Spring 2000

Characterization of Human T-Cell Leukemia Virus 1 Reverse Transcriptase

Pinky Gundayao Agbuya
Old Dominion University

CHARACTERIZATION OF HUMAN T-CELL LEUKEMIA VIRUS 1 REVERSE TRANSCRIPTASE

by

Pinky Gundayao Agbuya
B.S. May 1992, Philadelphia College of Pharmacy and Science

A Dissertation Submitted to the Faculties of Old Dominion University and Eastern Virginia Medical School in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

BIOMEDICAL SCIENCES

OLD DOMINION UNIVERSITY
EASTERN VIRGINIA MEDICAL SCHOOL
May 2000

Approved by:

_____
Laura K. Moen (Director)

_____
Mark S. Elliott (Member)

_____
Patricia A. Pleban (Member)

_____
Howard S. White (Member)

ABSTRACT

CHARACTERIZATION OF HUMAN T-CELL LEUKEMIA VIRUS 1 REVERSE TRANSCRIPTASE

Pinky Gundayao Agbuya
Old Dominion University and Eastern Virginia Medical School, 2000
Director: Dr. Laura K. Moen

Human T-cell leukemia virus 1 (HTLV-1) is a type C human retrovirus which has been the causative agent of Adult T-cell leukemia. Replication of the retrovirus requires a reverse transcriptase which converts the retroviral RNA into DNA which is later incorporated into the host's genome. Very little is known about the reverse transcriptase of HTLV-1. Researchers have attempted to purify HTLV-1 RT by isolating the enzyme from human cell lines. Because large amounts of protein could not be produced by this isolation method, the reverse transcriptase cannot be fully characterized. In this research, a recombinant protein expressed in *E. coli* was purified, which has an apparent molecular weight consistent with that from virion-derived reverse transcriptase. This recombinant protein was purified to approximately 90% homogeneity by a two-step chromatographic process on phosphocellulose and sepharose CL-6B.

The reverse transcriptase of HTLV-1 is synthesized as a Gag-Pro-Pol precursor protein and undergoes proteolytic processing during maturation. By using sequence comparisons from a number of retroviral *pol* genes as well as information about the location of the ribosomal frameshift, the location of the putative coding sequence for the enzyme has been identified. This information and PCR amplification was used to construct a clone, which spans a region of the *pro-pol* junction of HTLV-1, for overexpression in *E. coli*. The Pro-Pol protein was processed *in vitro* with a recombinant HTLV-1 protease to mimic the proteolytic processing of the Gag-Pro-Pol precursor in the virion. SDS-PAGE analysis of the cleavage sample revealed a 5.5 kDa fragment. This processed fragment of the Pro-Pol protein was further treated with trypsin and analyzed by capillary LC-MS and MS/MS. Analysis of the fragment by MS/MS revealed the N-terminal cleavage at Leu^{157}-Pro^{158} of the *pro* ORF. These results confirm the authentic

amino acid sequence of the reverse transcriptase of HTLV-1 RT. The data reported here provides a basis for further investigation of the functional and structural aspects of protein-nucleic interaction of the enzyme.

In loving memory of my grandmother, Estefania Gundayao, whose battle with illness has made my research much more meaningful

ACKNOWLEDGMENTS

First, I wish to thank my dissertation advisor, Dr. Laura Moen. My research project could not have been completed without your guidance and encouragement. Even in your absence, you were always willing to lend support and offer suggestions. Thank you for being there for me and convincing me that I WAS GOING TO GRADUATE! It's the "moth thing."

To my dissertation committee, Dr. Elliott, Dr. Pleban and Dr. White. Your advice and numerous suggestions have made a world of difference.

To my labmates, Barbara Conyers, Vivek Anantharaman, Dennis Moore, Hidayah Kendall, and Steve Scherer. Your presence has made an enjoyable working environment. Special thanks to Barbara for teaching me how to find humor in anything. I will always remember our afternoon laughter. Many thanks to Vivek and Hidayah for keeping me company in the lab on the weekends and also, your advice.

To my family. Mom and Dad, your love and support have meant more to me than you'll ever know. You are two truly amazing people. Pam, thank you for all your advice, faith and encouragement. Thank you for subconsciously reminding me that I needed to take a break once in awhile. I couldn't ask for a better big sister.

To my fiancé, Joe McCoy, Jr. Your unconditional love, understanding and patience have helped me make it through some of my toughest days of stress.

To my friends and fellow graduate students, Elizabeth Carey, Caroline Lauderbaugh, Sue Gitlin and Henri Parson. Our numerous conversations about "Life as a Graduate Student," should be made into a novel. Thank you for taking the time to listen.

To Joni Danganan and John Dashiell. You two were the best of roommates and are the best of friends. I know I can always count on you both to listen, no matter what time it is.

Finally, to Dr. Paul Ratz. You may not realize it, but you are the reason I have written this dissertation. Thank you for teaching me that I can be a scientist and that ultimately, it is my decision where I want to go and what I want to do.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Human T-cell Leukemia Virus 1 (HTLV-1) is a type C human retrovirus initially reported to be the causative agent of Adult T-cell leukemia (39-40, 96, 117, 136-138). The virus was first isolated in 1980 by Poiesz et al. from a cell line derived from a patient with Adult T-cell Leukemia (ATL) (14, 35). ATL is a mature T-cell non-Hodgkin lymphoma with a leukemic phase characterized by circulating $CD4^+/CD25^+$ T-cells (134). We now know that HTLV-1 infection has a much broader spectrum of disease manifestations including HTLV-1 associated myelopathy, tropical spastic paraparesis (HAM/TSP) (28, 38, 90, 100-101, 112, 120), uveitis (79) and an infectious dermatitis in children (60). Several inflammatory and immune-mediated conditions such as polymyositis, arthropathy, Sjögren's syndrome, and facial nerve palsy have been associated with HTLV-1, although, a clear etiolgoical relationship has not been established (7, 48, 128).

While the pathogeneses of these diseases is unknown, all involve activated, HTLV-1 infected $CD4^+$ T-cells. The human T-cell lymphotropic virus resides in and functionally alters immune cells of central importance for immunoregulation. Thus, infection of $CD4^+$ T-cells (immune cells) by HTLV-1 has several influences on the immune response: (1) interference with signaling pathways and transcriptional regulation; (2) activation of resting T-cells which propagates the infection; (3) activation of $CD8^+$ cytotoxic T-lymphocytes (CTLs) which alters cell death pathways in the host T-cell; and (4) induction of a strong antiviral immune response, which nonetheless is incapable of eradicating the infection (44).

EPIDEMIOLOGY OF HTLV-1

Worldwide, HTLV-1 infects between 11 and 20 million people. The infection is

---

The model for this dissertation is *Archives of Biochemistry and Biophysics*.

endemic in Southern Japan, the Caribbean basin (10), parts of South America (4), parts of Africa (19, 143), the Middle East, the Pacific Melanesian Islands (150) and Papua New Guinea. The prevalence in these endemic populations ranges from 5-30% (88, 93, 141). The seroprevalence among low-risk populations, such as those of the southern United States, is estimated to be 0.025% (70, 144). Patients who acquire HTLV-I are at risk of developing both infectious and noninfectious sequelae. Seropositive individuals have a 0.1 to 1.0% risk of developing HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP) (133). HAM is a chronic progressive inflammatory demyelinating myelopathy, often clinically confused with multiple sclerosis, that typically presents 10 to 30 years after infection with lower limb weakness, profound bladder dysfunction and minimal sensory loss.

Seropositive individuals have a 1 to 4% lifetime risk of developing ATL (133). Disease manifestations occur after a latent period and are classified by clinical and laboratory criteria as acute, chronic, lymphoma or smoldering ATL. Widespread or localized skin lesions such as large nodules, plaques, ulcers and a generalized papular rash are common. Immunosuppression is well documented, manifesting as bacterial and opportunistic infections which contribute to a poor prognosis (37, 120, 140).

Infectious dermatitis has been reported from several HTLV-I endemic populations including Japan, Trinidad, Brazil, and Columbia. The disease is characterized by a severe exudative dermatitis of the scalp, external ear and retroauricular aras, eyelid margins, paranasal skin, neck, axillae, and groin as well as generalized fine papular rash. The incidence and prevalence of this disease are undefined, as is the pathogenesis (61, 151).

Idiopathic uveitis was observed to be very common in the HTLV-I endemic areas of Kyushu, Japan, leading to speculation that HTLV-I infection might be the cause (79-80). Patients with HTLV-I associated uveitis will often complain of blurred or foggy vision and acute, sudden onset of floaters. Blurred vision is associated with acute onset of mild inflammation in the vitreous body followed by mild iritis and retinal vasculitis.

HTLV-I is, thus, the causative agent in clinical disorders affecting several organ systems. Excess morbidity and mortality have recently been reported among carriers (3)

indicating that the full public-health impact of HTLV-1 may be much greater than generally perceived.

MODES OF TRANSMISSION

The virus may be transmitted both horizontally through blood products, sexual contact and shared needles and vertically through breast milk, transplacentally, and intrapartum. Of these modes of virus transmission, transfusion is perhaps the most efficient; the probability of seroconversion in the recipient of contaminated blood is 40-60% (74, 89). HTLV-1 transmission by blood products can lead to rapid development of HAM or TSP and has been seen within 6 months of transmission (35). Today, routine blood-donor screening for antibodies to HTLV-1 is performed in the USA and other countries.

Most HTLV-1 infections are attributable to transmission from mother to child through prolonged breastfeeding or transmission by sexual contact later in life. Mother-to-child transmission of HTLV-1 occurs primarily through ingestion of milk-borne lymphocytes (142) whereby transmission efficiency is dependent on duration of breast feeding and the presence of maternal antibodies to HTLV-1.

METHODS OF DETECTION

The identification of HTLV-1 by the detection of antibodies in infected individuals is important because knowledge of HTLV-1 seropositivity may help to prevent the transmission between sexual partners, as well as transmission from mother to child and blood transfusion. Detection of antibodies to HTLV-1 is useful in establishing a diagnosis of ATL, HAM/TSP and other HTLV-1-associated diseases, and for screening of blood donors and pregnant women. Methods to detect HTLV-1 antibodies include gelatin particle agglutination (PA), indirect immunofluorescence (IF), immunoperoxidase staining (IP), enzyme-linked immunosorbent assays (ELISA) with disrupted whole virus, synthetic peptides or recombinant proteins, radioimmunoprecipitation assay and Western blotting (73, 147).

ELISAs using whole virus lysate are the most commonly used screening assays for human sera or plasma in the United States and Europe. In Japan, the particle

agglutination assay is most commonly used for screening. Guidelines from the US Public Health Service and other international groups (78) recommend that newly identified seropositive individuals have additional blood collected for repeat testing to eliminate possible technical errors. IF, IP, radioimmunoprecipitation assay (RIPA) and Western Blotting (WB) assays are commonly used as supplementary assays to confirm specificity of the antibodies to HTLV-1. The most common WB assays are constructed from whole virus lysate with the addition of recombinant envelope antigens and/or HTLV-1 and HTLV-2 specific envelope peptides to distinguish HTLV-1 to HTLV-2. HTLV-2 shares a 60% amino acid homology with HTLV-1 (24, 121). Consequently, it is difficult to distinguish the two unless virus-specific reagents are used. This distinction is important because HTLV-2 is less pathogenic than HTLV-1 (63).

Additionally, DNA analysis is used to diagnose ATL and other HTLV induced disease. Specifically, the polymerase chain reaction (PCR) has been used to identify type-specific proviral sequences in peripheral blood mononuclear cells from infected individuals to distinguish HTLV-1 from HTLV-2 and to detect DNA in tumor tissue and other biological specimens (95, 106).

Because HTLV-1 does not need a specific receptor for infection, it not only infects lymphocytes but also a variety of other cells. After infection, HTLV-1 randomly integrates its provirus into the chromosomal DNA. Since ATL cells show clonal proliferation in HTLV-1 infected T-lymphocytes, the demonstration of clonal integration of the HTLV-1 provirus is essential for the diagnosis of ATL (148). Southern blot analysis is used to identify the clonal proliferation of HTLV-1 cells. However, Southern blot analysis requires 5% infected cells for detection (147). The PCR method is much more sensitive, yet, it cannot distinguish whether HTLV-1 infected cells proliferate monoclonally or polyclonally. Inverse PCR (IPCR) enables the identification of unknown sequences flanking known sequences such as retroviral DNA. (76, 132). Therefore, IPCR is used to detect the monoclonality of HTLV-1 infected cells.

RETROVIRAL REPLICATION

A virus is a life form that lacks its own metabolism, and is therefore dependent on

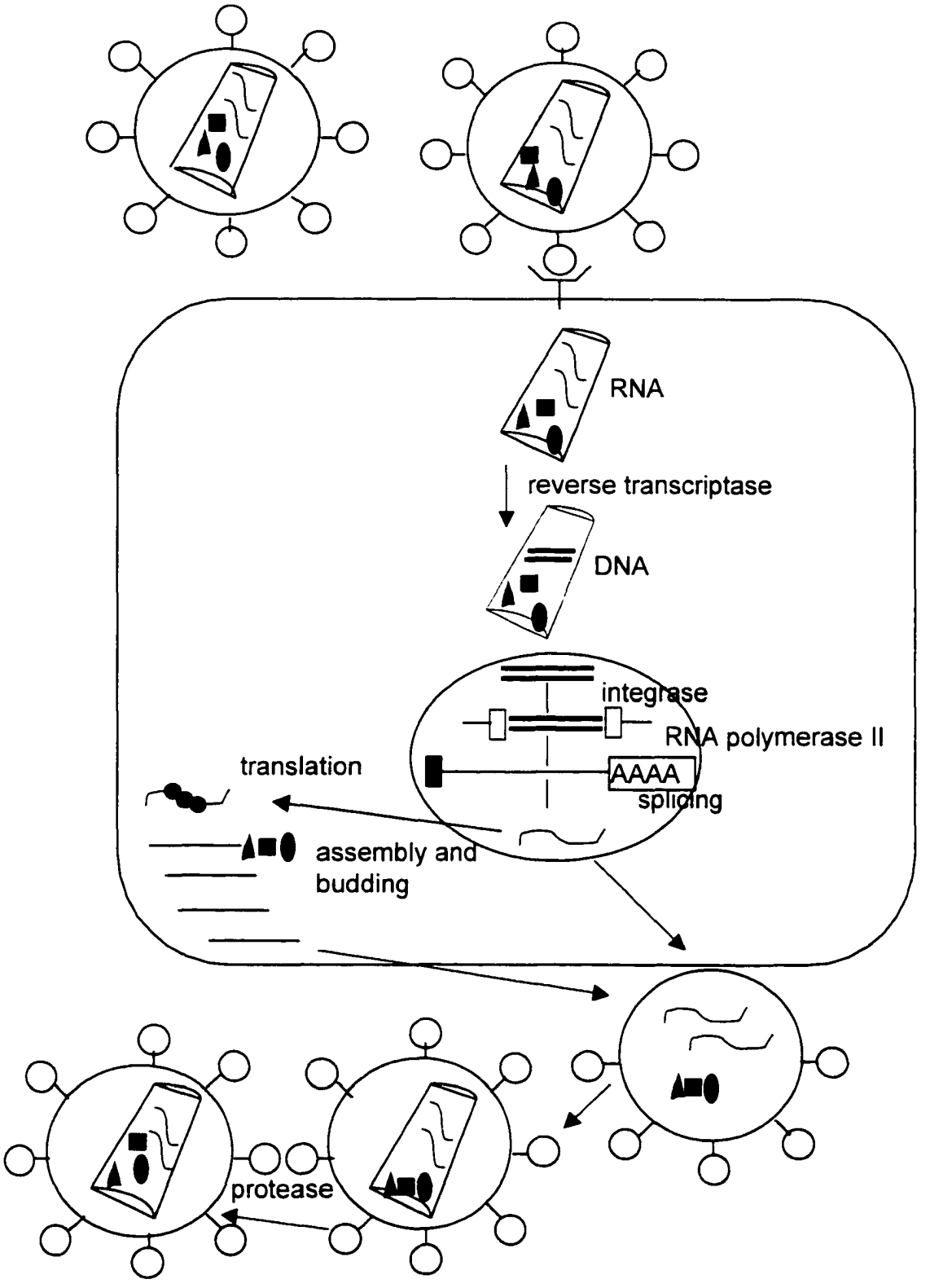


FIG. 1. Retroviral life cycle. The virus particle attaches to the host cell and through the catalytic activities of retroviral enzymes, a new virus is produced.

the metabolic apparatus of a host cell for its replication. The core structure of a retrovirus contains its genetic code in two RNA molecules. In addition, each virus particle contains enzymes and other proteins that are required for infection of host cells and conversions of them to virus factories. The surface of the virus particle contains glycoprotein molecules that mediate attachment of the virus to a receptor protein ($CD4^+$) located on the plasma membrane of the infected cell (Fig. 1). Upon entry into the host cell, the subviral complex containing the genomic RNA molecules and limited viral proteins is released into the cytoplasm (124). Single-stranded RNA chains are converted to double-stranded DNA by the enzyme, reverse transcriptase. The double-stranded DNA is then integrated into the host genome by another enzyme, integrase, to form the provirus.

