text)

2.3 Aspartic acid proteases: structure and mechanism

Retroviral protease structures follow a general template (Wlodawer and Gustchina, 2000), which is depicted in Figure 2.2. The secondary structure, from N-terminal to C-terminal, consists of the following: three to four anti-parallel beta sheets comprising a hairpin and a wide loop which contains the catalytic aspartic acid, followed by a helical region, an anti-parallel pair of beta sheets that form a hairpin observed as a flap, two to three anti-parallel beta sheets comprising a hairpin and wide loop, a helical region, and terminal beta sheet. The dimer contains a single active site formed by residues from the two identical monomers, a topology that is unique to retroviral proteases.

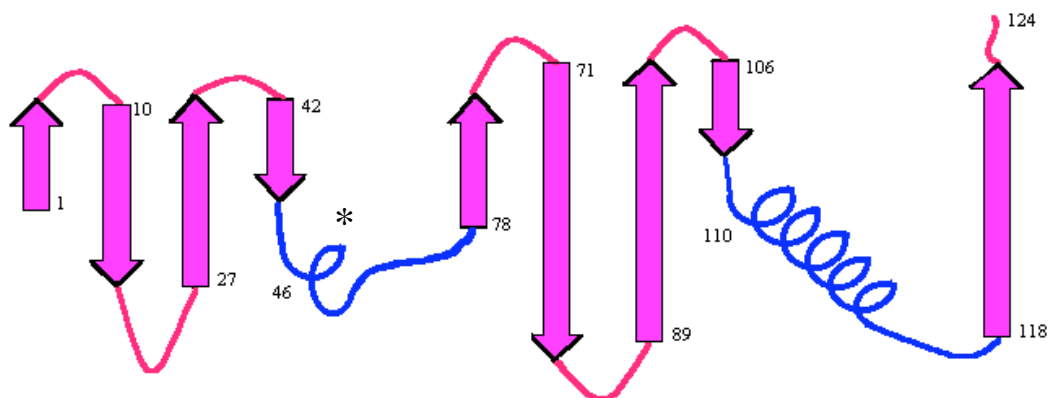


Figure 2.2. Schematic of secondary structural elements in retroviral proteases. Asterisk indicates position of catalytic aspartic acid.

The binding pockets of proteases correspond to amino acids in the substrate. There is a naming convention for binding sites and positions in the substrate adopted from the work of Schechter and Berger (Schechter and Berger, 1967; Schechter and Berger, 1968). From N-terminal to C-terminal

relative to the substrate, binding sites are named S4 to S4', and positions on the substrate are named P4 to P4'. The bond that will be cleaved, the scissile bond, is in the middle between P1 and P1' in the substrate. Figure 2.3 shows the relation of substrate, binding sites, and scissile bond.

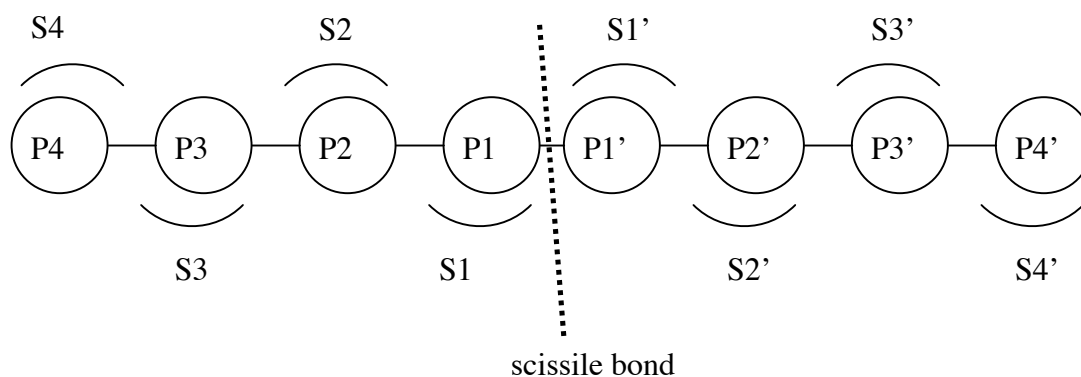


Figure 2.3. Schematic diagram of substrate positions and binding pockets. Circles represent substrate amino acids, and arcs represent sites on the protease. The scissile bond is the point of cleavage.

Structural and computer studies of HIV-1 protease have suggested a mechanism of catalysis (Liu, Muller-Plathe, and van Gunsteren, 1996). The mechanism is a general acid-base mechanism involving the two active aspartic acid residues and water. A tetrahedral intermediate is formed by water attacking the carbon of the amide bond to be cleaved. The conjugate base aspartate abstracts a proton from the intermediate, restoring the carbonyl group and displacing the nitrogen of the amide bond. The nitrogen simultaneously abstracts a proton from the aspartic acid. This mechanism is diagrammed in Figure 2.4.

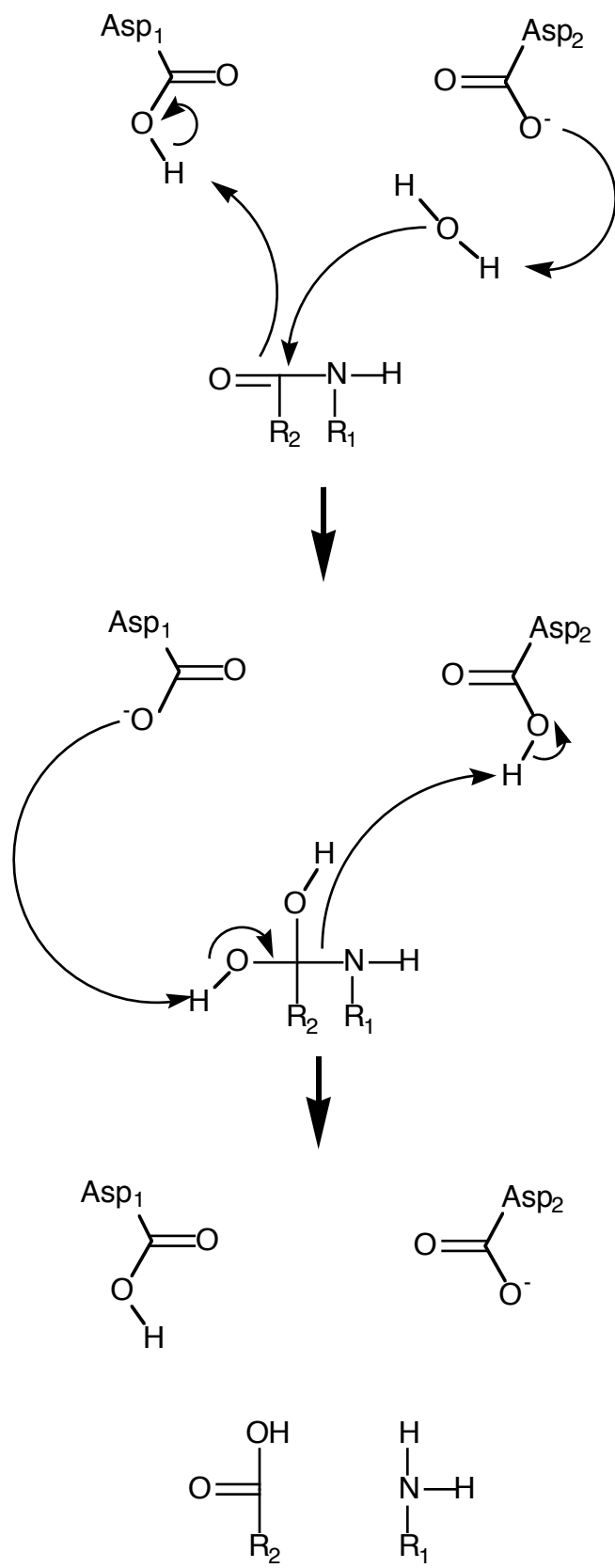


Figure 2.4. Mechanism of aspartic acid proteases.

2.4 HTLV-I protease: background

HTLV-I protease is a homodimer composed of two 125-amino acid monomers (Kobayashi et al., 1991). The DNA and amino acid sequence of this protease is shown in Figure 2.5. The monomers each carry one catalytic aspartic acid and fold together to form the active site and binding pockets. Little has been reported about the structure of HTLV-I protease, particularly the binding pockets. Knowledge of the binding pockets would provide useful information about the electrostatic and steric contacts that are necessary for substrate recognition and inhibitor binding. The following summarizes current hypotheses about the binding sites of HTLV-I protease. These are based principally on the two published theoretical structures. A summary of the substrate preferences of HTLV-I protease, and comparison to HIV-1 protease, is given in Table 2.1 (Tözsér et al., 2000).

P	V	I	P	L	D	P	A	R	R	P	V	I	K	A	
CCA	GTT	ATA	CCG	TTA	GAT	CCC	GCC	CGT	CGG	CCC	GTA	ATT	AAA	GCC	15
Q	V	D	T	Q	T	S	H	P	K	T	I	E	A	L	
CAG	GTT	GAC	ACC	CAG	ACC	AGC	CAC	CCA	AAG	ACT	ATC	GAA	GCT	CTA	30
L	D	T	G	A	D	M	T	V	L	P	I	A	L	F	
CTA	GAT	ACA	GGA	GCA	GAC	ATG	ACA	GTC	CTT	CCG	ATA	GCC	TTG	TTC	45
S	S	N	T	P	L	K	N	T	S	V	L	G	A	G	
TCA	AGT	AAT	ACT	CCC	CTC	AAA	AAT	ACA	TCC	GTA	TTA	GGG	GCA	GGG	60
G	Q	T	Q	D	H	F	K	L	T	S	L	P	V	L	
GGC	CAA	ACC	CAA	GAT	CAC	TTT	AAG	CTC	ACC	TCC	CTT	CCT	GTG	CTA	75
I	R	L	P	F	R	T	T	P	I	V	L	T	S	C	
ATA	CGC	CTC	CCT	TTC	CGG	ACA	ACG	CCT	ATT	GTT	TTA	ACA	TCT	TGC	90
L	V	D	T	K	N	N	W	A	I	I	G	R	D	A	
CTA	GTT	GAT	ACC	AAA	AAC	AAC	TGG	GCC	ATC	ATA	GGT	CGC	GAT	GCC	105
L	Q	Q	C	Q	G	V	L	Y	L	P	E	A	K	G	
TTA	CAA	CAA	TGC	CAG	GGC	GTC	CTG	TAC	CTC	CCT	GAG	GCA	AAA	GGG	120
P	P	V	I	L	STOP										
CCG	CCT	GTA	ATC	TTG	TGA										

Figure 2.5. HTLV-I protease DNA and amino acid sequence. The residue and corresponding codon in outline represent the catalytic aspartic acid.

Position	Preferred residues	
	HIV-1 protease	HTLV-I protease
P4	Ser, Thr (small, polar)	Thr, Val, Leu (small aliphatic side chain)
P3	Phe, Leu (large hydrophobic, aromatic)	Lys, Phe, Leu, Val, Ala
P2	Asn, Cys (small, polar)	Val, Leu, Ile (small, hydrophobic)
P1	Leu, Phe (hydrophobic)	Leu, Phe (hydrophobic)
P1'	Val, Leu, Ile, Phe (hydrophobic)	Val, Leu, Pro (hydrophobic, smaller than P1)

Table 2.1. Comparison of substrate preferences of HTLV-I and HIV-1 proteases.

The S4 and S4' binding sites are shallow, surface-exposed pockets. In most substrates, the residues at P4' and P4 are hydrophobic, with proline being most prevalent. A variety of amino acids are tolerated at these sites; for example, the capsid-nucleocapsid junction (CA/NC) has the polar residue threonine at P4. However, N-terminal truncation is not well tolerated. A synthetic nine-residue substrate without a P5 residue showed reduced cleavage, while an eight-residue substrate without a P4 residue exhibited no detectable cleavage (Tözsér et al., 2000).