The viral proteins take over the control functions of the host cell and can start making new viral particles. The reverse transcriptase enzyme of the host cell starts to produce multiple copies of messenger RNA encoding new viral materials. Translation of the mRNA results in the production of long peptide chains. The long peptide chains or polypeptide precursors and viral RNA are then exported to budding particles that form on the cell surface. Completion of the virion occurs by pinching off the budding particle at the cell surface. In the last stage of budding, the polyprotein precursors are proteolytically cleaved at domain boundaries by the viral protease to liberate mature proteins.

MOLECULAR ORGANIZATION OF HTLV-1

As previously stated, HTLV-1 has been associated with a wide spectrum of clinical manifestations including cancer, immunological manifestations and neurological disorders. Because of its clinical relevance, the genes of the retrovirus have become a focal point of retroviral research for scientists today. As depicted in Fig. 2, the viral genome of the HTLV-1 provirus consists of two long terminal repeats (LTR), the typical *gag, pro, pol*, and *env* genes, and a novel 3' end which encodes the two trans-regulatory proteins, tax and rex (120-121).

The LTR enables the genome to insert itself into the genome of a host cell. In the proviral DNA, the LTR contains signals which control and affect transcription, splicing, and packaging of RNA into virions.

The *gag* and *env* genes are known to be essential for the production of infectious progeny virus, in that, they encode proteins that constitute the virus structure. The gag gene encodes structural proteins, specifically the core proteins p19 (matrix protein) and p24 (capsid protein). The protease is primarily responsible for processing of the *gag* gene products. The *pol* gene, located between the *gag* protease genes and the *env* gene (52).

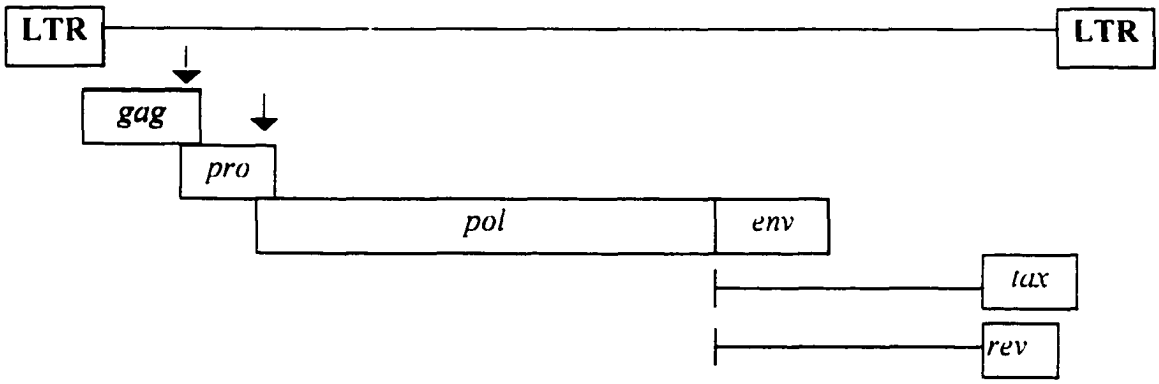


FIG. 2. HTLV-1 genome. Long terminal repeats (LTR) and open reading frames are depicted. (↓) denotes frameshift sites.

includes an RNA-directed DNA polymerase, the reverse transcriptase, an RNaseH region which degrades viral RNA in the immediate wake of its reverse transcription and an integrase or endonuclease region which integrates the newly synthesized DNA into the host genome.

The *env* gene encodes the precursor envelope protein, gp61. Gp61 is cleaved by an unknown cellular enzyme to yield the small transmembrane (gp21) and large external envelope (gp46) glycoprotein. Gp46 interacts with a still unidentified cellular receptor to facilitate viral entry (22).

There exists a long ORF between the *env* gene and the 3'-LTR which is called the pX region. Double spliced mRNA enables this region to encode two regulatory genes, *tax* and *rex*. These regulatory genes are essential in controlling viral RNA and protein synthesis (86). Expression products of these genes yield proteins not present in the virus itself. The Tax protein is located in the nucleus and enhances viral transcription by

activation of three 21 base pair enhancer elements located in the U3 region of the viral LTR (155). Tax does not actually bind to this region, rather, it activates different cellular transcription factors that in turn bind to this region. IL-2α and its receptor are responsive to transactivation by Tax and expression of this receptor in HTLV-1 transformed cells has been reported (22). The Tax protein also activates sequences to which the nuclear factor κB (NFκB) binds and thereby may induce a set of genes expressed during T-cell activation.

The Rex protein is also located in the nucleus and functions in regulating the levels of unspliced to spliced viral mRNA. Rex is responsible for exporting full length *gag/pol* mRNA and single spliced *env* mRNA from the nucleus to the cytoplasm (55, 106). Export of mRNA is achieved by binding of a phosphorylated Rex protein to cis-acting RNA sequences , called the *rex* responsive elements in the viral RNA (4, 8, 11, 3, 119, 152). The Rex protein is therefore regarded as a positive regulator of expression of structural proteins of HTLV-1. However, an overexpression of rex may lead to negative regulation. Thus, HTLVs possess a self-regulatory mechanism to control their replication by products of the *tax* and *rex* genes (49).

RIBOSOMAL FRAMESHIFTING

All of the genes in the retroviral mRNA produce proteins essential for the survival of the virion. However, not all virion particles that are produced are infectious. The proteins translated from the mRNA in the nucleus are long polypeptide chains or polyprotein precursors which are transported to the cell membrane and incorporated in the virion. The budding virion is not considered infectious until the protease cleaves the polyproteins into their respective structural proteins. The mature virion components are generated by proteolytic cleavage of the primary translational products of the genome sized mRNA, designated the Gag, Gag-Pro, Gag-Pol, and Gag-Pro-Pol polyproteins. The synthesis of Gag-Pro, Gag-Pol, and Gag-Pro-Pol fusion proteins, which all share the common *gag* initiator codon located at the 5'end of the *gag* gene, is translationally regulated. Two general mechanisms by which the synthesis of the fusion proteins can be achieved are known: readthrough suppression and ribosomal frameshifting.

In readthrough suppression, an amino acid is inserted (by a suppessor tRNA) where a stop codon is located. Through the actions of these suppressor tRNAs, some ribosomes are able to bypass the designed stop codon during translation of the mRNA and a fusion protein is produced instead of the Gag polyprotein.

In mammalian type C retroviruses, the *gag* amber terminator (UAG) is suppressed by a glutamine-tRNA to translate the *pol* gene (153). However, Nam, et al. found in HTLV-1, that the Gag-Pro fusion protein occurs by ribosomal frameshifting (85). Ribosomal frameshifting involves the slippage of a tRNA one nucleotide in a forward or reverse direction. As a result, the *gag* and *gag-pol* sequences are out of frame with respect to each other. In the HTLV family of retroviruses, Bovine Leukemia Virus (BLV), Mouse Mammary Tumor Virus (MMTV) and type D retroviruses, the *gag* and *pol* genes are separated by a *pro* ORF that overlaps both *gag* and *pol* (48, 81, 84, 98, 102, 108 117 127. The protease gene may be located at either the 3'end of *gag* or the 5' end of *pol* i.e. Human Immunodeficiency Virus 1 (HIV-1) and Rous Sarcoma Virus (RSV) (115. Thus, HTLV-1 utilizes a slight variant of the frameshift mechanism to form their Pol proteins. Two successive frameshifts must occur: (1) moving from the *gag* reading frame into a short intermediate gene encoding the protease, and (2) moving from that reading frame into the *pol* frame (Fig. 2). These frameshifts occur by a slippage of asparagine tRNA 0in the −1 direction (83-84).

PROTEOLYTIC PROCESSING

After the HTLV-1 fusion proteins are translated, they are proteolytically processed by the retroviral protease to produce the Gag, Pro and Pol proteins. The viral protease is encoded in a separate reading frame overlapping the *gag/pol* coding sequences (23, 32) (Fig. 2). Thus, the protease must, itself, be excised from a precursor protein. The protease may autoprocess from within its precursor protein by cis catalysis. Alternatively, the protease may be cleaved by trans catalysis through the actions of a neighboring Gag-Pol precursor (23).

Through site-directed mutagenesis studies, Nam et al. (1988) determined that the catalytic center of the HTLV-1 protease, consisting of one of the two conserved regions

common to retroviral proteases, was identical to that of an aspartyl protease (85). Other reports confirm that mutations in Asp^{125} abolish protease activity (58, 82). Retroviral aspartic proteases have a optimal pH range of 4-6 while eukaryotic aspartic proteases have an optimal activity range of pH = 2-4. The difference in optimal activity range has been attributed to the lack of hydrogen bonding between Asp and Thr or Ser in eukaryotic proteases. In retroviral proteases, the corresponding Thr or Ser is replaced by Ala. No corresponding hydrogen bonding can occur; thus, the active Asp becomes less acidic, and the pH optimum increases (33).

The mechanism by which the HTLV-1 protease cleaves its substrate is modeled after that of the HIV-1 protease since they are both aspartic proteases (46-47, 54) (Fig. 3). The aspartic residues in the active site of each monomer share a proton. The lytic water molecule forms hydrogen bonds to the charged carboxylate. The C_β-C_γ bond of Asp^{125} rotates to allow its proton to hydrogen bond to the carbonyl oxygen of the substrate. The Asp^{25} becomes more basic than Asp^{125} due to the loss of the shared proton. The proton from the lytic water can be abstracted to the carbonyl oxygen, forming a protonated amide species, EA'H. The carbonyl carbon of EA'H becomes an electrophilic substrate for the water molecule. The amide hydrate intermediate, EXH, is then formed. Two proton transfers lead to the tetrahedral intermediate EX'H. Finally the peptide bond is broken and the fragments are created. The enzyme is then released in free form and may process more substrate.

HTLV-1 REVERSE TRANSCRIPTASE

While a considerable amount of information is known about the structure of the protease and gag proteins, not much is known about the structure and proteolytic processing of the reverse transcriptase of HTLV-1 RT. Therefore, to further our knowledge of molecular events during retroviral DNA synthesis, it is essential to understand the architecture and functional organization of the reverse transcriptase molecule. Information about HTLV-1 RT, would allow for better understanding of the biology of these viruses.

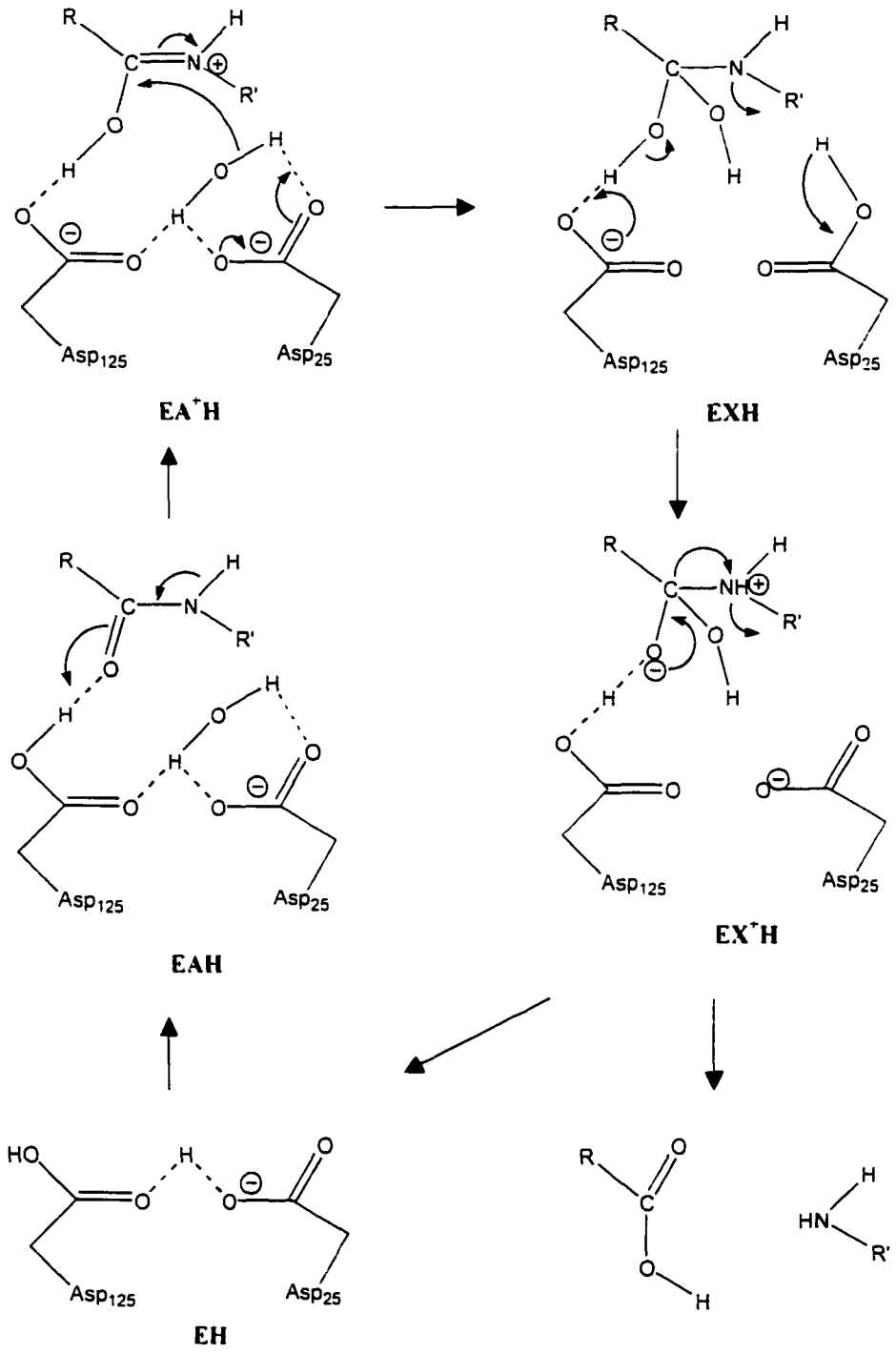


FIG. 3. Proteolytic mechanism for the hydrolysis of the Pro-Pol protein. The aspartic residues of the free enzyme (EH), share a proton. After attack of the substrate (EAH), a protonated amide species is formed (EA*H). The carbonyl carbon of EA*H is attacked by water to form the amide hydrate intermediate. EXH. After two proton transfers, a tetrahedral intermediate is formed (EX*H). Finally the peptide bond is broken and the products are free to dissociate. The enzyme can then trap another water molecule.

The reverse transcriptase (deoxynucleoside triphosphate, DNA deoxynucleotidyl transferase, RNA-directed E.C. 2.7.7.49) of all retroviruses is encoded by the *pol* gene and synthesized by the translation of a full length viral mRNA, identical in structure to the genomic RNA packaged into virions. It is never expressed as a separate protein, but rather translated as part of a large Gag-Pro-Pol protein (polyprotein precursor) (65). The reverse transcriptase exhibits three enzymatic activities: RNA-dependent DNA polymerase, Ribonuclease H (RNaseH) and DNA-dependent DNA polymerase (54). All of these enzymatic activities are essential for transcribing the single-stranded genomic RNA found in the virions into the double-stranded linear DNA that is the precursor to the integrated provirus.

The process of reverse transcription is depicted in Fig. 4. As in the case of all DNA polymerases, RT needs a primer carrying a free 3'-OH group to initiate DNA synthesis. Retroviruses use host-encoded tRNAs as primers (64, 67, 72, 75). The tRNA anneals to the primer-binding site (PBS) located near the 5' end of the viral plus-strand RNA. The RT recognizes the tRNA/RNA complex and initiates reverse transcription of the minus-strand DNA by extending the 3' end of the primer. The recessed template strand guides DNA synthesis (6, 50). In copying the U5 and R regions of the genome, the RT forms a strong stop cDNA. Coupled with the synthesis of DNA is the degradation of the RNA strand of the newly formed RNA/DNA hybrid by the RNaseH (25, 34, 114).