The S3 and S3' binding sites are large and able to bind a variety of amino acids. Hydrophobic residues such as phenylalanine and hydrophilic residues such as lysine are accommodated here with similar catalytic efficiency (Tözsér et al., 2000). The S2 and S2' binding sites are hydrophobic. These sites are typically occupied by small aliphatic hydrophobic residues such as valine, leucine, and isoleucine.

The S1 binding site is hydrophobic and appears to be sterically hindered. Leucine is the only residue that appears in this position in known natural substrates. Synthetic substrate studies have shown that phenylalanine is also tolerated at this site; this is interesting considering that phenylalanine is much larger than leucine, and it is common in HIV-1 protease substrates (Shuker et al., 2003). Steric hindrance is indicated by the exclusion of beta-branched residues from the P1 position. This may occur because the substrate backbone fits in a narrow channel between Gly30 and Leu34 on opposite strands, excluding beta-branched amino acids from the S1 site. In contrast, the S1' binding site prefers proline, but the beta-branched residue valine is common here, as well as glycine and serine.

2.5 Substrates and inhibitors of HTLV-I protease

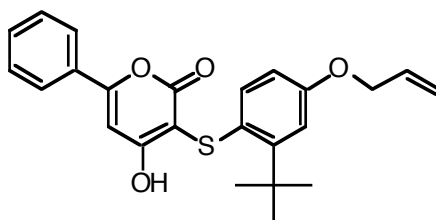
HTLV-I protease has a unique substrate specificity. The specificity of HTLV-I protease restricts the protease from binding and cleaving non-viral peptide sequences; however, it also prevents it from binding to inhibitors designed for other proteases. The native cleavage sites of HTLV-I protease have been experimentally determined. Table 2.2 summarizes the native substrates of HTLV-I protease.

JUNCTION	CLEAVAGE SITE
MA/CA	APQVL/PVMHP
CA/NC	KTKVL/VVQPK
TF1/PR	PASIL/PVIPL
PR/TFP	PPVIL/PIQAP
TFP/RT	APAVL/GLEHL
RT-RH/IN	PVLQL/SPADL

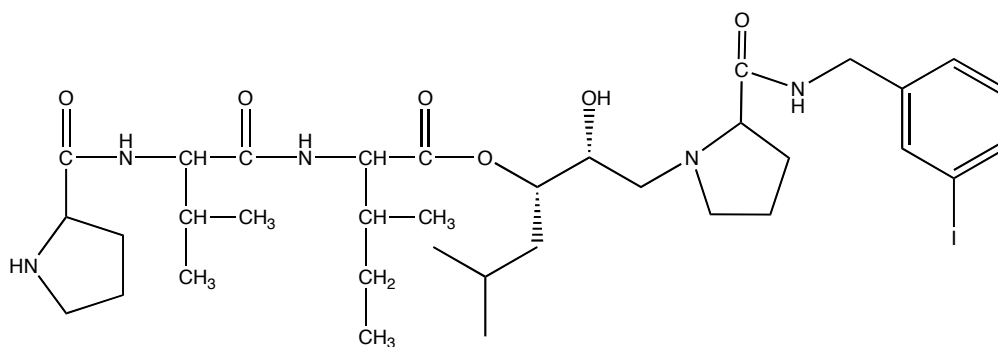
Table 2.2. Native cleavage sites of HTLV-I protease.

HTLV-I protease is weakly inhibited by pepsin inhibitors, because it is also an aspartic acid protease, and by HIV-1 inhibitors because of structural similarity. Previous works have reported data for inhibition of HTLV-I protease with pepsin inhibitors and HIV-1 inhibitors (Shuker et al., 2003); several effective inhibitors are presented in Table 2.3. Since HTLV-I protease has different substrate specificity from pepsin and HIV-1 protease, it follows that HTLV-I will not be effectively inhibited by compounds designed to target these proteases. Previous work shows that peptide-like compounds provide superior inhibition of all the compounds studied to date. In particular, JG-365 (Figure 2.6a), an HIV-1 protease inhibitor, is currently the most effective inhibitor, with K_i of 6.0 nM. A peptide-based, HTLV-I-specific inhibitor, shown in Figure 2.6b, was recently reported (Akaji, Teruya, and Aimoto, 2003), with K_i of 38 nM. However, few non-peptide inhibitors have been tested with HTLV-I protease; MES13-099 (Figure 2.6c), with K_i of 243 nM, is the only one currently featured in the literature (Ding, Rich, and Ikeda, 1998).

(a)



(b)



(c)

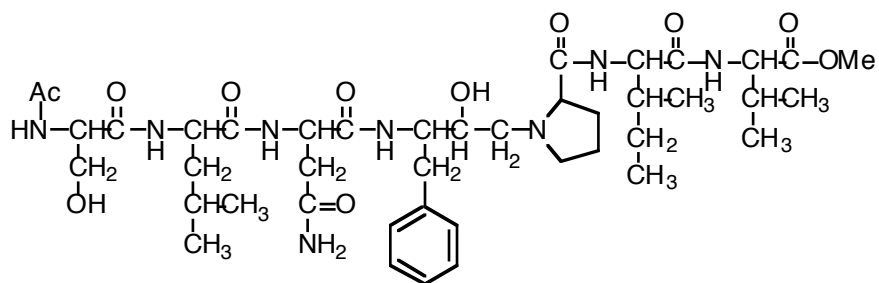


Figure 2.6. HIV-1 protease inhibitors. (a) MES13-099, (b) Compound 31 (Akaji *et al.*), (c) JG-365

Inhibitor (target)	K _i (pepsin) (nM)	K _i (HIV-1 protease) (nM)	K _i (HTLV-I protease) (nM)
CS-I-25 (pepsin)	19.2	N/A	7.2
CS-I-27 (pepsin)	0.22	N/A	142
MES13-099 (HIV-1)	N/A	7	243
JG-365 (HIV-1)	N/A	0.66	6

Table 2.3. Effect of inhibitors on three aspartic acid proteases: pepsin, HIV-1 protease, and HTLV-I protease.

2.6 HTLV-I protease expression, purification and assay

Several groups have reported cloning and purification of HTLV-I protease. HTLV-I protease has been totally synthesized (Teruya et al., 2002), but it is also common to express the protease in a bacterial system (Ding et al., 1998). Previous work has reported yields up to 150 mg/L of culture by expression in *E. coli* (Ha et al., 2002). Mutations have also been tested and found to improve protease stability. In particular, mutation of Leu40 to Ile abolishes autoprocessing at this position. The native sequence around Leu40 (outlined here) is MTVL/PIAL; VL/PI is present in other native cleavage junctions. Substitution of Ile at the P1' site inhibits cleavage. Mutations to replace cysteines at positions 90 and 109 do not affect catalytic activity (Louis, Oroszlan, and Tozser, 1999).

Two methods of purification from cell culture have been reported. The first includes size exclusion chromatography followed by HPLC (Louis, Oroszlan, and Tozser, 1999). Other work reports functionalizing HTLV-I protease with a polyhistidine tag in order to purify the protease by metal affinity chromatography (Ding et al., 1998). In both cases, purification under denaturing conditions is

necessary, followed by refolding into buffer in which the protease is active. Previous work also describes a method of determining protease concentration by measuring absorbance at 280 nm (Ha et al., 2002). Determination of protease concentration is essential for kinetic experiments.

Two protease assays have been devised to determine kinetics and inhibition of HTLV-I protease. Analytical HPLC is commonly used to measure cleavage (Ding, Rich, and Ikeda, 1998; Louis, Oroszlan, and Tozser, 1999; Tözsér et al., 2000). The other major method involves a fluorogenic substrate. In this case, a peptide matching a native cleavage sequence is functionalized with a fluorophore and quencher. The intact substrate self-quenches and emits a low background fluorescence, a process called fluorescence resonance energy transfer or FRET (Selvin, 1995). Cleavage by the protease separates the fluorophore and quencher, allowing detectable fluorescence. Dabcyl and EDANS were attached to the matrix-capsid junction PQVLPVMH to produce a fluorogenic substrate for HTLV-I protease, which is depicted in Figure 2.7. This substrate is excited at 340 nm, and emits light at 490 nm upon cleavage.

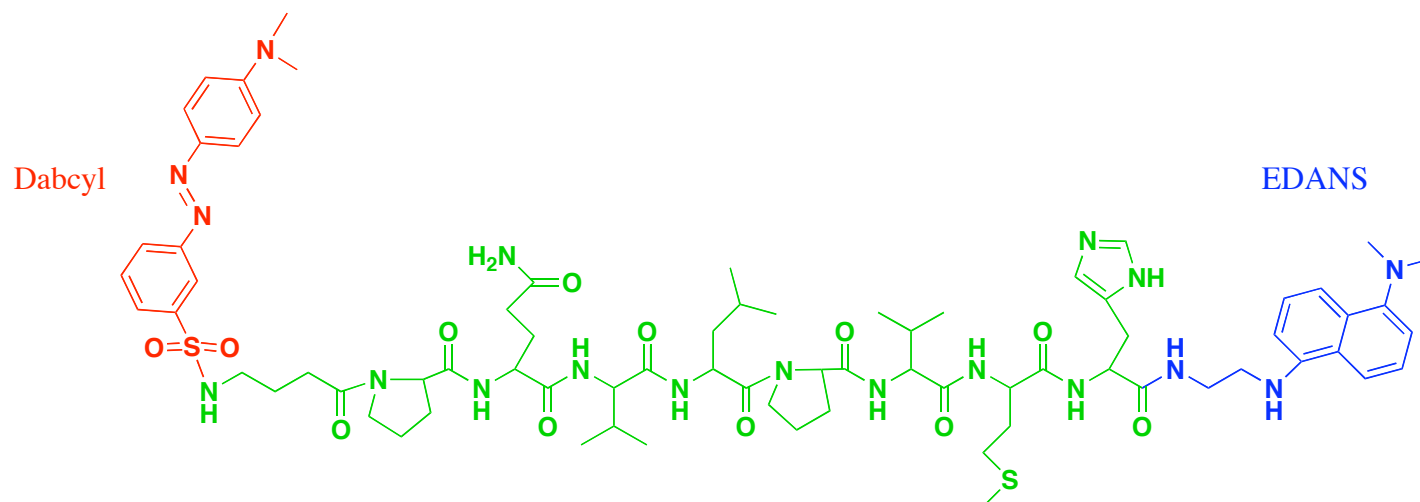


Figure 2.7. Fluorogenic substrate for HTLV-I protease. This substrate is based on the matrix-capsid junction, PQVLPVMH (green), with fluorophore EDANS in blue, and quencher dabcyl in red.

2.7 Significance of present work

The present work develops more information of the enzyme kinetics of HTLV-I protease. In particular, the present work describes a new theoretical model of the protease. Experiments evaluate the effect of individual residues on substrate binding and catalysis. These data clarify the mechanism of substrate recognition. This information can be used to verify and refine structural models, as well as to identify important binding interactions that can be exploited in designing protease inhibitors. As with other retroviruses, the protease of HTLV-I is essential for maturation of viral proteins. Inhibition of the protease will slow or prevent virus replication. Therefore, HTLV-I protease is an attractive pharmaceutical target.