The strong stop DNA is transferred to the 3'end of genomic RNA in order to continue synthesis of the minus-strand DNA. This first strand transfer is facilitated by a complementary R region at the 3'end of the viral RNA. The two complementary R sequences allow hybridization of minus-strand DNA with plus-strand RNA, permitting DNA synthesis to resume. As previously, the RNaseH activity of the RT degrades the RNA template. Synthesis of minus-strand DNA also entails the creation of purine rich fragments that function as RNA primers for plus-strand DNA synthesis (13-14, 45, 56, 146). Thus, the minus strand U3-R-U5 still attached to the tRNA primer is copied to form the plus-strong-stop DNA. As plus-strand DNA synthesis proceeds to the PBS, the original primer tRNA is degraded by the RNaseH activity.

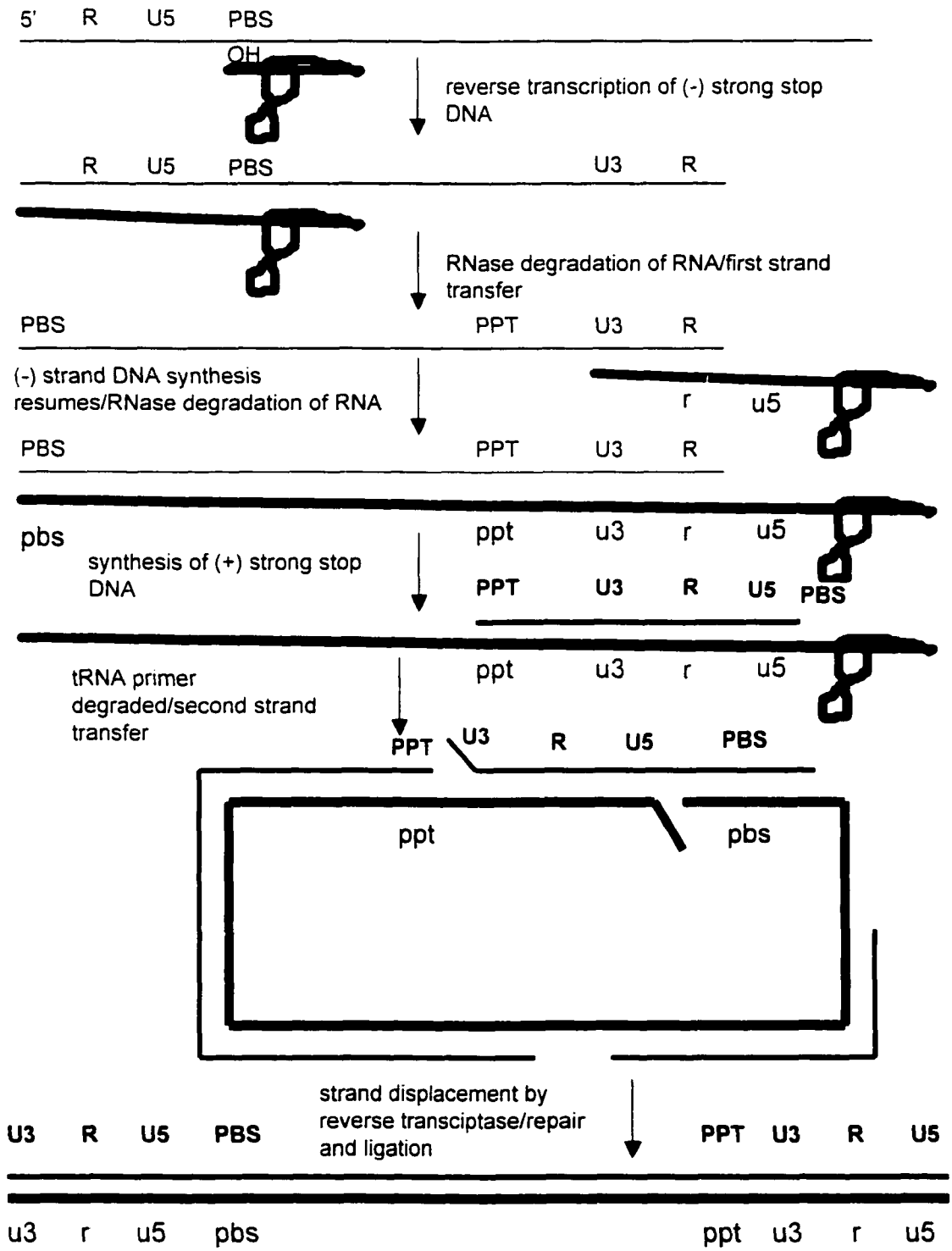


FIG. 4. Retroviral reverse transcription. The reverse transcriptase produces a double-stranded DNA from a single-stranded RNA through a sequence of steps involving two stand transfer events. The resulting linearized DNA is called the proviral DNA and is inserted into the host to form the provirus.

Once the primer tRNA is released, the second strand transfer occurs. The PBS on the plus-strong-stop DNA anneals to the complementary PBS on the 3' elongated minus strand and ligation of the circular intermediate produces a linear duplex with LTRs (54). The LTRs are duplicate R, U5, and U3 sequences. Thus, the proviral DNA differs in length compared to the original RNA genome whose LTR contain the R region and either the U5 or U3 region, respectively. The proviral DNA is then integrated into the host to form the provirus.

There are common biochemical properties of RT from different viruses, i.e. primer-templates and divalent cations. But, they also differ in molecular weight, in subunit composition, function, and in their sensitivities to polymerase inhibitors. For example, Murine Leukemia Virus RT is a single polypeptide chain that carries both reverse transcriptase activity and RNaseH activity, while some retroviral reverse transcriptases may be divided into two domains each exhibiting only one enzymatic activity. Avian Myoblastosis Virus RT consists of two subunits, alpha and beta. The larger subunit carries the integrase function as well as the RNaseH and polymerase activities; the smaller subunit carries polymerase and RNaseH activities (43, 125)

PREVIOUS RESEARCH

To date, not much is known about the reverse transcriptase of HTLV-1. Retroviral polymerases from different taxonomic classes show a preference for certain synthetic template-primer combinations. Transcription also requires a divalent cation: Mg^{2+} or Mn^{2+}. The majority of retroviruses preferentially use Mg^{2+} (65). In the first reported information about HTLV-1 RT, Rho et al.(101) characterized the enzyme from the virion. They found that this RT has an apparent MW of 95,000, that Mg^{2+} is the preferred divalent cation, and that poly(rC)•oligo(dG) is the preferred primer-template *in vitro*. However, this group could not obtain much protein from their isolation method from human cell infected lines. Further characterization of the enzyme was not performed. Subsequently, Hoffman et al. found that HTLV-1 RT isolated from infected MT-2 cells preferred Mn^{2+} with both a poly(rA)•oligo(dT) and poly(rC_m)• oligo(dG) primer template, but Mg^{2+} for a poly(rC)•oligo(dG) template primer (43). Most recently, Owen

et al. expressed a recombinant HTLV-1 protein with reverse transcriptase activity. This protein has an approximate MW of 60,000-65,000, and prefers a Mg^{2+} divalent cation on a poly(rC)•oligo(dG) template (91).

The exact sites of protease cleavage have been identified for a number of viral polyproteins by comparing the amino acid sequences of the precursors, deduced from the proviral DNA sequence. The N- and C-terminal amino acid sequences of mature viral proteins were determined by direct protein sequencing techniques. However, the actual identity of the authentic N-terminus from virion-produced HTLV-1 RT has not been identified from mature, processed protein.

Two very recent reports have been published on the expression of the RT of HTLV-1. In the first, Owen et al. expressed an enzymatically active HTLV-1 RT in bacteria. They began their construct such that translation was initiated at Pro^{33}, which is located within the proposed pro-pol frameshift site (91). This group made no effort to determine the correct N-terminal sequence (proteolytic processing). In the second report, Trentin et al. hypothesized that the N-terminus of the RT should immediately follow the known C-terminus of the protease. By using sequence alignments with other retroviral sequences, they demonstrated that there was no known function for the sequence encoded between the C-terminus of the protease and Pro^{33}. Based on the similarity of the sequence of other RTs, the protein coding sequence should begin at Pro^{33}. The protein resulting from their construct also had RT activity (131). However, they used an *in vitro* transcription/translation system which could have an impact on the overall folding and potential subunit structure. In addition, Trentin et al. assumed that the HTLV-1 RT sequence is similar to that of the RT of avian sarcoma virus (ASV). Based on this, they deduced an oligomeric structure of the RT and claimed to have obtained an exclusive enzymatically active RT form, the α_3/β tetramer. This is a molecular arrangement never encountered before in retroviral RTs (16, 41, 125). Taken together, there are discrepancies between the two published reports on recombinant HTLV-1 RT. Furthermore, Perach et al. (92) found that the amino acid sequence requirements and molecular arrangements of the recombinant BLV RT are substantially different from those of both HTLV-1 RT studies. These unique differences among relatively close RTs

should be helpful to future studies designed to study structure-function relationships of these RTs.

HTLV-1 viruses show very limited sequence variation in comparison to a much greater variation seen among different isolates of each lentivirus in the HIV/SIV family (28, 30, 57, 71, 77). Ratner et al. reports a 1.2%-3.3% nucleotide sequence difference among HTLV-1 isolates (100). The identification of possible regions of functional importance can be unequivocally determined by comparative sequence analysis because of the genomic stability of HTLV-1 and minimal strain variation between isolates. As each of the functional aspects becomes better understood, targets for antiviral intervention against HTLV-1 associated diseases will emerge.

DEVELOPMENT OF THE RESEARCH METHODOLOGY

In light of the facts presented, the ultimate goal of this project is to characterize the reverse transcriptase enzyme of HTLV-1. Successful expression of this replication enzyme would allow for a detailed biochemical characterization as well as a comparison of structure-function relationships with reverse transcriptases from HIV and other retroviruses. The expression system for the HTLV-1 RT was developed in 1995 by my advisor, Dr. Laura Moen. The plasmid, pRCH (generously provided by Dr. Lee Ratner of Washington University), which contained the cDNA from a replication competent isolate of HTLV-1, was used to create the desired protein expression system. The intact HTLV-I *pol* gene was cloned using a pET vector for expression. The desired sequence was amplified using PCR. To mimic the second frameshift in the pol ORF, an extra C was inserted at the second frameshift site when the primers were made. Since the expression vector used to insert our clone (pET 24) contained the Bam and EcoR1 restriction sties, our clone was made such that the clone had a Bam site at its N-terminus (Pro^{158}). After the expression vector is cut open with Bam and EcoR1 restriction enzymes, the clone is ligated in the opened vector. This coding sequence was cloned in frame with the 10 amino acid T7 tag sequence in pET24b to facilitate expression and detection on a Western.

Originally, two clones were created because the structure of the RT and the cleavage sites in *pol* were unknown. The *pol* ORF of HIV-I also encodes for reverse transcriptase.

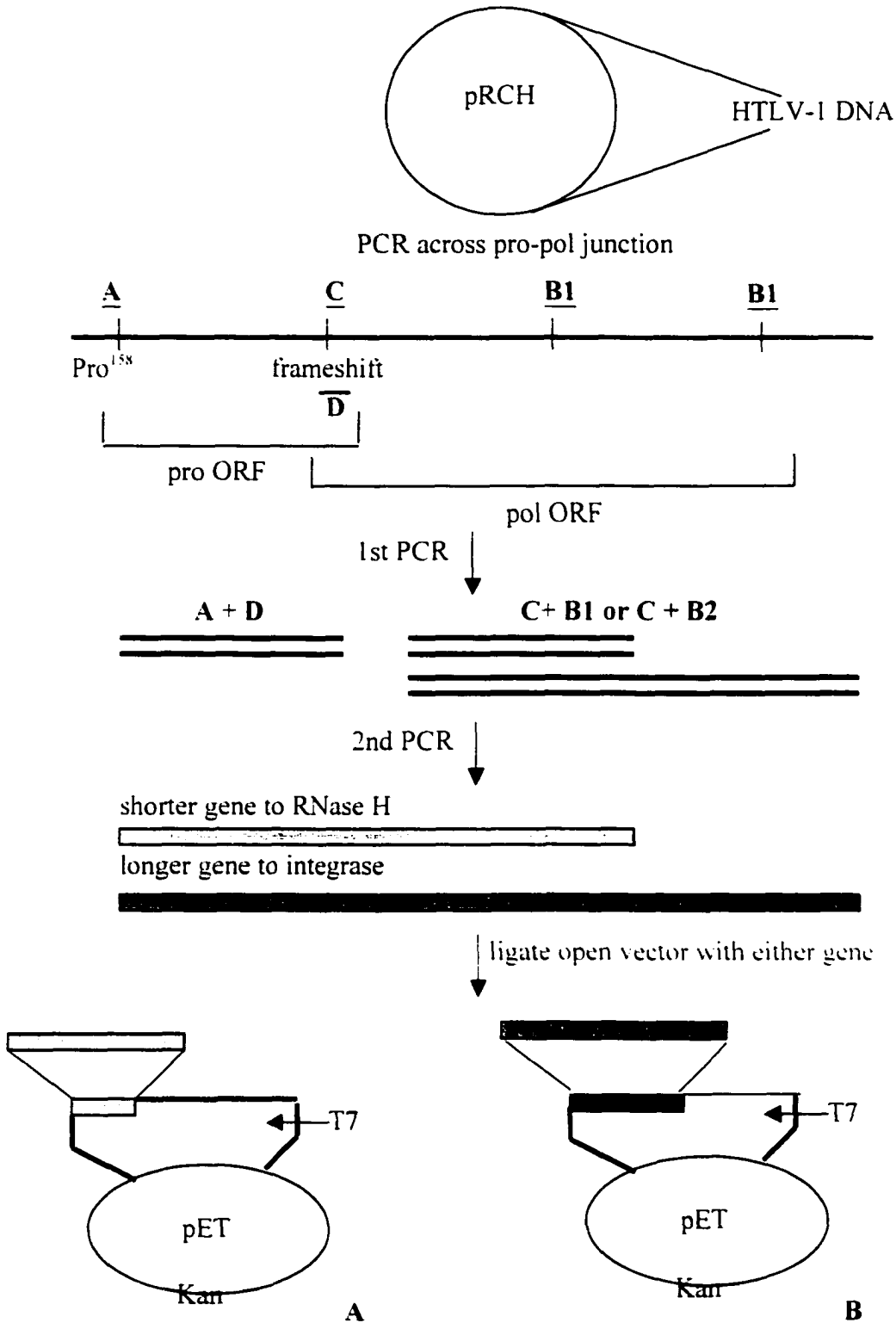


FIG. 5. Strategy for the amplification of the intact HTLV-1 RT coding sequence. (A) Expression system for reverse transcriptase with RNaseH; $M_r = 65{,}000$. (B) Expression system for reverse transcriptase with integrase: $M_r = 99{,}000$.

RNaseH and integrase (endonuclease) domains. This Pol polypeptide is further processed to produce a mature reverse transcriptase and a separate mature endonuclease protein (23). The HTLV-I RT protein may act in a similar fashion. On the other hand, the HTLV-I RT protein may have similar functions to other reverse transcriptases such as Rous Sarcoma Virus RT which possesses both a reverse transcriptase and endonuclease domain (23). Given this background information, the clones shown in Fig. 5 were produced: (1) The short clone was constructed with a stop codon at amino acid 591 which terminated the coding sequence at the end of the RNaseH domain. (2) The large clone included the integrase domain and terminated at the native *pol* stop codon.

Since the sites of protease cleavage in the *pol* region have not been identified other than those required for the production of mature protease, the clones were created with a putative N-terminus. The actual identity of the authentic N-terminus from virion-produced HTLV-1 RT has not been identified and confirmed. Dr. Laura Moen and her post-doctoral associate, Imogene Richardson, carried out a sequence alignment with several reverse transcriptases which included that from Bovine Leukemia Virus, the closest characterized relative to HTLV-1. A portion of that alignment, shown in Fig. 6, suggests that the N-terminus of HTLV-1 RT is likely to start where the viral protease C-terminus ends. Therefore, both of the clones were constructed to contain this sequence at the N-terminal region.

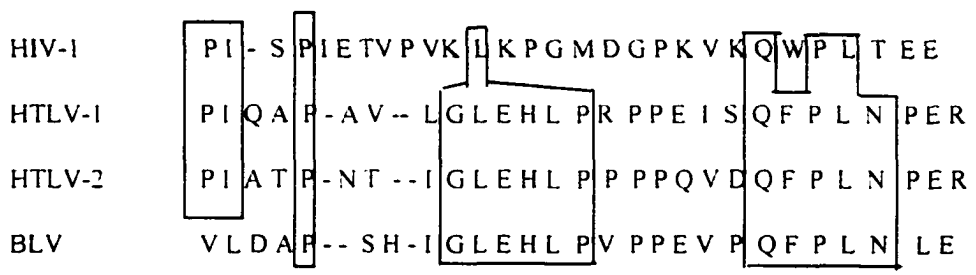


FIG. 6. Sequence to determine the N-terminus of HTLV-1 RT. The amino acid sequences at the N-terminal regions of HIV-1 and BLV RTs are aligned with the N-terminal regions for HTLV-1 and HTLV-2. The last boxed region contains the sequences surrounding the frameshift in the *pol* reading frame (between L and N of HTLV-1 and HTLV-2).

RESEARCH OBJECTIVES

As previously stated, Owen et al. reported that a smaller recombinant protein from an MT-2 HTLV-1 cDNA exhibits RT activity (91). The coding sequence of this protein begins after the ribosomal frameshift and ends at the end of the RNaseH domain. Efforts to produce the short clone which would be analogous to that reported by Owen et al. did not succeed: the smaller protein was rapidly degraded as it was synthesized. The instability of this protein that was observed may be due to as yet undetermined protein folding differences between the two proteins and perhaps by the different expression systems employed. The results of expression of the longer clone which included the integrase domain, however, showed significantly higher levels of expression on Western blots. It is with this longer clone that a purification scheme was developed. Further investigation was necessary to clarify the reasons for folding differences and to unequivocally identify the mature form of the authentic HTLV-1 RT. The specific aims of this research are to:

1. Purify a recombinant HTLV-1 RT

With the availability of primary sequences for RNA-dependent DNA polymerases, it has become possible to compare and delineate some of the highly conserved regions found in polymerases from diverse origins. Sequence alignments among these polymerases have revealed four conserved motifs (94). Genetic studies of these conserved regions have demonstrated a critical role for these residues in the catalytic function of these polymerases (9, 31, 62, 97, 109). For example, the invariant DD motif in motif 3 of these polymerases is thought to be at the polymerase active site. For related polymerases, published procedures can be used as a starting point and modifications made later to improve the specific activity, purity, and recovery of the enzyme. Since the location of the active site of the HTLV-1 RT is not known and thus, the amino acids involved in catalysis, other retroviral reverse transcriptase purification schemes were used as a basis. Consequently, several different ligands were analyzed.

Several sensitive methods exist for protein characterization. These chromatographic methods include ion exchange, DNA affinity, metal chelate affinity, hydrophobic interaction, precipitation, ammonium sulfate, polyethyleneimine and high

performance liquid chromatography (HPLC). All of these methods were used to empirically determine the optimal conditions for adsorption of the RT. Analysis of purity of the final product was assessed using SDS-PAGE analysis, gradient gels, gel filtration, Western Blotting, and measurement of reverse transcriptase activity. Further characterization involved template specificity and stability studies.

2. Confirm the proteolytic processing site of the N-terminus of HTLV-1 RT

Results from the stability studies were in agreement with others in that the activity of the enzyme is consistently low. An earlier report by Hizi, et al. demonstrated that the first 20 amino terminal acids of HIV-1 RT are required for an active enzyme (42). The low activity observed in this study and/or others may be due to an incorrect coding sequence for the HTLV-1 RT. Moreover, the N-terminal proteolytic cleavage site of HTLV-1 RT has not been determined. Therefore, to confirm the HTLV-1 RT purified in this project contains the authentic N-terminus, clones coding for the expression of a protein fragment spanning the *pro-pol* junction were constructed. The clone was purified using methods described in objective 1 and cleaved using recombinant HTLV-1 protease.

The peptide fragments were analyzed by LC-MS and MS/MS at the Biomolecular Research Facility at the University of Virginia by Dr. Nicholas Sherman. In recent years, mass spectrometric methods have emerged as powerful tools for protein characterization (124). Protein amino acid sequences can be determined by tandem mass spectrometry using a triple-quadrupole mass spectrometer. In this process, protonated peptide ions are selected with the first mass spectrometer and are transmitted to a collision cell. Low energy collision with neutral inert gases results in fragment ions that are then detected in a second mass spectrometer. The fragmentation patterns of peptides obtained are indicative of their amino acid sequences and thereby allow for the determination of peptide sequences as well as determination of any posttranslational modifications occurring within the protein. Analysis of the Pro-Pol protein fragments by LC-MS and MS/MS was used to determine the location of the cleavage site which is presumably used to generate mature reverse transcriptase from Gag-Pro-Pol polyproteins expressed *in vivo* in HTLV-1. Since the sequence of the Pro-Pol protein is already

known, confirmation of the peptide fragments can be rapidly performed by mass measurement.

PROJECT SUMMARY

The retroviral replicative cycle has been intensively studied in order to identify possible targets for designing inhibitors, which may become powerful therapeutic agents. While HTLV-1 is not as prevalent as other retroviruses i.e. HIV, coinfection with HIV may have a serious impact. In order to deal with the consequences of a widespread infection and disease before an epidemic occurs, we can study the replicative mechanisms of HTLV.

The reverse transcriptase plays a crucial role in the early stages of viral replication. To understand the architecture and functional organization of the reverse transcriptase molecule, a recombinant enzyme expressed in *E.coli* will be purified. By using a recombinant source of HTLV-1 RT, some of the biochemical properties of HTLV-1 RT can be determined. Confirmation of the protein as authentic will provide a basis for further studies on the structural aspects of protein nucleic acid interaction.

CHAPTER II

MATERIALS AND METHODS

PREPARATION OF MEDIA

To optimize the growth of the two different types of *E. coli* strains, nutrient rich media was prepared. Luria Broth (LB: Difco, Detroit, MI) was prepared by dissolving 25 g of LB powder in 1 L MilliQ water to a final concentration of 10 g bacto-tryptone, 5 g bacto-yeast extract and 10 g NaCl per liter of LB. The solution was autoclaved (Sterilmatic, Market Forge Co., Everitt, MA) at 240°C for 20 minutes.

1.4 g of agar (Fisher, Indianapolis, IN)/100 mL LB was autoclaved to create solid media for pouring plates. After cooling to 50°C, the solution was poured into sterile petri dishes (Fisher)) and allowed to polymerize at room temperature. If antibiotics were required in the media to enhance the growth of plasmid DNA and/or *E. coli* cultures, they were added to the media upon cooling to 50°C.

PREPARATION OF COMPETENT *E. coli*

The preparation of BL21(DE3)pLysS (Novagen, Madison, WI) and W3110(DE3) competent *E. coli* was performed using the $CaCl_2$ method outlined in Current Protocols in Molecular Biology (16). The BL21(DE3)pLysS host strain carries a plasmid with the T7 lysozyme gene which confers resistance to chloramphenicol. In which case, LB media was supplemented with chloramphenicol to a final concentration of 34μg/mL in all steps of this procedure. Unless otherwise stated, any protocol described hereafter in which the BL21(DE3)pLysS host *E.coli* strain is grown, utilizes LB supplemented with 34 μg/mL chloramphenicol as its growth media. The 34 mg/mL chloramphenicol stock solution was made by dissolving 34 mg of chloramphenicol (Fisher) in 1 mL 100% ethyl alcohol (AAPER Alcohol and Chemical Co., Shelbyville, KY).

An LB plate was streaked with the *E. coli* strain of choice using a sterile inoculation loop. The plate was incubated in a 37°C incubator (Fisher) for approximately 14-16 hours. The following day, a single colony was used to inoculate 3 mL of LB

media. The single colony was grown overnight at 37°C while shaking at 250 rpm (Orbit Environ Shaker, Lab-Line Instruments, Inc.). On the third day, a 1 mL aliquot of the overnight culture was inoculated into 100 mL sterile LB media in a 500 mL flask. The culture was grown at 37°C, while shaking at 250 rpm, to an $O.D._{590} = 0.375$. The culture was divided into 50 mL aliquots by pouring the culture into two pre-chilled, sterile, polypropylene centrifuge tubes. After leaving the tubes on ice for 10 minutes, the suspension was centrifuged for 7 minutes at 3000 rpm (Beckman GPR, Fullerton, CA). At all centrifugation steps in this protocol, the tubes were allowed to decelerate without braking. The supernatant was discarded and each pellet was resuspended in 10 mL ice-cold, filter sterilized $CaCl_2$ (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES, pH 7). The suspension was centrifuged for 5 minutes at 2500 rpm. The supernatant was again discarded and each of the pellets resuspended in 10 mL ice-cold $CaCl_2$. The suspended bacteria were kept on ice for 30 minutes and were centrifuged again for 5 minutes at 2500 rpm. In the last step, the supernatant was discarded and each of the pellets was resuspended in 2 mL ice-cold $CaCl_2$. These competent bacteria were stored in 16-250 µL aliquots in 1.5 mL sterile, polypropylene tubes (Fisher) at -70°C.

EXPRESSION SYSTEMS

A plasmid pRCH (generously donated by Dr. Lee Ratner of Washington University) containing the entire genome from an HTLV-1 strain was used to amplify the entire coding region for the polymerase and the protease. The cDNAs were prepared by Dr. Laura Moen and Barbara Conyers using a Perfect Prep® (5 Prime→ 3 Prime, Boulder, CO) DNA preparation kit. All three expression systems were developed using pET vectors (Novagen). These cloning vectors contain specific antibiotic resistant genes to allow for growth selection on media plates. Both pET24 and pET29 carry the kanamycin resistant gene. pET19 carries the ampicillin resistant gene.

The RT coding sequence was cloned in-frame with the 10-amino acid T7 tag sequence in pET24b which facilitated the detection of the expression product by Western blot. The protease coding sequence was cloned in-frame with an N-terminal polyhistidine tag (His-tag) in pET19, which facilitates purification with an immobilized

nickel affinity column. A Pro-Pol protein clone, which spans a region where the protease is known to terminate and continues past the *pro-pol* frameshift, was cloned in-frame with the N-terminal 16 amino acid S-tag in pET29 which facilitates detection of the expression product by Western blot. The coding sequence also contained a C-terminal His-tag.

Before transformation in *E. coli*, the plasmid DNAs were digested with the appropriate restriction enzymes and analyzed by agarose gel electrophoresis to confirm that the correct plasmids were obtained. Because the DNA is cut at specific sites, a distinct pattern of bands for each cDNA is apparent with ethidium bromide staining. The plasmid DNAs, pET24ab3#12 (RT), IK19 (protease), and peptide (Pro-Pol protein) were stored as glycerol stocks in either XL-Blue or AG1 cells.

ISOLATION OF PLASMID DNA

A BIGGER prep® Plasmid DNA Preparation Kit (5 Prime → 3 Prime, Inc.) was used to prepare plasmid DNAs for the HTLV-1 RT, the HTLV-1 protease and the Pro-Pol protein expression constructs. Unless otherwise stated, all steps in the procedure were performed at room temperature. The DNA stock was streaked on an LB/agar plate supplemented with appropriate antibiotic and grown for 14-16 hours in a 37°C incubator. 25 mg/mL stock kanamycin was made by dissolving 25 mg of kanamycin (Sigma, St. Louis, MO) in 1 mL MilliQ water. If supplemented with kanamycin, the 25 mg/mL stock was added to final concentration of 50 μg/mL in media. To supplement media requiring ampicillin, the appropriate number of 2.5 mg ampicillin tablets (Stratagene, La Jolla, CA) was dissolved directly in media to give a final concentration of 100 μg/mL ampicillin.

The following day, the plate was stored at 4°C until late afternoon, at which time, a single colony was inoculated in 500 mL LB with the appropriate antibiotic and incubated overnight at 37°C with shaking at 250 rpm. The bacteria was pelleted in 2-250 mL centrifuge bottles by centrifugation (Beckman Model J 2-21) at 10,000 x g for 3 minutes. The supernatant was discarded and the plasmid DNA was isolated from the pellet by following the manufacturer's instructions. In brief, the pellet was treated with a buffered RNase A solution, lysed with alkaline lysis solution, and neutralized with

potassium acetate solution. Unless otherwise stated, all solutions were provided in the preparatory kit. The plasmid DNA in the neutralized bacterial lysate was then bound to a DNA binding matrix suspension. The resulting DNA binding matrix/plasmid DNA solution was washed with several aliquots of buffered salt solution followed by elution of the purified plasmid DNA with heated TE solution, pH 8.0. The eluted purified plasmid DNA was then precipitated by the addition of 5 M NaCl and 95% ethanol. The solution was mixed and the plasmid DNA was pelleted by centrifugation at 16,000 x g for 5 minutes. After carefully removing the supernatant, the plasmid DNA pellet was washed twice with 70% ethanol and air dried overnight to remove the ethanol. The next day, the pellet was resuspended in 100 µl sterile water to rinse the side of the tube. The purified plasmid DNA was transferred to a 0.5 mL sterile, siliconized, eppendorf tube (Fisher) and stored at -20°C.

QUANTITATION OF DNA

An aliquot of the DNA sample was diluted either 1:50 or 1:100 with sterile water. The amount of DNA in the samples was determined spectrophotometrically using the Cary UV/VIS Spectrophotometer (Varian Model 3-Bio, Palo Alto, CA). The absorbance of the dilute sample was read at 260 nm and 280 nm. The reading at 260 nm allows for the calculation of the concentration of nucleic acid in the sample. An O.D. = 1 corresponds to approximately 50 µg/mL for double-stranded DNA. Thus, the following equation was developed to calculate the concentration of nucleic acid in the sample:

$$A_{260} \text{ x dilution factor x } 50 \text{ µg/mL} = X \text{ µg/mL DNA}$$

The ratio between the readings at 260 nm to 280 nm provides an estimate for the purity of the nucleic acid. If this ratio was near 1.8, the original DNA sample was stored at -20°C and used for the transformation of plasmid DNA in *E. coli*.

TRANSFORMATION OF COMPETENT *E. coli* WITH PLASMID DNA

Because the competent *E. coli* were stored in $CaCl_2$, they were exposed to calcium ions which allow the cells to take up DNA. The bacterial cell membrane is permeable to chloride ions, but is non-permeable to calcium ions. As the chloride ions

enter the cell, water molecules accompany the charged particle. This influx of water causes the cells to swell and is necessary for the uptake of DNA. The exact mechanism of this uptake is unknown. It is known, however, that the calcium chloride treatment must be followed by heat.

A 250 μL aliquot of competent *E. coli* was allowed to thaw on ice. 100 μL of *E. coli* was transferred to a 14 mL sterile, round bottom, polypropylene tube (Fisher) which was also kept on ice. The bacteria were incubated with 1.7 μl of a 1:10 dilution of β-mercaptoethanol (Fisher) for 10 minutes with gentle swirling every two minutes. Approximately 1 μg of DNA (~ 1-2 μL) was added to the cells and the solution was allowed to incubate on ice for 30 minutes.

When *E. coli* are subjected to 42°C heat, heat shock genes are expressed which aid the bacteria in surviving at such temperatures. The heat shock step is necessary for the uptake of DNA. At temperatures > 42°C, the bacteria's ability to uptake DNA becomes reduced, and at extreme temperatures the bacteria will die. Therefore, in the next step the tube was placed in a 42°C water bath for 45 seconds. The tube was then quickly placed in ice for 2 minutes. 900 μL of sterile SOC media (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) was added to the tube and the tube was incubated for 1 hour at 37°C with shaking at 250 rpm. All of the bacteria were pelleted by centrifugation at 1000 rpm for 10 minutes. The supernatant was gently discarded and the remaining slurry of pelleted *E. coli* DNA was plated on an LB/antibiotic plate using a spreader to evenly distribute the bacteria. The spreader was initially sterilized by dipping the spreader in ethanol, passing the spreader through a gas flame to ignite the ethanol, and allowing the flame to burn out. The spreader was cooled by touching the surface of the LB plate. The plate was incubated overnight in a 37°C incubator.

The next morning the plate containing several isolated colonies was stored at 4°C until ready to be cultured. Ideally, the plates should not be stored for more than one week.