2.8 References

- Akaji, K., Teruya, K., and Aimoto, S. (2003). Solid-phase synthesis of HTLV-1 protease inhibitors containing hydroxyethylamine dipeptide isostere. *J Org Chem* **68**(12), 4755-63.
- Ding, Y. S., Owen, S. M., Lal, R. B., and Ikeda, R. A. (1998). Efficient expression and rapid purification of human T-cell leukemia virus type 1 protease. *J Virol* **72**(4), 3383-6.
- Ding, Y. S., Rich, D. H., and Ikeda, R. A. (1998). Substrates and inhibitors of human T-cell leukemia virus type I protease. *Biochemistry* **37**(50), 17514-8.
- Gustchina, A., Kervinen, J., Powell, D. J., Zdanov, A., Kay, J., and Wlodawer, A. (1996). Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci* **5**(8), 1453-65.
- Ha, J. J., Gaul, D. A., Mariani, V. L., Ding, Y. S., Ikeda, R. A., and Shuker, S. B. (2002). HTLV-1 protease cleavage of p19/24 substrates is not dependent on NaCl concentration. *Bioorganic Chemistry* **30**(2), 138-144.
- Kervinen, J., Lubkowski, J., Zdanov, A., Bhatt, D., Dunn, B. M., Hui, K. Y., Powell, D. J., Kay, J., Wlodawer, A., and Gustchina, A. (1998). Toward a universal inhibitor of retroviral proteases: comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. *Protein Sci* **7**(11), 2314-23.
- Kobayashi, M., Ohi, Y., Asano, T., Hayakawa, T., Kato, K., Kakinuma, A., and Hatanaka, M. (1991). Purification and characterization of Human T-Cell Leukemia- Virus Type I protease produced in Escherichia coli. *Febs Letters* **293**(1-2), 106-110.
- Liu, H., Muller-Plathe, F., and van Gunsteren, W. F. (1996). A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease **261**(3), 454-69.
- Louis, J. M., Oroszlan, S., and Tozser, J. (1999). Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. *J Biol Chem* **274**(10), 6660-6.
- Miller, M., Leis, J., and Wlodawer, A. (1988). Preliminary crystallographic study of a retroviral protease. *J Mol Biol* **204**(1), 211-2.
- Ohlendorf, D. H., Foundling, S. I., Wendoloski, J. J., Sedlacek, J., Strop, P., and Salemme, F. R. (1992). Structural studies of the retroviral proteinase from avian myeloblastosis associated virus. *Proteins* **14**(3), 382-91.
- Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* **62**(3), 597-635.
- Rawlings, N. D., and Barrett, A. J. (1995). Families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods Enzymol* **248**, 105-20.
- Rawlings, N. D., O'Brien, E., and Barrett, A. J. (2002). MEROPS: the protease database. *Nucleic Acids Res* **30**(1), 343-6.

- Rose, R. B., Craik, C. S., Douglas, N. L., and Stroud, R. M. (1996). Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry* **35**(39), 12933-44.
- Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**(2), 157-62.
- Schechter, I., and Berger, A. (1968). On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem Biophys Res Commun* **32**(5), 898-902.
- Selvin, P. R. (1995). Fluorescence resonance energy transfer. *Methods Enzymol* **246**, 300-34.
- Shuker, S. B., Mariani, V. L., Herger, B. E., and Dennison, K. J. (2003). Understanding HTLV-I Protease. *Chem Biol* **10**(5), 373-80.
- Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. (1990). Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J Mol Biol* **214**(1), 143-70.
- Teruya, K., Kawakami, T., Akaji, K., and Aimoto, S. (2002). Total synthesis of [L40I, C90A, C109A]-human T-cell leukemia virus type 1 protease. *Tetrahedron Letters* **43**(8), 1487-1490.
- Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., and Anderson, P. C. (1993). Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures. *Proc Natl Acad Sci U S A* **90**(18), 8387-91.
- Tözsér, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W., and Weber, I. T. (2000). Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *European Journal of Biochemistry* **267**(20), 6287-6295.
- Tyndall, J. D., Reid, R. C., Tyssen, D. P., Jardine, D. K., Todd, B., Passmore, M., March, D. R., Pattenden, L. K., Bergman, D. A., Alewood, D., Hu, S. H., Alewood, P. F., Birch, C. J., Martin, J. L., and Fairlie, D. P. (2002). Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macrocyclic inhibitors of HIV-1 protease **43**, 3495-3504.
- Voet, D., Voet, J. G. (1995). "Biochemistry." 2 ed. John Wiley and Sons, New York City.
- Wlodawer, A., and Gustchina, A. (2000). Structural and biochemical studies of retroviral proteases. *Biochim Biophys Acta* **1477**(1-2), 16-34.
- Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W., and Weber, I. T. (1998). Structural basis for specificity of retroviral proteases **37**, 4518-4526.

CHAPTER 3

DEVELOPMENT OF A THEORETICAL MODEL AND FUNCTION OF THE TEN C-TERMINAL RESIDUES

3.1 Similarity of HTLV-I protease to other retroviral proteases

HTLV-I protease is a member of the retroviral aspartic acid protease family. As previously noted, these proteases have similar sequences and similar overall folds and structures, but varying substrate specificity. No experimental structure of HTLV-I protease has been reported.

Structural information is essential to understanding substrate binding and inhibitor design. It is possible to use proteases with known structure as a template for the prediction of HTLV-I's structure by comparing their sequences. The quality of the predicted structure depends on the homology or identity of the proteases being studied. The homology of retroviral proteases is presented in Table 3.1 (Shuker et al., 2003).

	HTLV-I	HTLV-2	BLV	MuLV	RSV	HIV-1	HIV-2	SIV	FIV	EIAV
HTLV-I	-	50%	36%	21%	25%	28%	29%	32%	23%	26%
HTLV-2	50%	-	33%	25%	23%	26%	24%	26%	21%	25%
BLV	36%	33%	-	27%	23%	24%	25%	25%	21%	26%
MuLV	21%	25%	27%	-	21%	23%	24%	27%	25%	24%
RSV	25%	23%	23%	21%	-	30%	26%	25%	26%	26%
HIV-1	28%	26%	24%	23%	30%	-	48%	50%	25%	29%
HIV-2	29%	24%	25%	24%	26%	48%	-	86%	24%	33%
SIV	32%	26%	25%	27%	25%	50%	86%	-	22%	35%
FIV	23%	21%	21%	25%	26%	25%	24%	22%	-	30%
EIAV	26%	25%	26%	24%	26%	29%	33%	35%	30%	-

Table 3.1. Homology, or percent identical sequence, of various retroviral proteases.

3.2 Sequence alignment and generation of theoretical models

The software package MODELLER is able to generate sequence alignments between proteins of known and unknown structure. It can then use these sequence alignments, with known structures as templates, to predict the structure of the unknown protein (Fiser, Do, and Sali, 2000; Marti-Renom et al., 2000; Sali and Blundell, 1993). In the absence of an experimentally determined structure of HTLV-I protease, MODELLER was to generate a theoretical model of the protease for prediction of structural features.

A multiple alignment of several retroviral proteases is shown in Figure 3.1. In order to generate a structural model, two proteases with known structure were selected: RSV protease (PDB file 1BAI) (Wu et al., 1998), and HIV-1 protease (1D4L) (Tyndall et al., 2002). These proteases and structures were selected based on homology and quality of the published structures.

The sequence alignment used by MODELLER, in MODELLER alignment file format, is shown in Figure 3.2(a). This file is essentially a machine-readable form of the above noted multiple alignment, including only the sequences for HTLV-I, HIV-1, and RSV proteases. The MODELLER script used to generate the theoretical model is shown in Figure 3.2(b). This script reads the alignment file, sets default parameters, and generates the model.

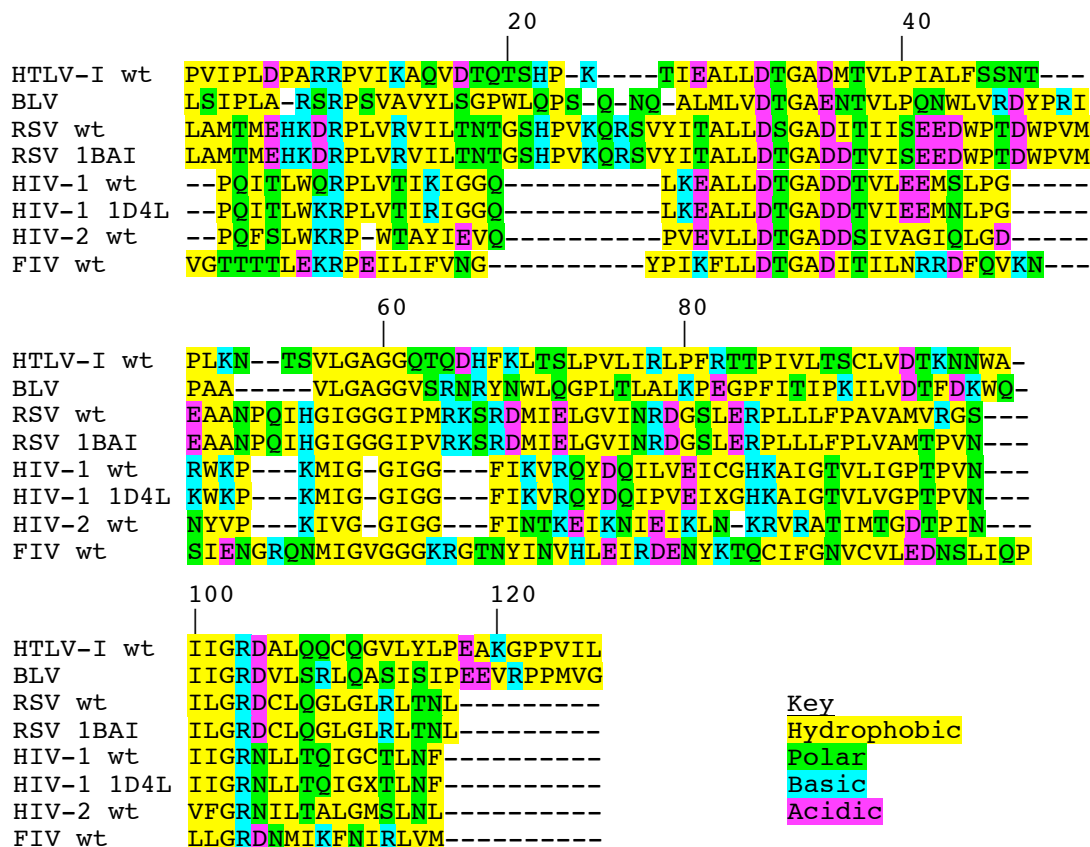


Figure 3.1. Sequence alignment of several retroviral proteases.

```

(a)
C; alignment in the PIR format
>P1;1d4l
structureX:1d4l:1      :A:99      :B:HIV-1 protease:human: 1.75: 0.19
--PQITLWKRPLVTIRIGGQ-----LKEALLDTGADDTVIEEMNLPG-----KWKP--
KMIGGIGG-----FIKV
RQYDQIPVEI.GHKAIGTVLVGPTPVN---IIGRNLLTQIG.TLNF-----/
--PQITLWKRPLVTIRIGGQ-----LKEALLDTGADDTVIEEMNLPG-----KWKP--
KMIGGIGG-----FIKV
RQYDQIPVEI.GHKAIGTVLVGPTPVN---IIGRNLLTQIG.TLNF-----*
>P1;1bai
structureX:1bai:@      :A:X      :B:HIV-1 protease:human: 2.40: 0.17
LAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDTGADDTVISEEDWPTDWPVMEANPQIHGIGGGI
PVRKSRDMI
ELGVINRDGSLERPLLLFPLVAMTPVN---ILGRDCLQGLGLRLTNL-----/
LAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDTGADDTVISEEDWPTDWPVMEANPQIHGIGGGI
PVRKSRDMI
ELGVINRDGSLERPLLLFPLVAMTPVN---ILGRDCLQGLGLRLTNL-----*
>P1;3htl
sequence:3htl:      : :      : :HTLV-1 protease:human: 2.00:-1.00
PVIPLDPARRPVKAQVDTQTSHK-K----TIEALLDTGADMTVLPALFSSNT---PLKN--
TSVLGAGGQTQDHFGLT
SLPVLIRLPFRTPPIVLTSCSLVDTKNNWA-IIGRDALQQCQGVLYLPEAKGPPVIL/
PVIPLDPARRPVKAQVDTQTSHK-K----TIEALLDTGADMTVLPALFSSNT---PLKN--
TSVLGAGGQTQDHFGLT
SLPVLIRLPFRTPPIVLTSCSLVDTKNNWA-IIGRDALQQCQGVLYLPEAKGPPVIL*

```

```

(b)
# A sample TOP file for fully automated comparative modeling

INCLUDE                                # include MODELLER
routines
SET ALNFILE      = 'htlv-dimer.ali' # input file w/ templates and
target
SET KNOWN        = '1d4l' '1bai' # templates' PDB codes
SET SEQUENCE     = '3htl'        # target code
SET ATOM_FILES_DIRECTORY = './' # directory with input atom files
SET STARTING_MODEL= 1
SET ENDING_MODEL= 5
SET OUTPUT_CONTROL = 1 1 1 1 2
CALL ROUTINE      = 'model'          # get alignment and a model

```

Figure 3.2. Scripts used by MODELLER to generate the theoretical model. (a) Sequence alignment file. (b) Script to generate the model from the alignment and template PDB files.