PREPARATION OF BACTERIAL PELLETS

Reverse Transcriptase

Ten flasks containing 500 mL Luria Broth, 50 μg/mL kanamycin were inoculated with a single colony each of W3110(DE3) containing the pET24ab3#12 expression plasmid. The flasks were incubated with shaking at 37°C until the culture reached an $O.D._{600}$ of 0.6-0.9, at which point 800 mM isopropyl-β-D-thiogalactoside (IPTG; Promega, Madison, WI) was added to a final concentration of 1 mM. 2 hours after the addition of IPTG, cultures were pelleted by centrifugation at 10,000 RPM for 20 minutes at 4°C. The cultures were harvested in 250 mL bottles. The supernatant was discarded with care so as not to disturb the bacterial pellet. Two 250 mL pellets were scraped from the bottle and combined into one 500 mL pellet on a tared piece of aluminum foil so that the pellet's wet weight could be measured. Finally, the pellet was wrapped in the aluminum foil, labeled with its wet weight and stored at -70°C. All 10-500 mL pellets were harvested in the same manner.

Protease

A flask containing 100 mL Luria Broth, 100 μg/mL ampicillin and 34 μg/mL chloramphenicol was inoculated with a single colony of BL21pLysS(DE3) containing the protease expression plasmid. The flask was incubated with shaking at 37°C until the culture reached an $O.D._{600}$ of 0.6-0.9, at which point 800 mM IPTG was added to a final concentration of 1 mM. 2 hours after the addition of IPTG, cultures were pelleted by centrifugation at 10,000 RPM for 20 minutes at 4°C. The supernatant was discarded with care so as not to disturb the bacterial pellet. The 100 mL pellet was scraped from the bottle and transferred to a tared 2.0 mL round bottom, self-standing cryovial (Fisher). The cryovial was labeled with the pellet wet weight, frozen in liquid N_2 and stored at -70°C.

Pro-Pol protein

Two flasks containing 500 mL Luria Broth, 50 μg/mL kanamycin and 34 μg/mL chloramphenicol were inoculated with a single colony each of BL21pLysS(DE3) containing the Pro-Pol expression plasmid. The flasks were incubated with shaking at

37°C until the culture reached an O.D.$_{600}$ of 1.0-1.2, at which point 800 mM IPTG was added to a final concentration of 1 mM. 45 minutes after the addition of IPTG, cultures were pelleted by centrifugation at 10,000 RPM for 20 minutes at 4°C. The cultures were harvested in 250 mL bottles. The supernatant was discarded with care so as not to disturb the bacterial cell pellet. Two 250 mL pellets were scraped from the bottle and combined into one 500 mL pellet on a tared piece of aluminum foil so that the pellet's wet weight could be measured. Finally, the pellet was wrapped in the aluminum foil, labeled with its wet weight, frozen in liquid N_2 and stored at -70°C. Both 500 mL pellets were harvested in the same manner.

PREPARATION OF CHROMATOGRAPHIC RESINS

Phosphocellulose (P-11)

To determine the amount of phosphocellulose cation exchanger (Whatman, Maidstone, England) resin to pre-cycle, the approximate bed volume was approximated using the standard formula for the volume of a cylinder:

$$\text{volume} = \pi \times r^2 \times h$$

r = radius of the column

h = height of the column

With a column height = 20 cm and column i.d. = 5 cm, the bed volume would be 392 mL. The amount of dry phophocellulose will swell to double its volume in MilliQ water. Therefore, approximately 200 mL of dry phosphocellulose resin was placed in a 1 L graduated beaker and allowed to swell in 300 mL MilliQ water. After the resin had settled, the water was carefully decanted. Two-500 mL washes of 1 N NaOH (Fisher) were added to the beaker to start the basic pre-cycling. The resin was allowed to settle for 5 minutes between washes. The 1 N NaOH was decanted into a waste beaker. The resin was washed with 250 mL aliquots of MilliQ water until a pH $\leq$ 10 was reached. Each time, the resin was allowed to settle for 5 minutes before decanting.

Acid pre-treatment followed by adding two 500 mL volumes of 1N HCl (Fisher) to the resin. The resin was allowed to settle for 5 minutes and the acid was decanted. The resin was washed with 250 ml aliquots of MilliQ water until a pH $\geq$ 4 was reached.

The resin was washed once more with 500 mL MilliQ water. The final phophocellulose slurry was incubated with 1 L of 10X Reverse Transcriptase Buffer (RTB: 500 mM KCl 500 mM Tris, pH 7.5, 25 mM EDTA, 1% Triton X-100) and stored in a 1 L bottle at 4°C. All reagents in the buffer were purchased from Fisher.

After the phosphocellulose slurry had equilibrated in the 10X RTB for at least two days, the buffer was decanted, and the bottle of phosphocellulose slurry was warmed in a 30°C water bath for 1 hour. Approximately 500 mL of RTB-50 (1X RTB; 50 mM KCl , 50 mM Tris, pH 7.5, 2.5 mM EDTA, 0.1% Triton X-100) was added to the media slurry. Meanwhile, 500 mL RTB-50 was poured into an empty column (BioRad, Hercules, CA) to wash the column and to determine if the stopcock was functioning properly. The column outlet was opened to allow 5% of the buffer to flow out and to remove air from the space beneath the column bed support and the outlet tubing.

After 5% of the buffer had passed through the column, the column was packed with the phosphocellulose slurry. The 1 L bottle of slurry was swirled several times and the slurry was added to the column in one single operation. The resin was allowed to settle by gravity with the column outlet closed. After the bed was formed, the outlet was opened and the excess liquid on the top of the column was allowed to enter the bed, until approximately 1 cm remained above the bed surface. The column was equilibrated in RTB-50 by running the buffer through the column until the pH of the eluate was 7.0.

Sepharose CL-6B

The gel filtration media, sepharose CL-6B (Pharmacia Biotech, Uppsala, Sweden), was resuspended by shaking the bottle. 700 mL of the resuspended gel was transferred to a 1 L flask. The gel was allowed to settle, the excess liquid was decanted and the gel was resuspended in 500 mL RTB-50. The excess buffer was decanted after settling, and 300 mL of RTB-50 was added to the gel to make a slurry The slurry was then de-gassed by bubbling in N_2 for 1 hour.

Meanwhile, 1 L of RTB-50 was poured into the outlet tubing of an empty column (XK 50/60, Pharmacia), until it passed through the bed support net. The column outlet was closed when the dead space under the net was properly filled. The entire slurry was poured into the column in a single operation using a reservoir. The outlet valve was

opened and the packing of the column was initiated by gravity flow. When all the gel had sedimented to the bottom, the reservoir was removed and the adapter was inserted. The flow adapter was adjusted to the surface of the gel bed when the gel was thoroughly packed into the column. The column was equilibrated with 2 L of RTB-50 by running the buffer through the column overnight in a 4°C cold box. The column remained at 4°C for the entire standardization and purification procedures.

Gel filtration molecular weight markers (Sigma) ranging from 12,000 -200,000 Da were used to calibrate the column. The procedure for determining molecular weight outlined in the Sigma bulletin is a modification of the methods of Whitaker (135) and Andrews (2). Molecular weight determinations of unknown proteins are made by comparing the ratio of V_e/V_o for the protein in question to the V_e/V_o of protein standards of known molecular weight (V_e = elution volume;V_o = void volume). A calibration curve can then be prepared by plotting the logarithms of the known molecular weights of protein standards versus their respective V_e/V_o values. The apparent molecular weight of the HTLV-1 protein was calculated from the standard curve (Appendix A) determined by calibration of the column with molecular weight standards: blue dextran (M_r = 2,000,000), carbonic anyhydrase (M_r = 29,000), alcohol dehydrogenase (M_r = 150,000), β-amylase (M_r = 200,000), and cytochrome C (M_r = 12,400).

Table I shows the molecular weights and concentration of markers applied to the column. The final concentrations of each marker, suggested by the manufacturer, would give an A_{280} of approximately 1.0 in the peak fraction. The void volume was determined based on the volume of effluent required for the elution of the large molecule, blue dextran (M_r = 2,000,000). 2 mL of 2 mg/mL blue dextran was carefully applied to the column. The column was connected to an Econo Pump (Model EP-1,BioRad), an Econo UV monitor (Model EM-1, BioRad), an Econo chart recorder (Model 1325, BioRad), and a fraction collector (Model 2110, BioRad). To minimize mixing, all pieces of the apparatus connected to the outlet were placed as close as possible to the outlet tubing. RTB-50 was pumped through the column at a rate of 0.5 mL/min to elute the blue dextran. The chart recorder, set at 200 mV, 1cm/hr, recorded signals from the UV

TABLE I

Preparation of Protein Standards for Sepharose CL-6B Calibration Curve

Protein Standard	Molecular Weight (Daltons)	Concentration (mg/mL)[a]
blue dextran	2,000,000	2
β-amylase	200,000	4
alcohol dehydrogenase	150,000	5
albumin	66,000	10
carbonic anhydrase	29,000	3
cytochrome C	12,400	2

[a]Individual protein standards were dissolved in RTB-50 containing 5 % glycerol. 2 mL of individual samples containing the following concentrations, as mg solid per mL of buffer, should give an $A_{280} = 1.0$ in the peak fraction.

monitor programmed to read absorbances at 280 nm. The fraction collector was set at 8 mL/tube. This void volume was calculated using the following formula:

cm peak **x** hr/cm (chart speed) **x** 60 min/hr **x** 0.5mL/min (pump speed) = volume (mL) migrated

To confirm the void volume, the absorbance of the tubes containing the peak fractions and the tubes located before and after the peak were read again manually at 280 nm using the department's Cary UV/Vis Spectrophotometer.

Each of the protein standards was applied as described. The column was allowed to equilibrate in 1 L of RTB-50 at 0.5 mL/min between each run. The elution volumes of the standards were denoted and the calibration curve (Appendix A) was graphed using CA Cricket Graph III Software (Computer Associates).

PURIFICATION OF HTLV-1 REVERSE TRANSCRIPTASE

Unless otherwise indicated, all steps were carried out at 4°C. Each 500 mL bacterial pellet was resuspended in 20 mL of STET buffer (50 mM Tris; 8% sucrose; 50 mM EDTA and 5% Triton X-100) supplemented with 10 mg/mL lysozyme and a protease inhibitor cocktail (Boehringer Mannheim, Atlanta, GA) containing aprotinin, leupeptin, pepstatin, phenylmethylsulfonylfluoride (PMSF), antipain, chymostatin, peflabloc and E-64. Stock solutions were prepared according to Table II. The bacterial pellet in STET solution was kept on ice for 1 hour. The thawed pellet was vortexed vigorously and DNA was sheared with a sonic dismembrator (Fisher Sonic Dismembrator 60) using 4-15 second pulses. Finally, the suspension was cleared by centrifugation for 30 minutes at 12,000 RPM to remove insoluble debris. The supernatant from each 500 mL pellet was combined to make the 200 mL crude lysate.

Phosphocellulose chromatography of HTLV-1 RT

The crude lysate was applied at a flow rate of 0.5 mL/min to a phosphocellulose column (20 cm x 5.0 cm i.d.) equilibrated in RTB-50. The column was eluted with an 800 mL linear gradient of 0.1 to 1 M KCl in RTB-50 supplemented with 10% glycerol (glycerol), 1 mM PMSF and 2 mM dithiothreitol (DTT, Fisher). Fractions of 8 mL were collected and assayed for reverse transcriptase activity as described later in this section.

TABLE II

Preparation of Protease Inhibitor Cocktail

Protease Inhibitor	Concentration of Stock Solution[a]	Diluent Buffer	Concentration in STET Buffer
antipain	50 μM	MilliQ water	0.1 μM
aprotinin	76 μM	MilliQ water	0.1 μM
chymostatin	500 μM	glacial acetic acid	1 μM
E-64	1mM	1:1 H_2O/100 % ethanol	1 μM
leupeptin	1mM	MilliQ water	1 μM
peflabloc	84 mM	MilliQ water	0.4 mM
pepstatin	700 μM	methanol	1 μM
PMSF	100 mM	100 % ethanol	1 mM

[a]All stock solutions were stored as 200-300 μL aliquots in 1.5 mL sterile, eppendorf tubes at -20°C.

Ammonium Sulfate Precipitation

The fractions containing the majority of enzyme activity were pooled in a 100 mL beaker, which was then placed on a stir plate in the 4°C cold box. The pool was precipitated with saturated ammonium sulfate (Fisher) to 65 % saturation (0.3432 g/mL; a saturated solution = 4 M $(NH_4)_2SO_4$ = 528 g/L = 0.528 g/mL). The solution was mixed with a magnetic stir bar until all of the ammonium sulfate was dissolved. The pool was transferred to a centrifuge tube and centrifuged at 10,000 RPM for 20 minutes. The supernatant was discarded and the pellet was resuspended in 1.5-2.0 mL RTB-50. The suspension was placed in 12,000-14,000 MW cut-off Spectra/Por® dialysis tubing (Fisher) and dialyzed against 1 L RTB-50 for 4 hours. The buffer was changed at 2 hours. The sample was removed from the dialysis tubing and centrifuged again at 40,000 RPM for 20 minutes to remove cellular debris.

Sepharose CL-6B chromatography

The supernatant was loaded on a column (52 cm x 5 cm i.d.) of Sepharose CL-6B, equilibrated in RTB-50. The column was developed using the parameters predetermined in the calibration procedure. The apparent molecular weight was computed from the elution volume of the peak absorbance and the equation for the linear regression line of the standard curve was determined by calibration of the column with molecular weight standards: blue dextran (M_r = 2,000,000), carbonic anyhydrase (M_r = 29,000), alcohol dehydrogenase (M_r = 150,000), β-amylase (M_r = 200,000), and cytochrome C (M_r = 12,400). The peak fractions were also assayed for reverse transcriptase activity as described later in the next section.

REVERSE TRANSCRIPTASE ASSAYS

10 µL of each fraction was pipetted into 40 µL of reaction cocktail (final concentration: 50 mM Tris-HCl (pH 8.8), 80 mM KCl, 6 mM $MgCl_2$, 8 mM DTT, 0.05% Triton X-100, 80 mM dGTP or dTTP (Promega), 1 mM [^{3}H]dGTP (30 Ci/mmol) or [^{3}H]dTTP (104 Ci/mmol)(Amersham, Arlington Heights, IL) and 20 µg/mL poly(rC)•oligo(dG) template or poly(rA)•oligo(dT) template (Pharmacia), respectively.

Two control reaction mixtures were also run with each set of assays. . The positive control contained 10 μL AMV RT and the negative control contained 10 μL sterile water. The controls were included as samples for each run, and all samples were run in triplicate. Using a timer, start times were staggered 15 seconds apart to give enough time to pipet each sample correctly. The reaction mixture was incubated for 1 hour in a 37°C water bath. The reactions were terminated by the addition of 800 μL of 10% trichloroacetic acid (TCA, Fisher). Acid precipitable material was collected on Whatman GF/C fiber discs, washed with 10 mL of cold 10% TCA and a final wash of 2 mL of 95% ethanol. The filters were loaded into scintillation vials (Fisher) , 6 mL liquid scintillant (ICN Biochemical, Costa Mesa, CA) was added to each vial, and the samples were counted in a scintillation counter (Beckman Model LS1701). Background counts due to the reaction cocktail were determined by counting 10 μL of reaction cocktail in 6 mL liquid scintillation cocktail. The following formula was developed to determine the units of specific activity associated with each sample:

$$\frac{\text{CPM of sample}}{\text{CPM of 10 μl of reaction cocktail}} \times (1.5 \times 10^{3}) \text{ pmoles} = \text{pmoles activity in the sample}$$

There are 1.5×10^{3} pmoles of [^{3}H]dNTP/10 μL of reaction cocktail.

DETERMINATION OF PROTEIN CONCENTRATION

Protein concentration was determined by the Bradford assay technique using Pierce Protein Reagent (Rockford, IL) (14). Bovine serum albumin (BSA) was used as the standard. To determine the protein concentration of reverse transcriptase purification samples, protease and Pro-Pol protein purification samples, the 2 mg/mL BSA stock solution was used to make a 200 μg/mL BSA working solution in RTB-50 or citrate buffer (100 mM sodium citrate, pH 5.3, 5 mM EDTA, 1 M NaCl, and 1 mM DTT), respectively. The 200 μg/mL solution was added to a series of glass, disposable, borosilicate test tubes (Fisher) containing MilliQ water to make standards within a 1-25 μg range. The final volume of each of these tubes was 1 mL. Simultaneously, 10-200 μl

of each purification sample was added to a set of tubes filled with MilliQ water to a final solution volume of 1 mL. To determine the background due to buffer, a control sample of 10-200 µL of RTB-50 or citrate buffer was added to MilliQ water to a final volume of 1 mL. 1 mL of protein assay reagent was added to all of the standards and the samples. The solutions were mixed carefully, and each sample was transferred to a disposable, plastic cuvette (Fisher). The absorbance of each standard was read at 595 nm where these values were used to create a standard curve using CA Cricket Graph III Software. The absorbance due to the buffer was subtracted from the absorbance of each sample read at 595 nm. The difference was applied to the equation for the linear regression line of the standard curve to determine the concentration of the protein alone.