The theoretical model of the full-length, wild-type HTLV-I protease is shown in Figure 3.3. It is noted from the alignments that the last ten C-terminal residues do not correspond to any portion of the sequence of HIV-1 protease or RSV protease. The ten C-terminal residues were omitted from the alignment, and the MODELLER script was run again to generate a structure of the truncated protease. This structure is shown in Figure 3.4. The core structure was not significantly altered by the truncation, and data presented in Section 3.5 indicate that it is not essential for catalysis. The truncated protease structure was used to prepare the published PDB model.

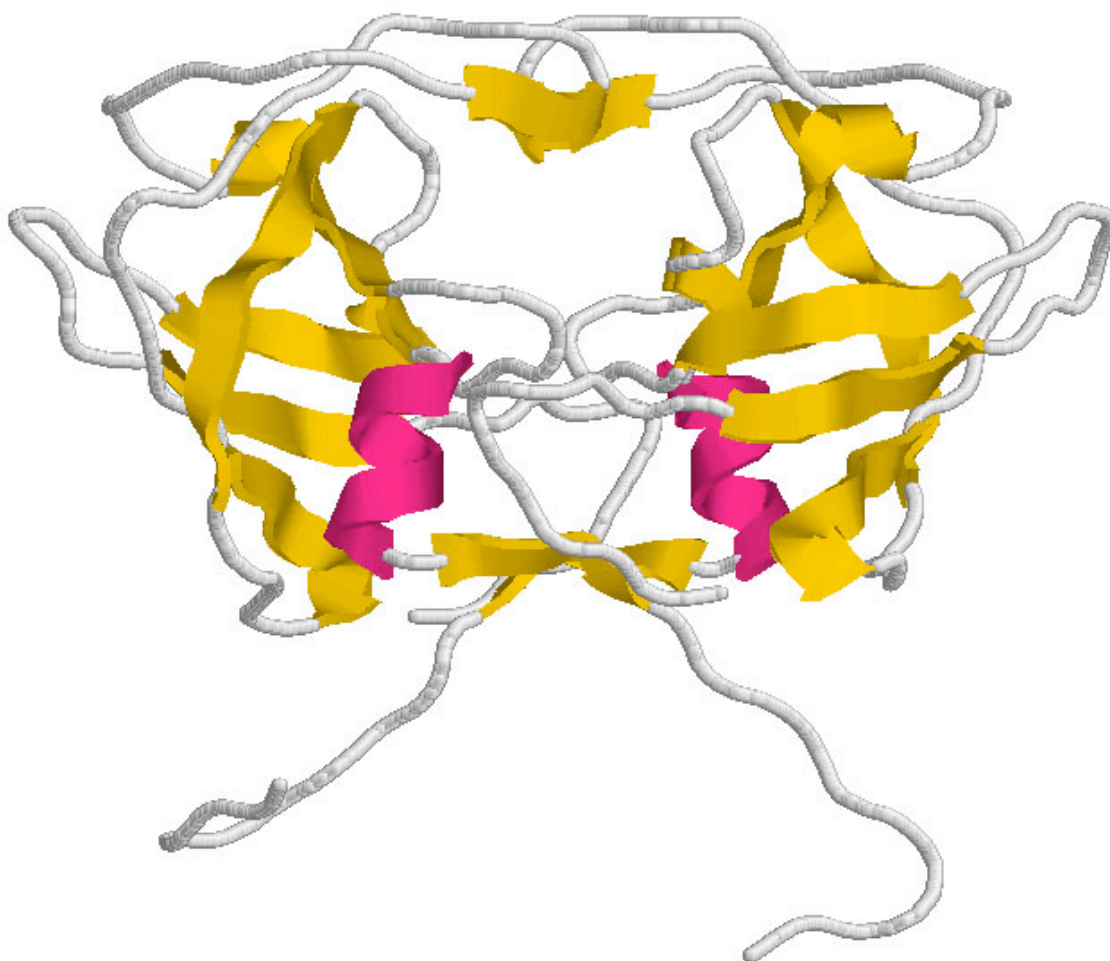


Figure 3.3. Theoretical model of full-length HTLV-I protease.

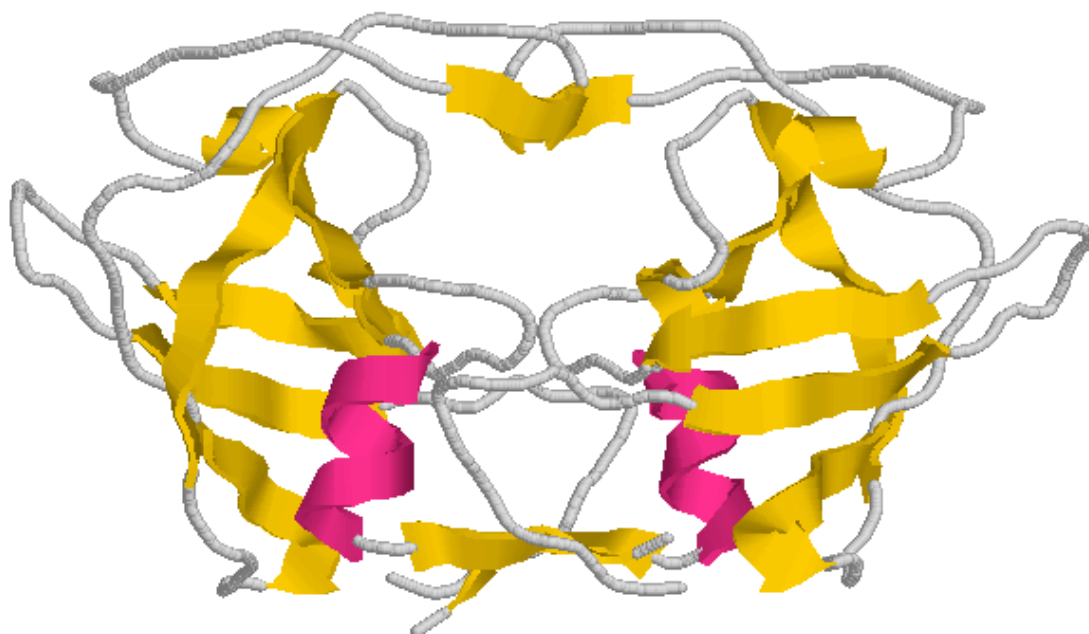


Figure 3.4. Theoretical model of C-terminal truncated HTLV-I protease.

3.3 Docking a native substrate into the active site

Most experimentally determined structures of proteases are prepared using a substrate or inhibitor in order to visualize binding. Substrate binding contacts in HTLV-I protease were demonstrated by docking the substrate TKVLVVQP in the putative active site of HTLV-I protease, which was modeled using the commercial software MOE (Chemical Computing Group, Montreal, Canada). The resulting model was refined by software analysis. Substrate design and docking work were performed by Kelly Joy Dennison. The final structure, with substrate, was published in the Protein Data Bank as file 1O0J, and is represented in Figure 3.5.

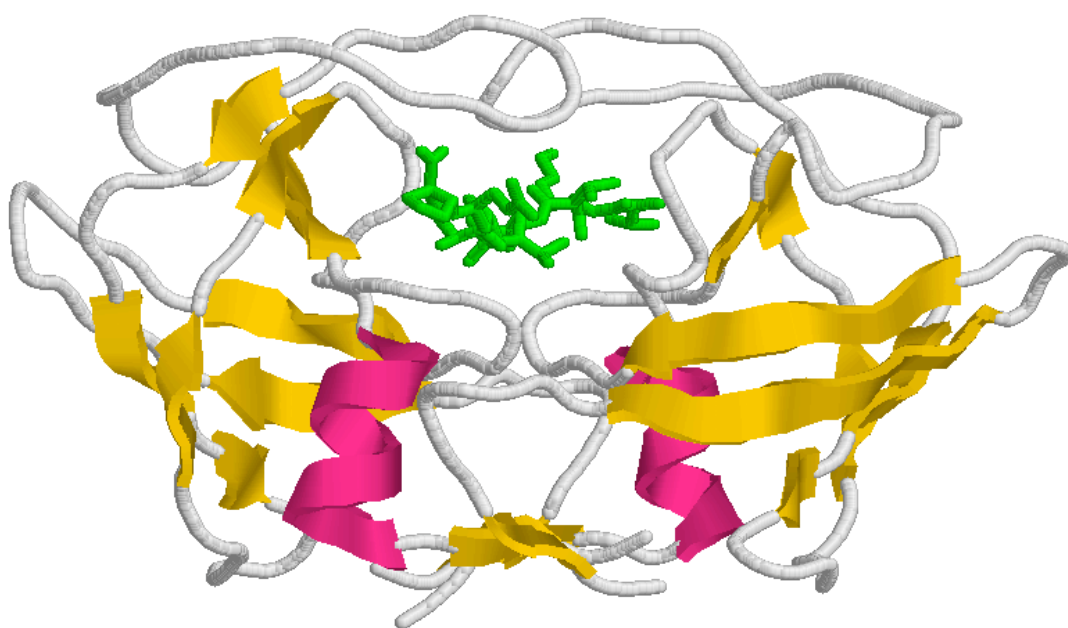


Figure 3.5. HTLV-I protease with docked CA/NC substrate. PDB file 1O0J.

3.4 The ten C-terminal residues

It is clear from the sequence alignment presented in Figure 3.1 that the ten C-terminal residues do not have a counterpart in the immune deficiency retroviral protease sequences. However, there is a corresponding C-terminal extension in BLV protease. Work on HTLV-I protease suggested that, while residues 121-125 could be deleted, residues 116-120 of the C-terminal extension were necessary for catalysis (Hayakawa et al., 1992). However, the authors did not perform an *in vitro* assay of purified truncated protease and substrate, instead relying on whole-cell protein analysis to detect cleavage product. It is possible that contaminants, including other cellular components or proteases, could have confounded the result. It is necessary to use purified protease and substrate to eliminate the possibility of other reactions that might give similar results.

Previous work showed that the ten C-terminal residues of BLV protease are not necessary for catalysis (Precigoux et al., 1993). Further, the generated model structure indicates that the C-terminal extension is not near enough to the active site or putative binding sites to significantly influence catalytic activity. In order to determine the effect of the C-terminal extension, a mutant was prepared that expressed only residues 1-115 of the protease.

3.5 Activity of full-length and truncated HTLV-I protease

A plasmid expressing L40I protease was prepared in previous work (Ha, 2001). An expression plasmid carrying truncated protease, also called “wo10” or “L40I-wo10”, was prepared by amplifying only the DNA encoding residues 1-115

by PCR. The PCR fragment was engineered to have NheI and XhoI restriction sites and a stop codon at the 3' end. This fragment was digested and ligated to similarly digested pET-28b plasmid. The sequence of the resulting plasmid was verified, and the plasmid was transformed into *E. coli* BL21(DE3) for expression.

Activity of full-length L40I protease and truncated L40I-wo10 protease was assayed against the previously described fluorogenic substrate (Figure 2.7). The enzyme concentration was standardized to 50 nM. Cleavage of varying substrate concentrations was monitored by observing substrate fluorescence at 490 nm following excitation at 340 nm. Eadie-Hosftee plots (Dowd and Riggs, 1965) of the resulting data are shown in Figures 3.6 (L40I) and 3.7 (L40I-wo10).

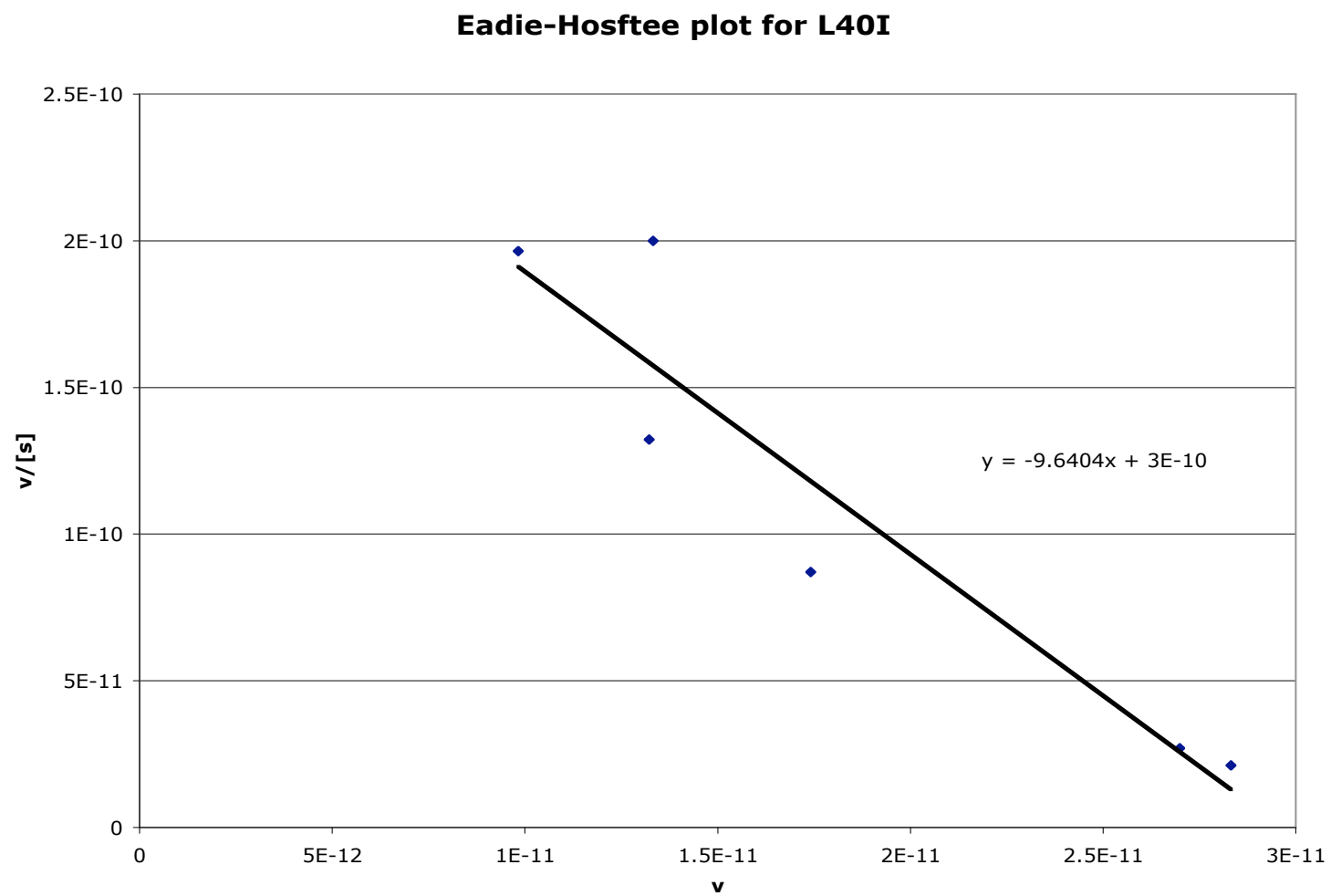


Figure 3.6. Eadie-Hosftree plot of L40I kinetic data.

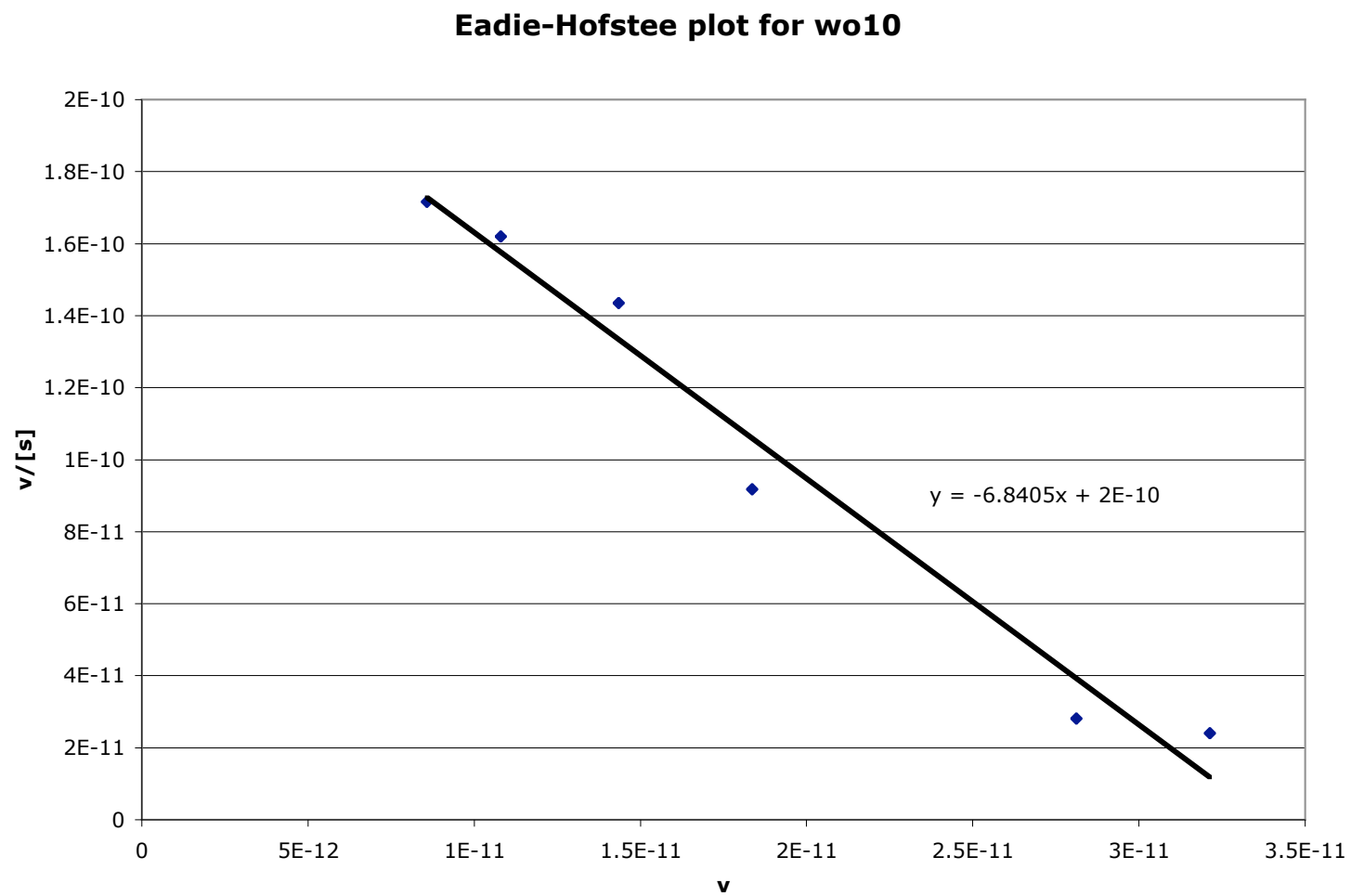


Figure 3.7. Eadie-Hofstee plot of L40I-wo10 kinetic data.

The analysis of the resulting data is summarized in Table 3.2. The catalytic efficiency, k_{cat}/K_M , was determined to be $593 \pm 48 \text{ M}^{-1} \text{ s}^{-1}$ for L40I, and $677 \pm 54 \text{ M}^{-1} \text{ s}^{-1}$ for L40I-wo10. While there is a marginal increase in catalytic efficiency of the truncated L40I-wo10 protease, the margins of error overlap, indicating that the difference may be due to experimental variability. It is reasonable to conclude from this data that the ten C-terminal residues play little role in catalytic activity.

Protease	$k_{\text{cat}} (\text{s}^{-1})$	$K_M (\text{microM})$	$k_{\text{cat}}/K_M (\text{M}^{-1} \text{s}^{-1})$
L40I	$.00572 \pm .000674$	9.64 ± 1.73	593 ± 48
L40I-wo10	$.00463 \pm .000238$	$6.84 \pm .575$	677 ± 54

Table 3.2. Summary of kinetic data for L40I and L40I-wo10 protease.

3.6 Discussion

The present work contrasts with previous work suggesting that residues 116-120 of HTLV-I protease are necessary for catalysis; however, the authors of this previous work did not perform an *in vitro* assay of purified truncated protease and substrate, instead relying on whole-cell protein analysis to detect cleavage product. It is clear from the theoretical model that the C-terminal tail is distant from the catalytic site and is probably not involved in catalysis. Kinetic data show that the truncated protease may have a slightly higher catalytic efficiency on the tested substrate, but the increase in efficiency does not exceed the error bounds. Therefore, the ten C-terminal residues have no significant effect on *in vitro* catalytic efficiency.

3.7 References

- Dowd, J. E., and Riggs, D. S. (1965). A Comparison of Estimates of Michaelis-Menten Kinetic Constants from Various Linear Transformations. *J Biol Chem* **240**, 863-9.
- Fiser, A., Do, R. K. G., and Sali, A. (2000). Modeling of loops in protein structures. *Protein Science* **9**(9), 1753-1773.
- Ha, J. J. (2001). Georgia Institute of Technology, Atlanta.
- Hayakawa, T., Misumi, Y., Kobayashi, M., Yamamoto, Y., and Fujisawa, Y. (1992). Requirement of N-terminal and C-terminal regions for enzymatic activity of Human T-Cell Leukemia Virus Type-I protease. *European Journal of Biochemistry* **206**(3), 919-925.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure* **29**, 291-325.
- Precigoux, G., Geoffre, S., Leonard, R., Llido, S., Dautant, A., Destaintot, B. L., Picard, P., Menard, A., Guillemain, B., and Hospital, M. (1993). Modeling, synthesis and biological activity of a BLV proteinase, made of (only) 116 amino acids. *Febs Letters* **326**(1-3), 237-240.
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints **234**(3), 779-815.
- Shuker, S. B., Mariani, V. L., Herger, B. E., and Dennison, K. J. (2003). Understanding HTLV-I Protease. *Chem Biol* **10**(5), 373-80.
- Tyndall, J. D., Reid, R. C., Tyssen, D. P., Jardine, D. K., Todd, B., Passmore, M., March, D. R., Pattenden, L. K., Bergman, D. A., Alewood, D., Hu, S. H., Alewood, P. F., Birch, C. J., Martin, J. L., and Fairlie, D. P. (2002). Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macrocyclic inhibitors of HIV-1 protease **43**, 3495-3504.
- Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W., and Weber, I. T. (1998). Structural basis for specificity of retroviral proteases **37**, 4518-4526.

CHAPTER 4

MODIFYING THE SUBSTRATE SPECIFICITY OF HTLV-I PROTEASE

4.1 Identifying the substrate binding sites of HTLV-I protease

The active form of HTLV-I protease is composed of two identical monomers, each of which contains one catalytic aspartic acid residue. The monomers bind to form the active site, which is composed of several pockets that recognize certain amino acids in the substrate. Previous studies have been undertaken to identify which amino acids are preferred in certain positions in the substrate (Mariani and Shuker, 2003; Tözsér et al., 2000). However, little work has been undertaken to identify the residues in HTLV-I protease that are involved in substrate recognition. Tözsér et al. prepared a theoretical model of HTLV-I protease (Tözsér et al., 2000) and speculated on which residues at each binding site might be important in recognizing substrate. However, no attempt has been made to validate which residues in the protease are important in substrate binding.