PREPARATION OF SAMPLES FOR SDS-PAGE

To prepare samples for electrophoresis, dilute samples were concentrated by TCA precipitation. An aliquot of the fractions and/or pools was placed in a 1.5 mL eppendorf tube. The sample was precipitated with a 100% solution of TCA to a final concentration of 10% TCA. The tube was vortexed and centrifuged for 5 minutes at 12,000 rpm. The supernatant was decanted into a waste beaker and the pellet was washed three times with 1 mL 95% ethanol. The tube was centrifuged for 5 minutes for each wash and the supernatant was discarded as before. The pellet was dried using the Speed Vac and resuspended in 20 µL Tricine SDS sample buffer (450 mM Tris HCl, pH 8.45, 12% glycerol, 4% SDS, 0.0025% Coomassie Blue G and 0.0025% Phenol Red) or NuPAGE™ SDS sample buffer (0.293 sucrose, 141 mM Tris Base, 106 mM Tris HCl, pH 8.5, 69.5 mM SDS, 0.51 mM EDTA, 0.22 mM Serva Blue G250 and 0.175 Phenol Red). 100 µL of β-mercaptoethanol was added to 1 mL of either sample buffer prior to use. The sample was then heated in a 90-100°C heat block for 10 minutes. For concentrated samples, samples were mixed 1:1 with sample buffer and heated.

Additionally, the T7 tag positive control (Novagen) was prepared by adding 5 µL control sample to 95 µL sample buffer (1:20 dilution). DNA Polymerase I (Promega) was prepared by preparing a 1:100 dilution of enzyme in sample buffer. 15 µL of these particular samples was loaded on the SDS-PAGE where indicated.

SDS-PAGE

Pools from each step of the HTLV-1 RT purification and fractions from the partial purifications of protease and Pro-Pol protein were analyzed by SDS-PAGE. SDS-PAGE was performed as described by Laemmli (59) or by a modification of the procedure described by Schaegger and von Jagow (113). 16% and 18 % acrylamide, Tricine gels were prepared according to Table III. Where indicated, some of the SDS-PAGE analyses were performed using pre-cast 10-20% acrylamide Tricine gels from Novex (San Diego, CA).

The Tricine running buffer was prepared as a 10X stock solution and stored at room temperature. 80 mL of the 10X solution was diluted with 720 mL MilliQ water to make a 1X Tricine running buffer solution (100 mM Tris, pH 8.3, 100 mM Tricine and 0.1 % SDS) prior to each run. The acrylamide gel(s) was placed in the Xcell II™ Mini-Cell (Novex) apparatus and the 1X Tricine running buffer was poured in both the upper and lower chambers of the gel box. The gel was run at a 125 constant voltage for approximately 90 minutes using the PowerPac200 (BioRad) power source.

A pre-cast NuPAGE ™ 4-12% Bis-Tris Gel (Novex) with 3-(N-morpholino)propane sulfonic acid (MOPS) running buffer was used in the final analysis of the HTLV-1 RT purification. The running buffer was prepared by adding 40 mL of 20X stock MOPS buffer (Novex) to 760 mL MilliQ water to give a final concentration of 50 mM MOPS, 50 mM Tris, pH 7.7, 3.5 mM SDS and 1 mM EDTA. 500 µL of NuPAGE™ antioxidant (Novex) was added to 200 mL of the running buffer. This solution was poured into the upper chamber of the gel box. The remaining 600 mL of running buffer was poured into the lower chamber. The gel was run at a constant 200V for 50 minutes.

After any type of run was complete, the gel was transferred to a clean glass container and pre-washed in 200 mL MilliQ water for 1 hour at room temperature, with rocking. The gel was then transferred to another tray filled with 20 mL Gelcode® Blue Stain Reagent (Pierce, Rockford, IL) and incubated for 1 hour, with gently rocking. Finally, the gel was transferred to another clean glass container and destained in 200 mL MilliQ water overnight.

TABLE III

Preparation of Separating and Stacking Gels for SDS-PAGE Electrophoresis

Solution	16 % Acrylamide Separating Gel	18 % Acrylamide Separating Gel	4 % Acrylamide Stacking Gel
50% Acrylamide/BIS (29:1)[a]	2.0 mL	2.25 mL	250 µL
1M Tris-HCl, pH 8.8[a]	2.35 mL	2.35 mL	---
10 % SDS[b]	62.5 µL	62.5 µL	---
0.375 M Tris-HCl, pH 6.8[a]	---	---	1 mL
TEMED	1.6 µL	1.6 µL	1.2 µL
50 mg/mL Ammonium Persulfate[c]	156 µL	156 µL	250 µL
50% Sucrose[b]	1 mL	1 mL	---

[a]Acrylamide and Tris-HCl solutions were made as stock solutions and stored at 4°C.
[b]SDS and sucrose were also made as stock solutions and stored at room temperature.
[c]Ammonium persulfate was prepared fresh for each gel.

WESTERN BLOTTING

Unless otherwise stated, all steps in the Western blot development procedure were performed at room temperature. A second SDS-PAGE gel for transfer was run simultaneously with the SDS-PAGE gel that was stained. After electrophoresis, proteins in the second gel were transferred to a polyvinylidene difluoride (PVDF, BioRad) membrane using a Trans-blot® SD Semi-Dry Electorphoretic Transfer Cell (BioRad). A piece of PVDF membrane was cut to approximately the same dimensions of the gel. The membrane was dipped in methanol and then equilibrated in Western transfer buffer (25 mM Tris Base, 192 mM glycine and 30% methanol) for 30 minutes, with rocking. The acrylamide gel was also equilibrated in buffer for 30 minutes in a separate container.

The gel sandwich was prepared for transfer by placing a pre-wet (with transfer buffer) piece of thick blot absorbant filter paper (BioRad) on the platinum anode. A disposable pipet was rolled over the surface of the filter paper to exclude all air bubbles. The pre-wetted PVDF membrane was placed on top of the filter paper. Bubbles were removed with the pipet. The equilibrated gel was placed carefully on top of the membrane, aligning the gel in the center of the membrane. Another sheet of pre-soaked filter paper was placed on top of the gel and air bubbles were removed again. The cathode was placed on top of the stack and the proteins on the gel were transferred to the membrane at 18V for 50 minutes.

The sandwich was taken apart, layer by layer. If all of the molecular weight markers (low molecular weight, BioRad) transferred to the membrane, the blot (membrane) was washed with Tris Buffered Saline (TBS, 20 mM Tris, pH 7.5 and 500 mM NaCl) for 5 minutes. To probe for HTLV-1 RT, the wash was removed from the container and the blot was blocked by incubating the blot with 40 mL of 3% gelatin/TBS solution for one hour, with rocking. The blocking solution was removed and the blot was washed again with Tween TBS (TTBS; TBS + 0.05% Tween-20) for 5 minutes. The wash was again removed and the blot was placed in 40 mL 1% gelatin/TTBS with 5 μL of T7 tag antibody (Novagen, 1:10,000 dilution) directed against the 10 amino acid gene leader sequence peptide expressed by the pET translation vector and incubated overnight with gentle rocking.

The next morning, the antibody solution was removed and the blot was washed twice with TTBS, 5 minutes each wash. The washes were removed and the blot was incubated with the second antibody solution (40 mL TTBS with 13.2 µL Goat Anti-Mouse IgG (H + L) alkaline phosphatase conjugate, 1:3000 dilution, BioRad) for one hour, with rocking. The solution was removed and the blot washed twice with TTBS and once with TBS, 5 minutes each wash. Detection of the antigen was performed using an ImmunoBlot assay kit from BioRad. 300 µL of Color Reagent A (NBT, nitroblue tetrazolium in aqueous dimethlyformamide, containing magnesium chloride) and 300 µL Color Reagent B (BCIP; 5-bromo-4-chloro-3 indoyl phosphate in dimethylformamide) were added to 30 mL alkaline phosphatase color development buffer. After the last TBS wash was removed, the blot was developed for 3 hours using the prepared alkaline phosphatase color development buffer. Purple bands stained on the blot represent the reverse transcriptase protein in the pET translation vector.

The Pro-Pol protein was transferred using the same methods, however, only one antibody incubation was required. After transferring the gel, the blot was blocked by incubating the blot with 40 mL of 1% gelatin/TTBS solution for 15 minutes, with rocking. The blocking solution was removed and the blot was placed in 25 mL TBS with 5 µL of S-protein alkaline phosphatase conjugate antibody (Novagen, 1:5,000 dilution) directed against the 15 amino acid S-Tag leader sequence peptide expressed by the pET translation vector and incubated for 30 minutes with gentle rocking.

The antibody solution was removed and the blot was thoroughly washed 5 times with 25 mL TTBS, 1-2 minutes each wash. Detection of the antigen was performed using an ImmunoBlot assay kit from BioRad. 332 µL of Color Reagent A (NBT, nitroblue tetrazolium in aqueous dimethlyformamide, containing magnesium chloride) and 336 µL Color Reagent B (BCIP, 5-bromo-4-chloro-3 indoyl phosphate in dimethylformamide) were added to 30 mL alkaline phosphatase color development buffer. After the washes were removed, the blot was developed for 15 minutes using the prepared alkaline phosphatase color development buffer. Purple bands stained on the blot represent the Pro-Pol protein in the pET translation vector.

Color development buffer was removed and the reactions were stopped by rinsing the blots with deionized water for 10 minutes, changing the water every 5 minutes. The blots were removed from the water and placed on paper towels to dry overnight. The next morning, the blot was wrapped in plastic wrap and taped into a laboratory notebook for future analysis.

PARTIAL PURIFICATION OF PRO-POL PROTEIN

The two 500 mL bacterial pellets were combined in a 50 mL conical centrifuge tube and resuspended in 6 mL of lysis buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 5 mM imidazole) supplemented with 0.1% Triton X-100 and 1mM PMSF. The solution was kept on ice for 1 hour. The thawed pellet was vortexed vigorously and DNA was sheared with a sonic dismembrator with 6-10 second pulses and 10 second cooling between pulses. The suspension was cleared by centrifugation for 20 minutes at 10,000 RPM at 4°C to remove insoluble debris. The crude supernatant was used for the subsequent purification steps.

Ni^{+2}-NTA Affinity Chromatography

Unless otherwise stated, all steps were carried out at room temperature. The crude supernatant was incubated, with rocking, with 1 mL of Ni^{+2}-NTA slurry (Qiagen, Valencia, CA). After one hour, the entire solution was poured in a column (8 cm x 1.5 cm i.d.) and the pass fraction was collected. The column was eluted with a stepwise gradient of 5 mL 10 mM imidazole, 5 mL 50 mM imidazole, four 1.25 mL 100 mM imidazole washes and 2 mL 500 mM imidazole. Each imidazole wash was prepared in buffer containing 150 mM NaCl and 50 mM Tris, pH 8.0). The fractions were analyzed by SDS-PAGE. The fractions containing the majority of the protein were pooled and dialyzed against 500 mL lysis buffer (without PMSF and Triton X-100) for four hours, where the buffer was changed at the 2 hour interval.

S-tag agarose chromatography

The entire protein pool from the Ni^{+2} purification step was incubated, with rocking, with 1 mL S-tag agarose beads (Novagen) for 30 minutes in a 14 mL round bottom, polystyrene tube. Recovery of S-tagged peptide was achieved by following the

manufacturer's protocol (Novagen). The pass fraction was collected by centrifuging the tube, containing the agarose beads and protein, at 500 x g for 10 minutes. The beads were allowed to settle and the supernatant was carefully removed with a pipet so as not to disturb the beads. The supernatant was transferred to a test tube and labeled as the pass fraction. The beads were then washed by resuspending the agarose beads in 5 mL bind/wash buffer (50 mM NaCl, 20 mM Tris, pH 8.0, and 0.1% Triton X-100). The mixture was mixed by repeated inversion, followed by centrifugation at 500 x g for 10 minutes. The supernatant was removed with a pipet and transferred to separate test tubes. The beads were resuspended twice more in the same manner, and the supernatants were saved as low salt washes.

To elute the Pro-Pol protein, the S-protein agarose beads were resuspended in 1.5 mL 3 M $MgCl_2$ and the mixture was incubated for 10 minutes, with gentle rocking. A 5-mL capacity spin filter, provided by the manufacturer, was placed in a 15 mL sterile, centrifuge tube. The incubated mixture was poured into this spin column. The bound protein was dissociated from the agarose by centrifuging the spin column/centrifuge tube assembly at 500 x g for 5 minutes. An additional 1.25 mL of 3 M $MgCl_2$ was added to the spin filter. The assembly was centrifuged again at 500 x g for 5 minutes. A 100 µL aliquot of the filtrate was saved for SDS-PAGE analysis while the remainder of the filtrate was dialyzed against 500 mL Bind/Wash buffer for 4 hours using 7,000 MWCO dialysis tubing (Pierce). The buffer was changed at two hours. The dialysate was then concentrated 10-fold using an 3,000 MWCO Centricon-3 microconcentrator (Amicon, Danvers, MA). The concentrator was placed in a 14 mL centrifuge tube. The dialysate was placed in the concentrator/centrifuge tube assembly and centrifuged at 6500 RPM for 3 hours. The final volume of the concentrate was 300 µL. Fractions from each step of the purification scheme were analyzed by SDS-PAGE and Western blotting.

EXPRESSION OF ACTIVE PROTEASE

The removal of inclusion bodies and the solubilization of the protease was performed using a modified protocol of a published procedure (19). Briefly, the bacterial pellet was resuspended in 1 mL of buffer A (500 mM NaCl, 20 mM Tris, pH 8.0, 5 mM imidazole).

The solution was kept on ice for 1 hour. The thawed pellet was vortexed vigorously and the suspension was centrifuged at 10,000 RPM for 25 minutes. The insoluble fraction was resuspended in another 1 mL of buffer A, vortexed vigorously and centrifuged again to further remove inclusion bodies. The insoluble fraction was resuspended in 2 mL buffer B (buffer A + 8 M urea) and the DNA was sheared with a sonic dismembrator using 6 -10 second pulses with a 10 second cooling period between pulses. Finally, the suspension was cleared by centrifugation for 20 minutes at 12,000 RPM at 4°C to remove insoluble debris. The supernatants from the first two centrifugation steps were saved for SDS-PAGE analysis. The final crude supernatant was used for the subsequent purification steps.

Ni^{+2}-NTA affinity chromatography

Unless otherwise stated, all steps were carried out at room temperature. The crude supernatant was incubated, with rocking, with 1 mL of Ni^{+2}-NTA slurry. After one hour, the entire solution was poured in a column (8 cm x 1.5 cm i.d.) and the pass fraction was collected. The column was developed with a stepwise gradient of four 1 mL 20 mM imidazole washes and the protein was eluted with 2 mL of 1 M imidazole. The imidazole washes were prepared in buffer containing 500 mM NaCl, 20 mM Tris, pH 8.0, 8 M urea.

Refolding and activation of protease

A 100 µL aliquot of the eluate fraction was saved for SDS-PAGE analysis. To refold and activate the protease in the eluate, a sequential dialysis was performed. The protease was dialyzed against 500 mL acetate buffer (10 mM sodium acetate, pH 3.5 and 1 mM DTT) for 4 hours using a 3 mL, 7,000 MWCO Slide-A-Lyzer® dialysis cassette from Pierce. The buffer was changed after 4 hours, and the protein was allowed to dialyze overnight for another 12 hours in acetate buffer. The next morning, the dialysis cassette was placed in 500 mL citrate buffer (100 mM sodium citrate, pH 5.3, 5 mM EDTA, 1 M NaCl and 1 mM DTT). The buffer was changed at the end of the day and the protein was again dialyzed overnight. The total acetate dialysis time was approximately 24 hours.