4.2 Determining residues involved in binding by sequence alignment

The similarity between retroviral proteases makes it possible to align the sequences to compare structural features. For example, HIV-1 protease, the most studied member of the retroviral protease family, can be used as a template in the study of similar proteases. The similarity between HIV-1 and RSV

proteases was utilized to prepare a mutant form of RSV protease that was able to process HIV-1 protease substrate (Wu et al., 1998). A theoretical model of HTLV-I protease has been previously reported (Tözsér et al., 2000). This model was developed by comparing the sequence of the published RSV protease S9 mutant (Wu et al., 1998), generating a model of HTLV-I protease using this structure, and modeling the substrate KVL/VVQP in the active site. However, this model was not deposited in the Protein Data Bank, and it is not readily available for examination.

In order to compare substrate specificity of HTLV-I protease with that of HIV-1 protease, a published alignment was compared to the HIV-1 and RSV protease alignment used by Wu et al. to prepare the mutant RSV protease. The alignment that was used to prepare the PDB model 1O0J was also compared to the HIV-RSV alignment.

The multiple alignment of HIV-1 protease, RSV protease, and the two HTLV-I protease alignments is shown in Figure 4.1. Residues highlighted in the RSV protease line were mutated to the residues at corresponding positions in HIV-1 protease. Based on this alignment, corresponding residues in HTLV-I protease were identified. The residues in HTLV-I protease that align with residues identified to be important for specificity in RSV protease were mutated to corresponding residues in HIV-1 protease.

```

HTLV-IS   PVIPLDPARPVIKAQVDTQT---S-HP--KTIEALLDTGADMTVLPIALFSSN-----TP-LK--N-TSVL
RSVwt     LAMTMEHKDRPLVRVILTNTGSHPVKQ-RSVYITALLDSGADITIISEEDWPT-DW-P-V-MEAANPOIHGI
HIV-1     PQITLW--KRPLVTIKIG-----GQLKEALLDTGADTVIEE--MS---LPGRW-KPK---MIGGI
HTLV-IT   PVIPLDPARPVIKAQVDTQT-----S-HPKTIEALLDTGADMTVLPIALFSS-NT-P-L-KN-T--SVLGA

HTLV-IS   GAGGQTQDHFK-LTSLPVLIRLPFRTTPIVLTSCLVDTKNNWAIIGRDALQOCQGVLYLPEAKGPPVIL
RSVwt     GGGIPMRKSRD-MIELGVINRDGSLERPLLLFPAVAMVRGS--ILGRDCLQGLGLRLTNL-----
HIV-1     GGFIKV-RQY-DQILIEIC-----GH-KAIGTVLVGPTPVN--IIGRNLLTQIGCTLNF-----
HTLV-IT   GGQTQDHFK-LTSLPVLIRLPFRTTPIVLTSCLVDTKNNWA--IIGRDALQOCQGVLYLPEAKGPPVIL

```

Figure 4.1. Multiple alignment of RSV protease (RSVwt), HTLV-I protease (Shuker alignment: S), HTLV-I protease (Tözsér alignment: T), HIV-1 protease. Residues in yellow in RSV protease were mutated by Wu et al. to the corresponding residues in HIV-1 protease.

4.3 Preparation of mutants

Mutations were performed by overlap extension (Ho et al., 1989; Horton et al., 1989). This two-step PCR-based protocol allows the preparation of single- or multiple-mutant DNA inserts that are then digested and ligated into an expression plasmid. This procedure requires four primers: a 5' primer matching the target gene, a 3' primer matching the target gene, and two complementary primers carrying the mutation flanked by 12-15 bases matching the gene sequence on either side. The 5' and 3' primers typically have restriction sites to facilitate insertion into a plasmid. In the first reaction, two separate amplifications are performed to prepare two fragments: the 5' half of the gene with a 3' mutation, and the 3' half of the gene with a 5' mutation. The mutated ends of the halves correspond such that when the two halves are amplified in the presence of 5' and 3' end primers, the mutated sections anneal, and the complete gene with mutation is amplified. Overlap extension is summarized in Figure 4.2. This process was repeated multiple times with different primers to produce genes with multiple mutations. The mutated gene was digested with restriction enzymes and ligated to a similarly digested plasmid, producing a construct to express the mutant gene.

Preliminary work in which percent activity was determined was performed by incubating protease samples at 37°C with HIV-1 substrate H-2930 (Molecular Probes), a FRET substrate that is excited at 340 nm and emits at 490 nm upon cleavage. Percent activity was calculated by dividing the initial rate observed for

each mutant by the initial rate observed for HIV-1 protease. Percent activity data from this study are presented in Table 4.1.

Original sequence:

-----AAA-----

Reaction 1, Step 1 PCR:

-F1->

-----AAA-----

<--CCC-Mut1

Reaction 2, Step 1 PCR:

Mut2-GGG-->

-----AAA-----

<-R1-

Products of step 1 are purified to remove template.

Step 2 PCR, part 1 – mutated halves overlap and anneal:

-----CCC---

--GGG-----

Fill-in by Vent DNA polymerase yields:

-----CCC-----

-----GGG-----

F1 and R1 are also present and selectively amplify the full-length sequence:

-F1->

-----CCC-----

-----GGG-----

<-R1-

Step 2 is purified, digested, and ligated to yield a construct.

Figure 4.2. Mutation by overlap extension. Fragments shown as arrows, e.g. -F1->, are primers.

<i>Protease variant</i>	<i>Percent activity</i>
HIV-1 protease	100%
HTLV-I protease, alternate Shuker	50%
HTLV-I protease, Shuker	30%
HTLV-I protease, Tözsér	5%
HTLV-I protease, wild-type	ND

Table 4.1. Preliminary kinetic data for specificity mutants. These data were obtained independently from the data used for Michaelis-Menten analysis.

4.4 Activity of HIV-1 substrate-specific mutant based on Tözsér alignment

The mutant form of HTLV-I protease based on Tözsér et al. alignment carries the following mutations: M37D, D65V, V92L, N96T, N97P, W98V, and A99N. These mutations are depicted in Figure 4.3(a). This mutant was incubated with a nitrophenylalanine-containing HIV-1 protease substrate. The rate of catalysis was determined by monitoring the absorbance at 290 nm and correcting for background variation.

4.5 Activity of HIV-1 substrate-specific mutant based on Shuker alignment

The mutant form of HTLV-I protease based on Shuker alignment carries the following mutations: M37D, T63V, C90L, K95P, and N96V. These mutations are depicted in Figure 4.3(b). This mutant was incubated with a nitrophenylalanine-containing HIV-1 protease substrate. The rate of catalysis was determined by monitoring the absorbance at 290 nm and correcting for background variation.

(a)

D

PVIPLDPARRPVIKAQVDQTQSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
 DHFKLTSLPVLIRLPFRTPIVLTSLVDTKNNWAIIGRDALQCCQGVLYLPEAKGPPVIL
 V L TPVN

(b)

D V

PVIPLDPARRPVIKAQVDQTQSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
 DHFKLTSLPVLIRLPFRTPIVLTSLVDTKNNWAIIGRDALQCCQGVLYLPEAKGPPVIL
 L PV

(c)

D

PVIPLDPARRPVIKAQVDQTQSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
 DHFKLTSLPVLIRLPFRTPIVLTSLVDTKNNWAIIGRDALQCCQGVLYLPEAKGPPVIL
 V L PV

Figure 4.3. Mutations made based on different alignments. Target residues are bold with mutations above or below the parent sequence. (a) Mutations based on Tözsér alignment; (b) mutations based on Shuker alignment; (c) mutations based on best fit regions in both alignments.

4.6 Activity of HIV-1 substrate-specific mutant based on alternate Shuker alignment

Review of the multiple alignment shows a two-residue variance between the Shuker and Tözsér alignments around Gly58. This difference appears to be caused by a variance in gaps between the two alignments. However, this alignment affects whether Thr63 or Asp65 should be aligned with the corresponding residue in RSV protease. This variance in alignment may alter the specificity of the mutant forms. In order to assess which alignment is more valid, a third mutant was prepared based on the Shuker alignment, but incorporating D65V rather than T63V. The revised Shuker alignment is shown in Figure 4.4.

```

HTLV-ALT  PVIPLDPARRPVIIKAQVDTQT---S-HP--KTIEALLDTGADMTVLPIALFSSN-----TP-LKN
HIV-1      PQITLW--KRPLVTIKIG-----GQLKEALLDTGADDTVIEE--MS---LPGRW-KPK-

HTLV-ALT  -TSVLGAGGQTQDHFK-LTSLPVLIR--LPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
HIV-1      --MIGGIGGFIKV-RQY-DQIIIEIC-----GH-KAIGTVLVGPTPVN--IIGRNLLTQIGCTLNF-----

```

Figure 4.4. Revised alignment of HTLV-I protease and HIV-1 protease. The yellow highlight indicates the similar region that was targeted in adjusting the alignments about Gly58 in the HTLV-I sequence. The green highlight shows Thr63 and Asp65 in the HTLV-I sequence and the corresponding valine in the HIV-1 sequence.

The mutant form of HTLV-I protease based on the alternate Shuker alignment carries the following mutations: M37D, D65V, C90L, K95P, and N96V. These mutations are depicted in Figure 4.3(c). This mutant was incubated with a nitrophenylalanine-containing HIV-1 protease substrate. The rate of catalysis was determined by monitoring the absorbance at 290 nm and correcting for background variation.

4.7 Summary and Discussion

The kinetic data from all HIV-1 substrate specific HTLV-I protease mutants, wild-type HIV-1 protease, and wild-type HTLV-I protease are summarized in Table 4.2.

<i>Mutant</i>	k_{cat} (s^{-1})	K_M (microM)	k_{cat}/K_M ($\text{mM}^{-1} \text{s}^{-1}$)
HIV-1 protease (Toth and Marshall, 1990)	0.29 ± 0.03	37 ± 8	7.8 ± 0.3
HTLV-I protease: combined mutant (ALT)	$3.6 \times 10^{-3} \pm 2.9 \times 10^{-4}$	64 ± 10	0.056 ± 0.0045
HTLV-I protease: Shuker mutant (S)	$1.6 \times 10^{-3} \pm 1.3 \times 10^{-4}$	52 ± 8	0.031 ± 0.0025
HTLV-I protease: Tözsér mutant (T)	$4.5 \times 10^{-4} \pm 4.4 \times 10^{-5}$	39 ± 8	0.012 ± 0.0011
HTLV-I protease: wild type	Not detected	Not detected	Not detected

Table 4.2. Summary of kinetic data from variants of HTLV-I protease and HIV-1 protease on HIV-1 protease substrate.

In terms of k_{cat}/K_M , or overall catalytic efficiency, it is clear that the specificity mutant based on the alternate Shuker alignment has the highest

efficiency at $0.056 \text{ mM}^{-1} \text{ s}^{-1}$, followed by the mutant based on the Shuker structure alignment at $0.031 \text{ mM}^{-1} \text{ s}^{-1}$, followed by the mutant based on the Tözsér alignment at $0.012 \text{ mM}^{-1} \text{ s}^{-1}$. The increase in k_{cat} is in line with preliminary percent activity data derived from a fluorescence experiment. However, the binding constant, K_M , increases with efficiency, indicating less specific binding despite increasing turnover, k_{cat} . The apparent decrease in binding constant could indicate a general decrease in specificity, rather than specificity to HIV-1 protease substrate. Nonetheless, this study indicates that certain residues are significant in binding substrate for catalysis, and that these residues can be identified through sequence alignment. The sequence alignments prepared in the present work better indicate which residues are significant in binding. These results further indicate that the structural model based on the Shuker alignment is a better representation of HTLV-I protease than the structural model based on the Tözsér alignment. With additional knowledge of the residues that are involved in specificity, we will be able to develop inhibitors that have contacts matching the pattern in the active site of HTLV-I protease.

4.8 References

- Ho, S., Hunt, H., Horton, R., Pullen, J., and Pease, L. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1), 51-59.
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**(1), 61-8.
- Mariani, V. L., and Shuker, S. (2003). Identification of the RT-RH/IN cleavage site of HTLV-I. *Biochem. Biophys. Res. Comm.* **300**, 268-270.
- Toth, M. V., and Marshall, G. R. (1990). A simple, continuous fluorometric assay for HIV protease. *Int J Pept Protein Res* **36**(6), 544-50.
- Tözsér, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W., and Weber, I. T. (2000). Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *European Journal of Biochemistry* **267**(20), 6287-6295.
- Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W., and Weber, I. T. (1998). Structural basis for specificity of retroviral proteases **37**, 4518-4526.