Protease activity assay

200 µL of refolded protease was pipetted in a 10 mm quartz cuvette (Starna Cells) and

incubated for 10 minutes at 37°C in a thermostatically adjusted Luminescence Spectrophotometer (Perkin Elmer, Model LS50B). The cuvette was removed from the cell chamber and the cleavage reaction was initiated by the addition of 480 µL 100 mM sodium citrate buffer and 5 µL substrate(15 mM). The reaction was mixed by gentle inversion and the cuvette was replaced in the cell chamber. Hydrolysis of the synthetic substrate Abz-KTKVLVVQPK-(3NO_2Y)A (custom made at the University of Michigan), which contains the cleavage site of HTLV-1 p24/p15 between the underlined leucine and valine, was characterized by measuring the fluorescence emission spectra of the sample at 450 nm/2.5 nm bandpass at an excitation of 320 nm/15 nm bandpass. Data points were collected over a period of 30 minutes using FL WinLab Software (Perkin Elmer, Norwalk, CT).

IDENTIFICATION OF THE PRO-POL PROTEIN PROCESSING SITE

100 µL of activated protease in citrate buffer (0.05 µg/µL) was incubated with 200 µL Pro-Pol protein (0.67 µg/µL) overnight in a 37°C heat block to cleave the Pro-Pol protein in lysis buffer. Control samples, 100 µL protease in citrate buffer only and 200 µL Pro-Pol protein only, were also incubated overnight at 37°C.

Hydrolysis of the substrate, which contains the putative proteolytic processing site at the N-terminus of HTLV-1 RT was analyzed by SDS-PAGE and Western Blotting. The cleavage reaction samples were separated by SDS-PAGE using either an 18% acrylamide Tricine gel or 10-20% acrylamide gradient Tricine gel. One gel was transferred to a PVDF membrane for Western blotting as described previously.

All steps in the following staining procedure were performed at room temperature. The second gel was fixed in 100 mL fixing solution consisting of 50% methanol, 10% acetic acid and 40% MilliQ water for 10 minutes, with rocking. The gel was then stained using solutions from a Colloidal Blue Coomassie™ Staining Kit (Novex). The gel was placed in staining solution containing 55 mL MilliQ water, 20 mL methanol and 20 mL Stainer A (ammonium sulfate and phosphoric acid) for 10 minutes. 5 mL of Stainer B was added to the staining solution and the gel was allowed to rock for 3 hours. The gel

was then transferred to a glass tray container with 200 mL MilliQ water and allowed to destain overnight.

The entire gel was sealed in a heat-sealable bag. The gel and a Polaroid picture of the gel denoting the band of interest was sent to the W.M. Keck Biomedical Mass Spectrometry Laboratory, Biomolecular Research Facility, University of Virginia, Charlottesville, Virginia. Dr. Nicholas E. Sherman cut the band out of the gel and digested the sample with trypsin. Extracts were combined and evaporated to 25 μL for LC-MS analysis.

Mass analysis and sequence analysis

The masses of the peptides of interest from the enzymatic digest were determined by LC-MS using a Finnigan LCQ ion trap mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 μm i.d. Phenomenex Jupiter 10 μm C18 reversed-phase capillary column. 0.5-5 mL volumes of the extract were injected and the peptides eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.25 μL/min. The nanospray source was operated at 2.8 kV. The digest was analyzed using the double play capability of the instrument to obtain full scan mass spectra and product ion spectra in sequential scans. In this manner, both peptide molecular weight and the amino acid sequence was determined. The data was analyzed by database searching using the SEQUEST search algorithm. Peptides that were not matched by this algorithm were interpreted manually and searched versus the EST databases using the SEQUEST algorithm.

CHAPTER III

RESULTS: PURIFICATION OF RECOMBINANT HTLV-1 RT

Detailed analysis of the properties of HTLV-1 RT have been limited by the small quantity of enzyme which can be isolated from retroviral virions. An alternate means to prepare the enzyme is to express the *pol* gene in *E. coli*, allowing overproduction of the protein and facilitating further characterization. Two clones of HTLV-1 RT were constructed by Dr. Laura Moen (Fig. 5) in our laboratory. Expression of the full length clone in *E. coli* strain W3110(DE3) was successful and this protein was used for the work described in subsequent sections.

The experimental approach to isolate the recombinant HTLV-1 RT utilized adsorption chromatography, a common feature for most polymerase purification procedures. Proteins bind to the exchanger resins by electrostatic forces. To determine the optimal conditions for adsorption, the stability range of the HTLV-1 RT, pH = 7.0-8.0 and its isoelectric point (p*I*) = 9.0 were considered. Scientists suggest using a cation exchanger for polymerases that are stable at neutral pH ranges and an anion exchanger for polymerases that are more stable at basic pH ranges. For a pI ≈ 8.5, a cation exchanger is also recommended (106, 112). Since the HTLV-1 RT has an optimum pH within a neutral and basic pH range, both a cation exchanger (phosphocellulose; P-11) and an anion exchanger (diethylaminoethyl cellulose; DEAE) were investigated.

Purification schemes of other reverse transcriptases (HIV-1 RT, BLV RT, AMV RT and MMTV) were also analyzed to determine a comparable purification scheme for recombinant HTLV-1 RT. Phenyl sepharose, heparin sepharose and DNA cellulose affinity columns and sepharose gel filtration media were also investigated for further optimization of purification parameters.. All of these techniques for purification of polymerases utilize a different property of interaction with the proteins to facilitate isolation. Thus, presumptions can be made as to the amino acid composition of the active site of the enzyme based on the information obtained from each application.

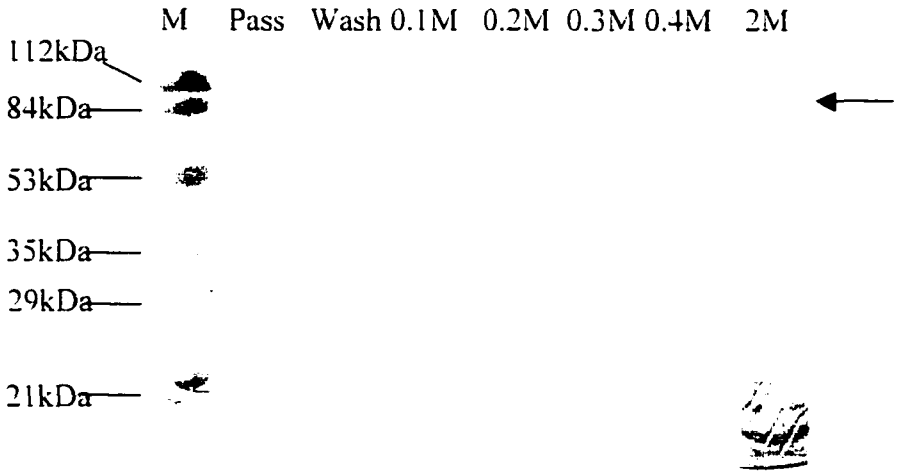
PRELIMINARY TRIALS

Initially, all trials with each type of resin were performed using small columns (1 mL -50 mL), stepwise gradients and 0.5-1 L bacterial cell pellets. As shown in Appendix B, the DEAE cellulose anion exchanger did not fare well with our protein. SDS-PAGE analysis and Western blotting indicate that the protein did not stick to the column. The majority of the RT activity was found in the pass fraction. SDS-PAGE analysis of a phosphocellulose (P-11) purification of the enzyme shows a much cleaner preparation compared to the DEAE purification (Fig. 7). Consequently, the DEAE column was not considered for the purification scheme.

SDS-PAGE analysis of the P-11 purification further indicates that the protein eluted at 0.2-0.3 M KCl using a stepwise gradient, however, the fractions were not pure. Further purification of the protein was attempted by utilizing either a heparin sepharose CL-6B, phenyl sepharose CL-4B or DNA cellulose. Each resin was prepared according to the manufacturer's instructions and poured as a media slurry in a small column (3-5cm x 1.5 cm i.d.). Either the 0.2 M or 0.3 M KCl fraction from the P-11 column was applied to the column.

Of the three resins that were investigated, the heparin sepharose gave the best results. As shown in Fig. 8A, a high molecular weight protein in agreement with the predicted size of the protein eluted in both the 0.1 M and 0.5 M KCl fractions. Western blot analysis demonstrates that the HTLV-1 RT appears only in the 0.5 M KCl fraction (Fig. 8B). Earlier efforts to use the heparin sepharose column indicated that the high molecular weight protein also eluted in the 0.5 M KCl fraction, and this protein was more pure than the 0.3 M KCl fraction from the P-11 step. However, the protein did not appear in the Western blot, indicating that the high molecular weight protein was not the HTLV-1 RT. To clarify this inconsistency and to determine the concentration at which the HTLV-1 RT elutes, a linear gradient of 0.1-1.0 M KCl was applied to the column instead of the usual stepwise gradient. Several different peaks of RT activity were observed in all of the preparations. Analysis of these peaks by SDS-PAGE and Western blotting proved inconclusive as no consistency in gel band patterns could be established. Also, the high molecular weight protein did not consistently appear on the Western blot.

A



B

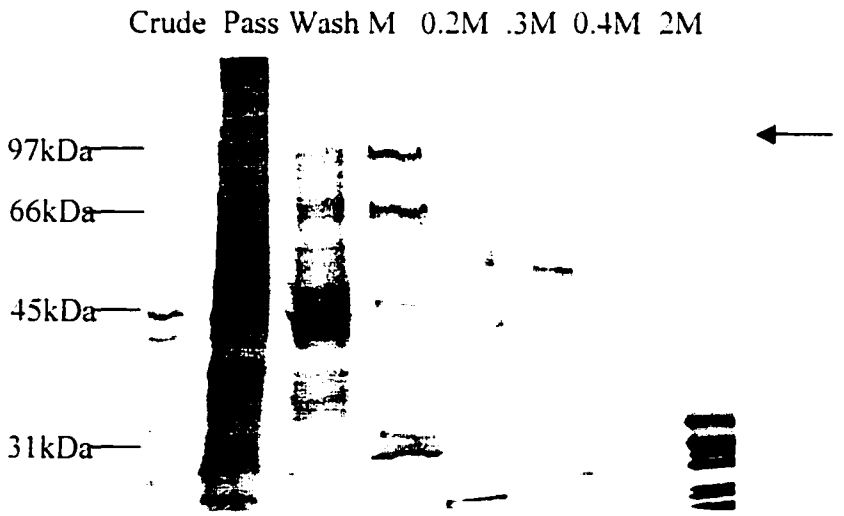
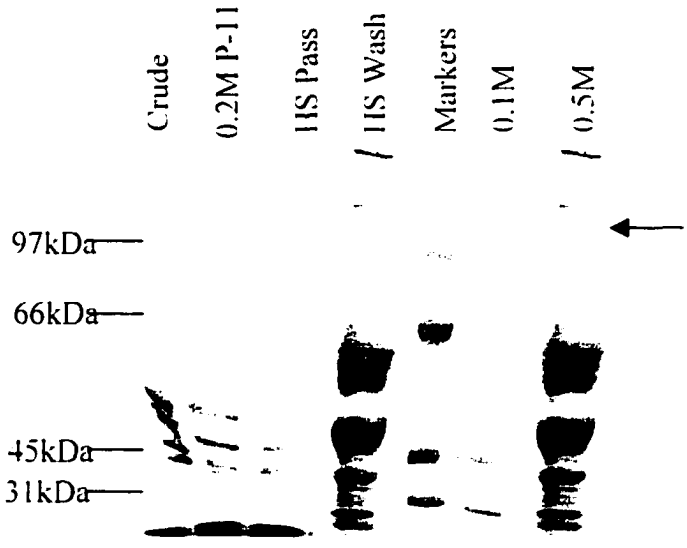


FIG. 7. SDS-PAGE analysis of a phosphocellulose (P-11) purification and a DEAE purification. Crude lysate is compared with eluate from each stepwise elution for P-11 (A) or DEAE (B). The arrow points to recombinant HTLV1-RT.

A



B

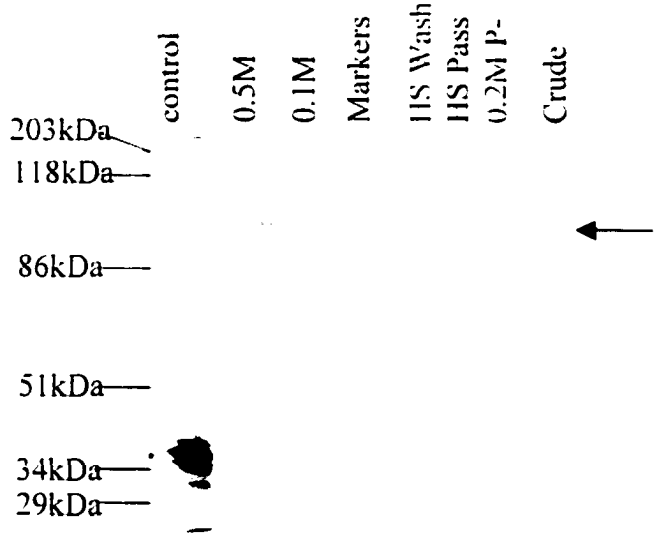


FIG. 8. SDS-PAGE analysis of heparin sepharose column fractions. Fractions were subjected to electrophoresis on 10% acrylamide gels and either (A) stained with Coomassie Brilliant Blue or (B) transferred to a PVDF membrane for Western blotting. The arrow points to the recombinant HTLV1-RT. Coomassie Brilliant Blue or (B) transferred to a PVDF membrane for Western blotting. The arrow points to the recombinant HTLV1-RT.

The phenyl sepharose column served two purposes in these preliminary trials: (1) to further purify the HTLV-1 RT, and (2) to determine if there were any hydrophobic regions in the polymerase. Coomassie stained SDS-PAGE gels of the phenyl sepharose fractions indicated that a protein of M_r = 99,000 was present in the 0.6 M- 0 M NH_4SO_4 fractions yet the protein did not light up on the Western blot (Fig. 9). After repeated cell preparations and similar results, it was concluded that the protein may have bound to the resin yet becomes too dilute throughout the elution process such that it does not appear on the blot. Another *E. coli* protein may have co-eluted with our protein making the 99,000 Da band appear concentrated. To evaluate this hypothesis, another preparation was run whereby the protein, if present, was eluted off the column with one 0 M NH_4SO_4 elution. This pool was concentrated and analyzed by SDS-PAGE and Western blotting. The protein band still did not appear in the concentrated fraction. Thus, use of the phenyl sepharose column as a second step for purification of the enzyme was terminated.

The SDS-PAGE analyses of the DNA cellulose fractions did not give any encouraging results. Most gels showed that the protein did not stick to the column as the protein was found in the pass fraction. Figure 10 also illustrates that the pass fraction is not much cleaner than the 0.2 M KCl fraction from the P-11 step.

After performing several preparations, the resin that would optimize purification of the P-11 fraction/pool was found to be the gel filtration media, sepharose CL-6B (Pharmacia). The overall purification scheme, depicted in Fig. 11, requires a 5 L bacterial culture preparation containing the expression plasmid. The protein was eluted from a 400 mL P-11 column using a linear gradient . The column volumes were increased 8-fold from the original 50 mL column in order to recover a sufficient amount of protein for biochemical characterization.

PURIFICATION OF THE RECOMBINANT PROTEIN

The purification procedure, summarized in Table IV, involves only two steps with minimum inactivation of enzyme activity during purification. This procedure results in the purification of the recombinant protein to approximately 90% homogeneity (Fig. 12).

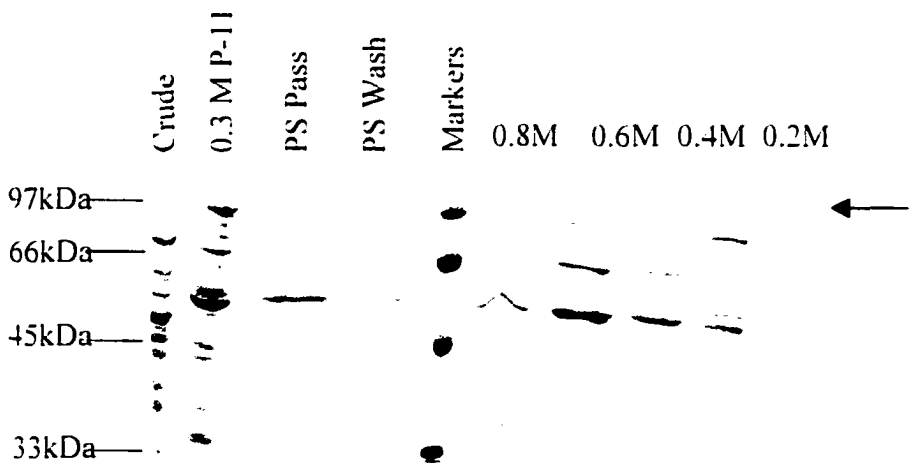


FIG. 9. SDS-PAGE analysis of phenyl sepharose column fractions. Fractions were subjected to electrophoresis on 10% acrylamide gels and stained with Coomassie Brilliant Blue. The arrow points to the recombinant HTLV1-RT.