CHAPTER 5

ALANINE SCAN OF HTLV-I PROTEASE

5.1 Purpose of alanine scan

Knowledge of which residues in the sequence of HTLV-I protease contribute to substrate binding is essential to understanding the structure and function of this protease. Determining which steric and electrostatic contacts are needed for catalysis can also help in the design of protease inhibitors; for example, identifying positive, negative, and hydrophobic contacts at each protease subsite allows rational selection of residues in peptide-based inhibitors. An alanine scan is an important diagnostic tool for identifying residues involved in catalysis (Cunningham and Wells, 1989). In this technique, a single residue is mutated to alanine, and the kinetic parameters of the resulting mutant are evaluated. Alanine is chosen because it has the smallest side chain other than glycine, but unlike glycine, it is chiral and maintains the conformation of the backbone. Several residues believed to be important in binding are mutated to determine the contribution of these residues to the catalytic activity of the protease. In the case of HTLV-I protease, several residues were hypothesized to be important in binding based on the theoretical structure. The following residues were studied in the present work: Arg10, Leu30, Thr94, Lys95, Asn96, Trp98, and Ile100.

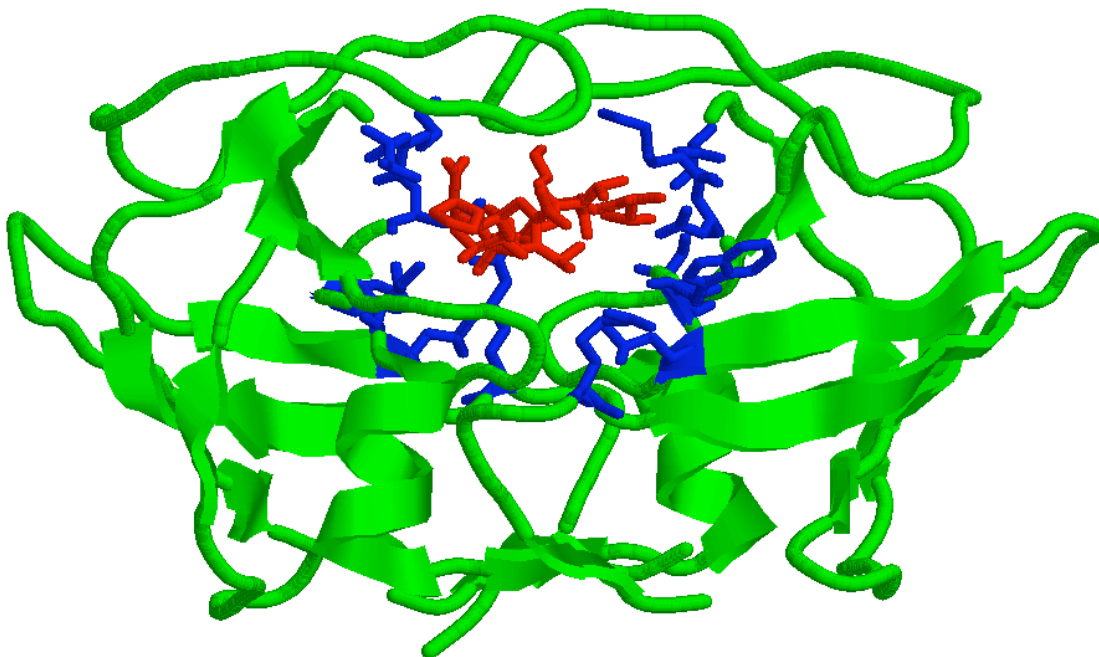


Figure 5.1. Positions of residues selected for alanine scan. Green and blue structure is protease; blue residues were mutated to alanine. Red structure is the CA/NC substrate.

5.2 Preparation of alanine mutants

Single alanine mutants of L40I HTLV-I protease were prepared by overlap extension (Ho et al., 1989; Horton et al., 1989). This process was summarized in Figure 4.2. The following residues were mutated to alanine: Arg10, Leu30, Thr94, Lys95, Asn96, Trp98, and Ile100. The position of these residues in the theoretical structure is shown in Figure 5.1. Each construct was transformed into *E. coli* BL21(DE3) cells, and the resulting mutant protease was expressed and purified as described. Kinetic studies were performed using an analog of the matrix-capsid substrate, APQVL(Nph)VMHPL, where Nph is 4-nitrophenylalanine. The substrate is shown in Figure 5.2. The activity of L40I protease on this substrate was also tested as a reference.

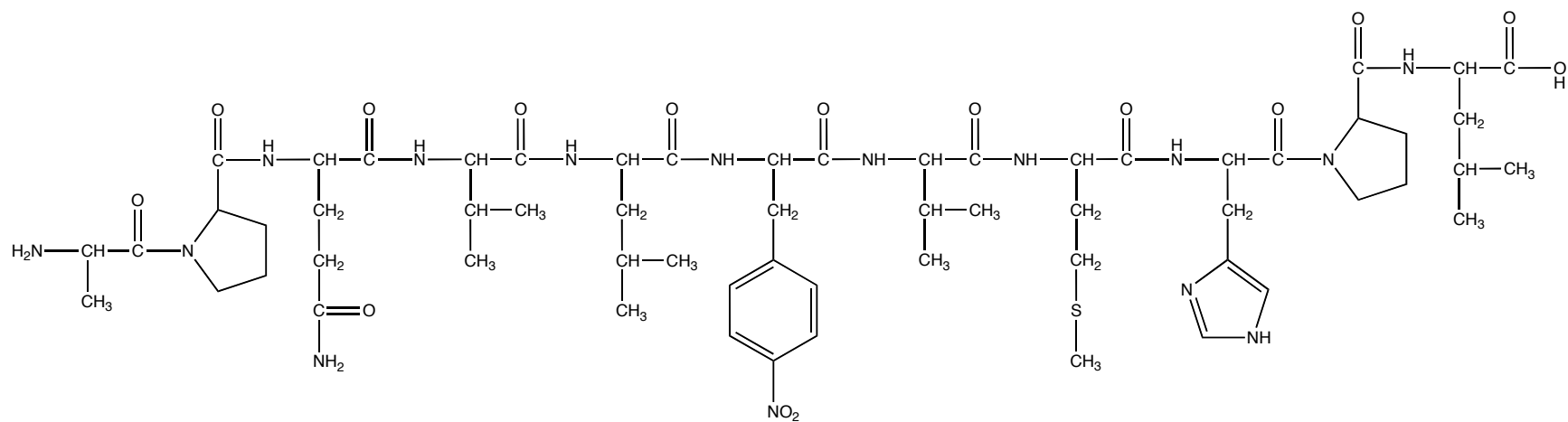


Figure 5.2. Substrate used in alanine scan kinetic study.

5.3 Results and discussion

Kinetic data for all alanine mutants as well as L40I are summarized in Table 5.1. Inferences about structural significance are based on the theoretical model, PDB file 1O0J, presented in Chapter 3. The reference L40I protease had k_{cat}/K_M of $63.5 \text{ M}^{-1} \text{ s}^{-1}$.

Mutant	$k_{\text{cat}} (\text{s}^{-1})$	$K_M (\text{microM})$	$k_{\text{cat}}/K_M (\text{M}^{-1} \text{s}^{-1})$
L40I	2.74×10^{-3}	43.2	63.5
R10A	7.18×10^{-3}	61.6	117
L30A	5.72×10^{-3}	71.6	79.9
T94A	3.01×10^{-3}	111	27.2
K95A	9.36×10^{-3}	73.6	127
N96A	6.03×10^{-3}	70.3	85.7
W98A	2.96×10^{-3}	8.33	355
I100A	4.43×10^{-3}	43.3	102

Table 5.1. Summary of kinetic data of alanine mutants.

Mutating Arg10 to alanine increased k_{cat}/K_M to $117 \text{ M}^{-1} \text{ s}^{-1}$, which is due to a significant increase in k_{cat} and a small increase in K_M . Arg10 appears to lie in the entrance to the active site channel and may provide a steric barrier and electrostatic contact to the substrate; replacing Arg10 with alanine appears to remove the steric barrier, allowing faster turnover. Leu30 lies very near the catalytic aspartic acid, Asp32, and probably contributes to the overall hydrophobicity of S1 and S1'. The position and size of Leu30 probably plays a significant role in restricting the type of residue that can bind in S1 and S1'. However, mutating Leu30 to alanine does not significantly change the efficiency of the enzyme; the k_{cat}/K_M of this mutant is $79.9 \text{ M}^{-1} \text{ s}^{-1}$. Other nearby residues,

notably the adjacent Leu31, may play a stronger role in selectivity in the active site, or the position of surrounding residues such as Leu31 may change to yield a similar effect.

Residues 94-100 lie on a loop on the interior of the protease. As noted in Chapter 4, residues in this loop play a significant role in selectivity. For example, mutation of Thr94 decreases the catalytic efficiency of the protease to k_{cat}/K_M of $27.2 \text{ M}^{-1} \text{ s}^{-1}$. Most of the decrease in efficiency is due to an increase in K_M , indicating weaker binding of the substrate. Thr94 may play a role in stabilizing the protease structure in the active site region. In the theoretical model, the hydroxyl group on the side chain of Thr94 lies near the backbone carbonyl oxygen of the P1 Leu. In the initial step of catalysis, the negatively charged oxygen of the catalytic aspartic acid attacks the carbonyl carbon. It is possible that the hydroxyl side chain of Thr94 stabilizes the partial negative charge on the carbonyl oxygen, facilitating the first step of cleavage. Figure 5.3 shows this process.

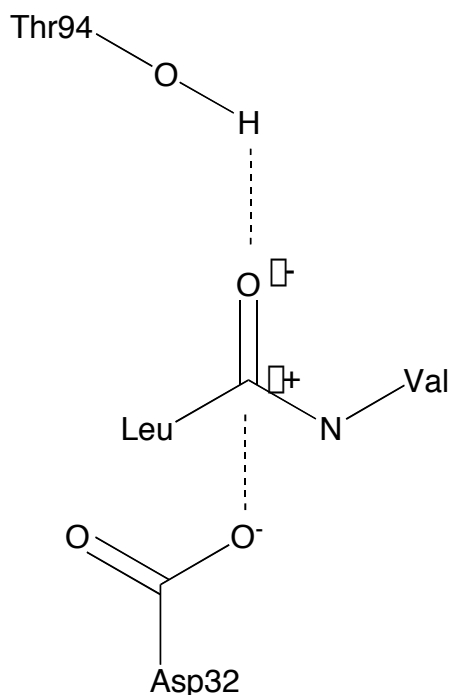


Figure 5.3. Stabilization of the backbone carbonyl oxygen by Thr94 prior to cleavage between Leu and Val.

Lys95 is a bulky residue with a charged amino group. Mutating this residue leads to a substantial increase in turnover, with overall k_{cat}/K_M of $127 \text{ M}^{-1} \text{ s}^{-1}$. The lysine side chain must move to accommodate substrate; therefore, removing the side chain results in faster binding. However, binding is slightly less specific as measured in terms of K_M . Mutation of Asn96 to alanine has little effect on catalytic activity, suggesting that this residue plays little role in structure or catalysis. Mutation of the bulky residue Trp98 to alanine dramatically improves the efficiency to k_{cat}/K_M of $355 \text{ M}^{-1} \text{ s}^{-1}$. Most of the change is in K_M , indicating tight binding of the substrate. However, this is likely due to the chosen substrate, which has nitrophenylalanine at P1'. The nitrophenylalanine likely interacts with the aromatic ring of Trp98, and removal of Trp98 apparently eliminates this steric constraint. Ile100 appears to contribute to the hydrophobic surface around S1

and S1'. The contribution of Ile100 is small; mutating Ile100 to alanine increases efficiency to k_{cat}/K_M of $102 \text{ M}^{-1} \text{ s}^{-1}$, with all of the effect due to an increase in k_{cat} . If Ile100 were present, the side chain would probably bend slightly in the process of binding; if this conformational change were not needed, as is the case in the alanine mutant, turnover increases slightly as observed.

5.4 References

- Cunningham, B. C., and Wells, J. A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**(4908), 1081-5.
- Ho, S., Hunt, H., Horton, R., Pullen, J., and Pease, L. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1), 51-59.
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**(1), 61-8.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

HTLV-I was discovered in the early 1980's in Japanese and American patients with cutaneous T-cell lymphoma (Gallo, 1996; Gessain, 1996). The protease of this retrovirus has been isolated and characterized (Nam and Hatanaka, 1986). However, no effective treatments exist that target the virus. Although HTLV-I relies on reverse transcriptase to transcribe its genome, reverse transcriptase inhibitors designed for HIV-1 infection show little therapeutic benefit. Protease inhibitors have been reported (Shuker et al., 2003), but there are no clinically approved protease inhibitors for the treatment of HTLV-I-related illness.

In the present work, information about the structure of HTLV-I protease is determined. A new sequence alignment with HIV-1 protease is described, and a theoretical model based on this alignment is presented, showing the structure of the protease with bound natural substrate. It was also determined that the ten C-terminal residues do not contribute to catalytic activity.

Sequence alignment among several retroviral proteases made it possible to identify residues that are important in specificity. Comparison of the sequences of RSV, HIV-1, and HTLV-I proteases allowed identification of several residues that are important in binding substrate. Mutation of residues in HTLV-I protease to the corresponding residues in HIV-1 protease yielded a form of

HTLV-I protease that could cleave HIV-1 protease substrate. This experiment also resulted in the selection of a different alignment of HIV-1 and HTLV-I proteases. Mutations prepared based on this new alignment yielded a protease that processed HIV-1 protease substrate more efficiently, indicating that this alignment corresponded better with HIV-1 protease.

An alanine scan was performed to elucidate the function of several residues. These residues were selected based on their position in the theoretical structure. Based on these results, we determined that Thr94 is significant in catalysis, and several other tested residues play some role in catalysis. In contrast, Asn96 appears to play little role in catalysis. No single residue in the test set was shown to be absolutely necessary for catalysis.

Future work on HTLV-I protease will focus on two key projects. First, the structure of the protease should be experimentally determined. This will confirm the studies described here, and give substantial information relating to the structure of the protease and how it binds substrate and inhibitor. Second, peptide- and non-peptide-based inhibitors of the protease should be developed. A collaboration exists between our group and the CDC, and it is hoped that this work will lead to protease inhibitors that are effective against infected cells in culture. This work may lead to the first clinically viable HTLV-I protease specific inhibitor.

References

- Gallo, R. C., Thomson, M. M. (1996). Introduction. *In* "Human T-cell Lymphotropic Virus Type I" (P. Hollsberg, Hafler, D. A., Ed.), pp. 325. Wiley and Sons, Chichester.
- Gessain, A. (1996). Epidemiology of HTLV-I and associated diseases. *In* "Human T-Cell Lymphotropic Virus Type I" (P. Hollsberg, and D. A. Hafler, Eds.), pp. 31-64. Wiley, Chichester, UK.
- Nam, S. H., and Hatanaka, M. (1986). Identification of a protease gene of human T-cell leukemia virus type I (HTLV-I) and its structural comparison. *Biochem Biophys Res Commun* **139**(1), 129-35.
- Shuker, S. B., Mariani, V. L., Herger, B. E., and Dennison, K. J. (2003). Understanding HTLV-I Protease. *Chem Biol* **10**(5), 373-80.

CHAPTER 7

MATERIALS AND METHODS

Preparation of plasmids expressing HTLV-I protease without 10 C-terminal residues. The gene for HTLV-I protease was cloned from the K30 strain of HTLV-I, which was prepared by Thomas Kindt (Mulligan and Berg, 1981; Zhao et al., 1995). A plasmid expressing HTLV-I protease with L40I mutation, called pPR203, was prepared in previous work (Ha, 2001). For the present work, a plasmid expressing protease without the last ten amino acids was prepared. This was accomplished by PCR amplification of pPR203 using primers to clone residues 1-115 with NheI restriction site at the 5' end and XhoI restriction site at the 3' end. PCR amplification was carried out using Vent DNA polymerase, Vent polymerase buffer and dNTP mix (NEB), and a temperature profile of 95°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute, repeated 30 times. The PCR-prepared insert was digested with NheI and XhoI (NEB) under buffer conditions recommended by New England Biolabs. The plasmid pET-28b was similarly digested. The digested insert was ligated to digested pET-28b using T4 DNA ligase (NEB), but using NEBuffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9) plus fresh DTT and ATP solutions to match the concentration of DTT and ATP in NEB T4 buffer (50 mM Tris-HCl, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 25 ug/mL BSA, pH 7.5). The ligation was carried out at 20°C for 3 hours. The ligated plasmid was transformed as noted below, and the sequence was verified by the Sanger dideoxy method at the Georgia Tech

Biology Core Facility.

Preparation of mutant forms of HTLV-I protease. For specificity and alanine scan mutants, mutations were performed by overlap extension (Ho et al., 1989; Horton et al., 1989). For alanine scan mutants, only one overlap extension was performed, while for specificity mutants, the process was repeated until all targeted residues were changed. PCR primers encoding 5' (with NheI site) and 3' ends (with XhoI site) and mutations were synthesized by Integrated DNA Technologies. PCR amplification was performed in two steps as described in literature, with a temperature profile of 95°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute for 30 cycles in the second overlap step. The mutant insert and pET-28b were digested and ligated as above. The ligated plasmid was transformed as noted below, and the sequence was verified by the Sanger dideoxy method at the Georgia Tech Biology Core Facility.

Transformation of cells with HTLV-I protease constructs. Competent *E. coli* cells were thawed on ice. A small volume, 1-5 uL, of plasmid solution was added to the cells. The cells were allowed to stand on ice for 30 minutes. The cells were heat shocked at 42°C for 45 seconds. The cells were chilled on ice for 1 minute. The cells were transferred to cool antibiotic-free LB or SOC broth, at least 5 times the volume of cells. This culture was grown at 37°C for 1 hour, and 100 uL of the culture was streaked on an LB agar plate with antibiotic corresponding to the plasmid. The agar plate was incubated for 37°C overnight. Alternatively, 100 uL of the antibiotic-free culture was transferred to 5 mL LB broth with the appropriate antibiotic. This liquid culture was then incubated overnight at 37°C.

Preparation of frozen stocks on transformed cells. Transformed cells were allowed to grow in antibiotic-labeled liquid media and incubated until the culture was turbid, at least 8 hours. One mL of culture was transferred to a labeled glass vial. One mL of 7% v/v DMSO in dH₂O was added to the culture. This mixture was frozen at -70°C immediately. Frozen stocks were reconstituted by scraping the frozen stock and transferring the ice to 5 mL of antibiotic-labeled LB broth and incubating overnight.

Expression of HTLV-I protease. All forms of HTLV-I protease were expressed in the same manner. *E. coli* BL21(DE3) cells transformed with an HTLV-I expression plasmid were grown in 5 mL LB broth at 37°C overnight. Between 1 and 5 mL of the overnight culture was transferred to a larger volume, and the culture was grown at 37°C to an OD₆₀₀ of 0.6. The culture was induced to express protease by the addition of IPTG to a final concentration between 0.5-1 mM. The culture was then grown an additional 2 to 3 hours. The cells were harvested by centrifugation at 10,000g for 10 minutes. The pellet was resuspended in Buffer A (20 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole), sonicated, and centrifuged at 10,000g for 10 minutes. The pellet was resuspended in Buffer B (20 mM Tris-HCl, 500 mM NaCl, 6 M urea, 5 mM imidazole), sonicated, and allowed to stand on ice for at least 30 minutes, but not more than 2 hours. This suspension was then centrifuged at maximum speed for 30 minutes. The supernatant was collected for purification, and the pellet was discarded.

Purification of HTLV-I protease. All forms of HTLV-I protease were purified by the same method. A column was packed with an amount of His-Bind nickel affinity chromatography resin (Novagen). The bed volume (hereafter, “volume”) was noted and used as a multiplier to determine the amount of each buffer to use in subsequent step. The column is equilibrated with 5 volumes ddH₂O, 5 volumes 50 mM NiSO₄, and 5 volumes Buffer B. The supernatant from the expression was then loaded. The supernatant was followed by 10 volumes of Buffer B and 6 volumes of Buffer C (20 mM Tris-HCl, 500 mM NaCl, 6 M urea, 20 mM imidazole). HTLV-I protease was eluted by addition of 5 volumes of Buffer D250 (20 mM Tris-HCl, 500 mM NaCl, 6 M urea, 250 mM imidazole) and collected in 1 mL fractions of eluent. The column was washed with 5 volumes of strip buffer. As needed, the column was cleaned by the method noted in the Novagen purification handbook. Column fractions were assayed for protein content and purity by SDS-PAGE gel. In some cases, presence of protein in the fractions was tested by Bio-Rad dye assay (10 uL dye + 30 uL water + 10 uL eluent fraction on Parafilm). Fractions that turned blue were retained. Purified HTLV-I protease was then dialyzed to 100 mM acetic acid at 4°C overnight. Following dialysis, protease concentration was determined by absorbance at 280 nm.

Assay of HTLV-I protease with and without the ten C-terminal residues. This work was performed by Victoria Mariani. Briefly, purified L40I (full-length) and L40I-wo10 (truncated) constructs were expressed and purified as described.

These constructs were assayed on varying concentrations of Dabcyl-PQVLPVMH-EDANS, a FRET substrate that absorbs at 340 nm and fluoresces at 490 nm on cleavage. Cleavage was monitored at 37°C in a PTI spectrofluorophotometer. Initial rates were determined from the fluorescence data, and kinetic parameters were determined by Eadie-Hofstee plot (Dowd and Riggs, 1965).

Assay of HTLV-I protease specificity mutants against HIV protease substrate.

The assay was performed with a fixed concentration of protease in 50 mM sodium acetate pH 5.2, 10% (v/v) glycerol, and varied concentrations of nitrophenylalanine-containing HIV-1 protease substrate H-5930 (Sigma) in DMSO. The rate of reaction was obtained by setting up reaction mixtures on a 96-well Costar microtiter plate, incubating at 37°C and monitoring the absorbance at 290 nm in a Molecular Devices SpectraMax Plus384 plate reader. Reaction mixtures with no protease added were also monitored as a control and to determine background effects. Initial rates were determined from the slopes of the absorbance curves, and kinetic parameters were calculated by Eadie-Hofstee plot.

Assay of single alanine mutants of HTLV-I protease against HTLV-I protease substrate.

The assay was performed with a fixed concentration of protease and varied concentrations of HTLV-I protease substrate APQVL(Nph)VHMPL (Synpep) in 50 mM sodium acetate, pH 5.2. Activity was measured by monitoring absorbance change at 290 nm. The resulting data were analyzed by

Eadie-Hofstee plot.

References

- Dowd, J. E., and Riggs, D. S. (1965). A Comparison of Estimates of Michaelis-Menten Kinetic Constants from Various Linear Transformations. *J Biol Chem* **240**, 863-9.
- Ha, J. J. (2001). Georgia Institute of Technology, Atlanta.
- Ho, S., Hunt, H., Horton, R., Pullen, J., and Pease, L. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1), 51-59.
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**(1), 61-8.
- Mulligan, R. C., and Berg, P. (1981). Selection for animal cells that express the Escherichia coli gene coding for xanthine-guanine phosphoribosyltransferase. *Proc Natl Acad Sci U S A* **78**(4), 2072-6.
- Zhao, T. M., Robinson, M. A., Bowers, F. S., and Kindt, T. J. (1995). Characterization of an infectious molecular clone of human T-cell leukemia virus type I. *J Virol* **69**(4), 2024-30.