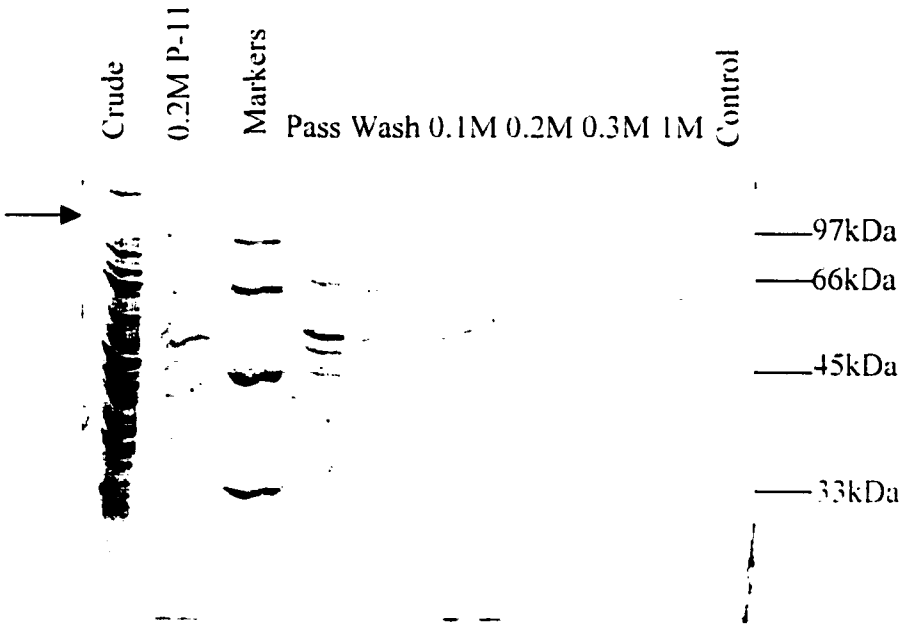


FIG. 10. SDS-PAGE analysis of DNA cellulose column fractions. Fractions were subjected to electrophoresis on 10% acrylamide gels and stained with Coomassie Brilliant Blue. The arrow points to recombinant HTLV1-RT.

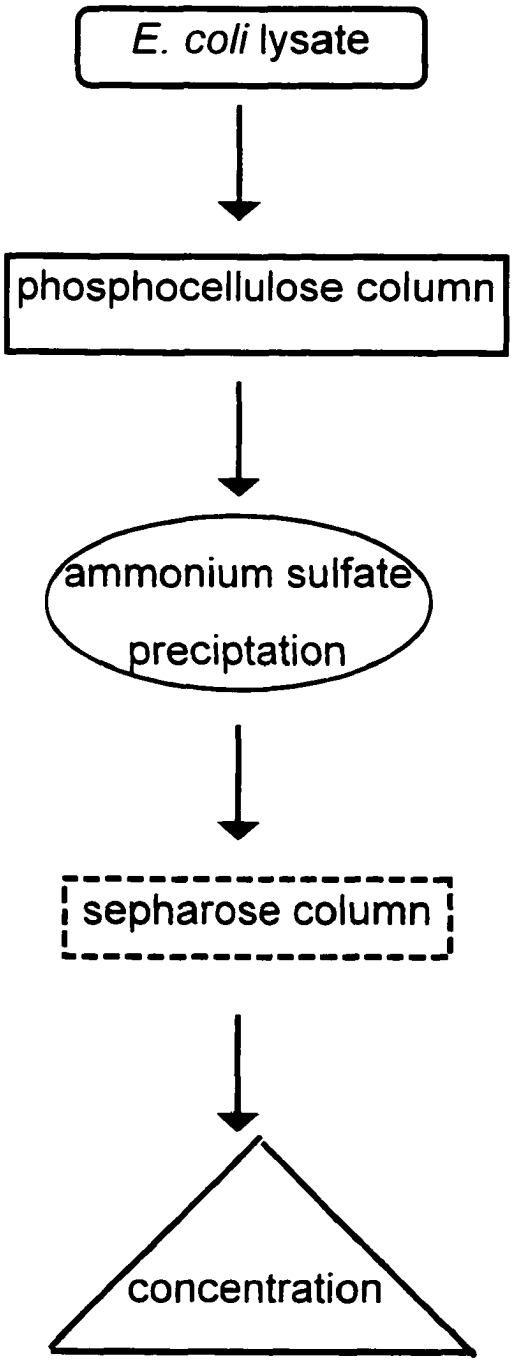


FIG. 11. HTLV-1 RT purification scheme. The purification of the recombinant HTLV-1 RT involves a two-step chromatographic process involving a cation exchanger, phosphocellulose and a gel filtration column, sepharose.

TABLE IV

Purification of Recombinant HTLV-1 RT and Associated Activity

Pool	Purification Step	Volume (mLs)	Total Protein[a]	% Total Protein	Specific Activity U/μg[b]
	E. coli lysate	170	1000 mg[b]	100	.11
I	phosphocellulose	69	856 μg	<0.1	24.8
II	sepharose	23	33 μg	<0.1	3.45

[a] Bradford Method used (14)

[b] One unit is the amount of enzyme required for the incorporation of 1 pmole of labeled dGMP in one hour at 37°C.

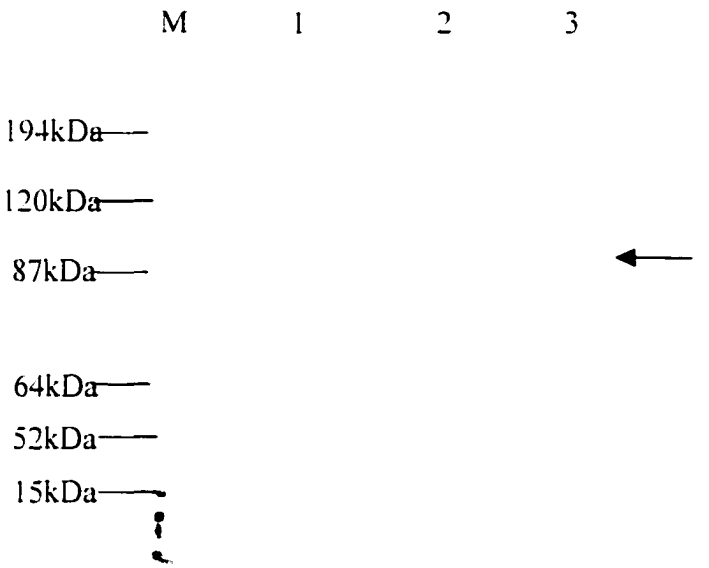


FIG. 12. Analysis and purification of HTLV-1 RT by SDS-PAGE. Samples from various stages of the purification were subjected to electrophoresis through a 4-12% Bis-Tris gel from Novex and the proteins were detected by staining with Gelcode blue as described (2). Lane M, Markers. Lane 1, crude extract . Lane 2, pooled phosphocellulose fraction, Lane 3, sepharose pool. The position of marker proteins are indicated at the sides. The arrow points to the band at the correct size of the full length translation product of pETab3.

Phosphocellulose chromatography

Initially, the bacterial cell pellet containing the recombinant HTLV-1 RT was lysed in 50mM Tris, 8% sucrose, 50mM EDTA, 5% Triton X-100, 10 mg/mL lysozyme and chromatographed on a P-11 column. The majority of the reverse transcriptase eluted as a peak at approximately 0.3-0.5M KCl (Fig. 13A). Phosphocellulose chromatography was the single most important step in the purification resulting in the greatest increase in specific activity; total recovery in this step was, unfortunately, rather low (<0.01%).

Sepharose Molecular Sieve

The phosphocellulose peak fractions of RT activity were pooled and applied to a sepharose CL-6B column. By comparing elution volumes of peaks to the standard curve shown in Appendix A, it was determined that the RT activity eluted as 2-3 separate peaks with varying molecular weights (Fig. 13B). To determine which of these pooled fractions contains the recombinant protein, samples from the protein fractions were analyzed by SDS-PAGE, and the proteins were visualized by staining with either Coomassie or Gelcode Blue. A major band (M_r = 82,000-104,000) was found in the third sepharose pool (Fig. 12). Since the location of the band in the gel is in agreement with the estimated size of our protein and has the expected activity, the pool was concentrated for further analysis.

STABILITY STUDIES

Figure 14A shows a comparison of the activities of pooled fractions at different stages of purification. A 60-fold increase in RT activity was consistently observed after the crude extract was passed through the phosphocellulose column. However, by the time the sepharose fractions (eight days) were pooled and recovered, the protein had lost 80-90% of its activity.

Time and Temperature

To determine whether the activity loss was due to the temperature conditions at which the samples were stored or if the activity loss was solely due to the duration of time that had passed, aliquots of the sepharose pools were stored at three different temperatures: 4°C, 20°C and -80°C. No significant difference in activity was calculated when sepharose

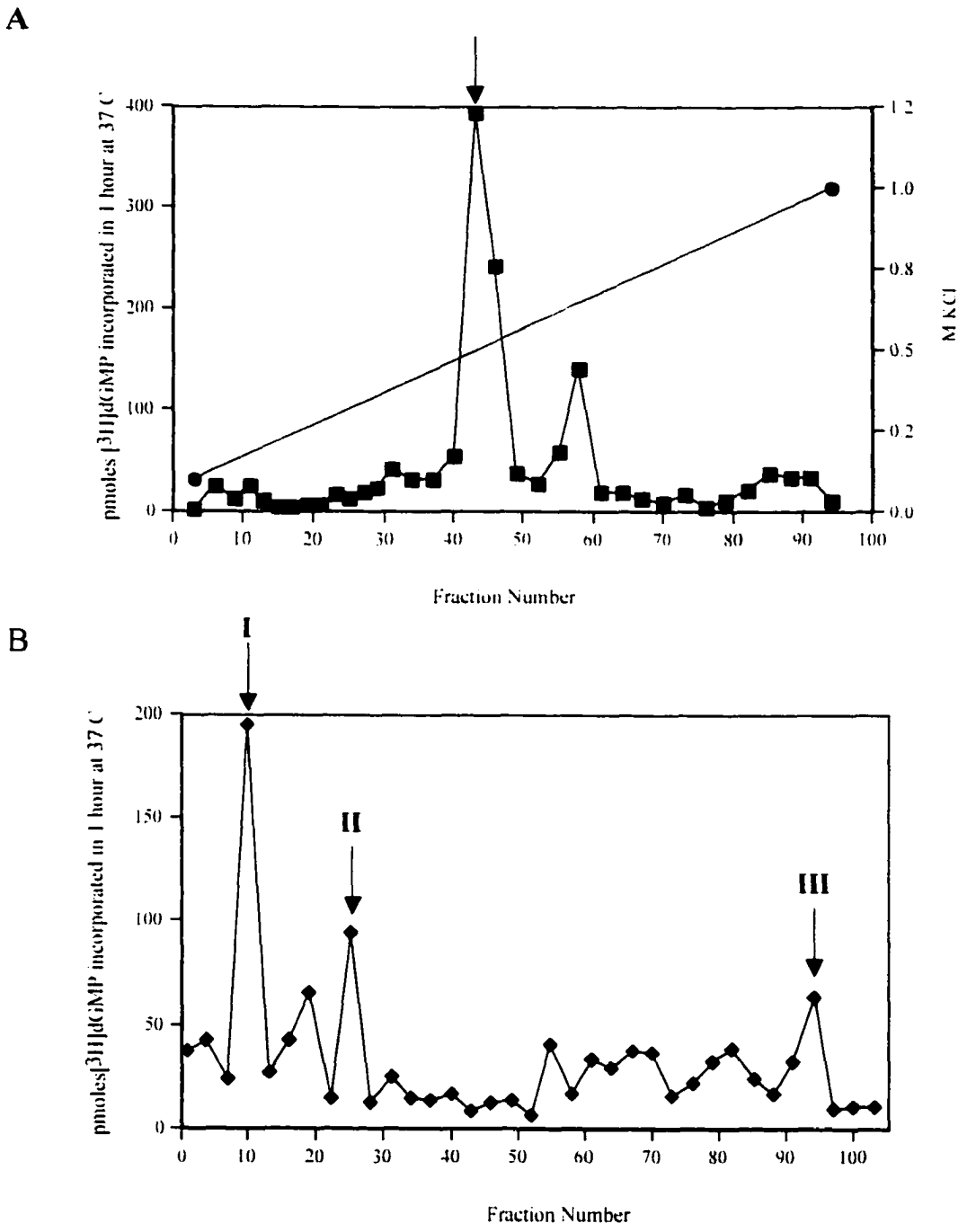
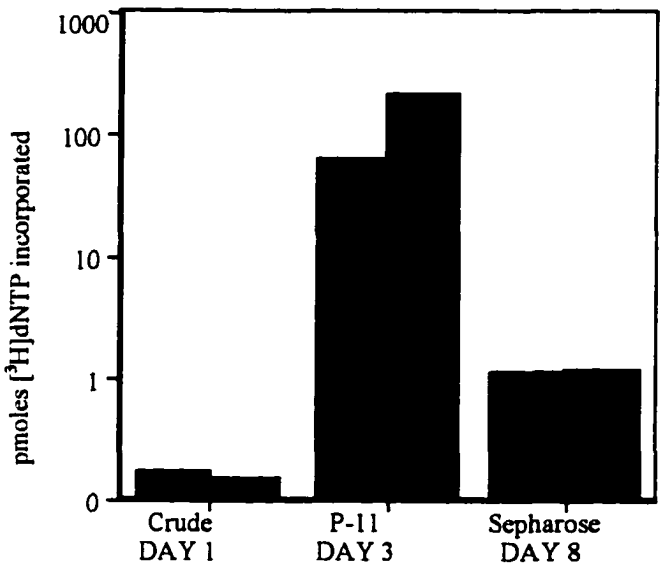


FIG. 13. Purification elution profiles. (A) Phosphocellulose elution profile. The crude extract was chromatographed on P-11 (20 x 5 cm i.d.) as described in the text. Activity of fractions (■) was determined over the range of fractions eluted with a linear gradient (●). The arrow indicates the fractions containing the recombinant protein. (B) Sepharose elution profile. The concentrated P-11 pool was applied to a sepharose 6CL-B column (52 x 5 cm i.d.). Aliquots of individual fractions were assayed for reverse transcriptase activity as described in Materials and Methods. The arrow indicates peak activities at (I) 640 kDa, (II) 540 kDa and (III) 99 kDa.

A



B

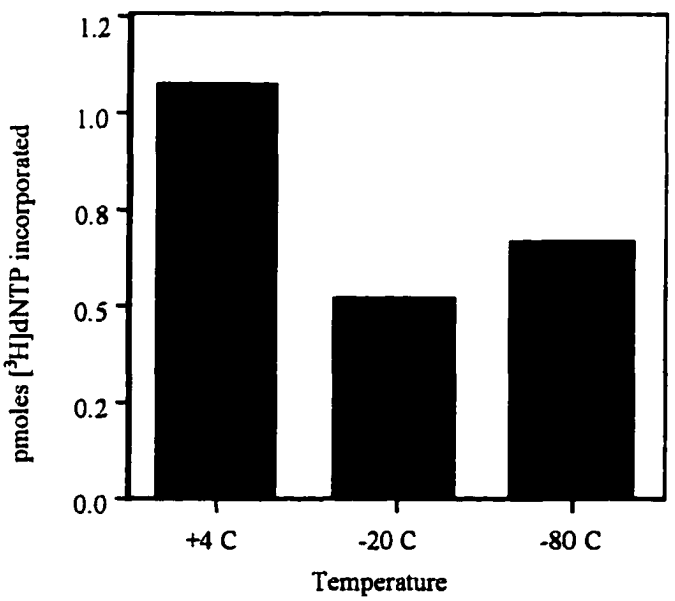


FIG. 14. Template and stability studies. RT activity was assayed as described in Materials and Methods. Black boxes represent poly(C)•oligo(dG) template. Gray boxes represent poly(A)•oligo(dT). (A) Stability over time was tested as described on 10 µL samples from each stage of purification. (B) Aliquots of the sepharose pool were stored at three temperatures: 4°C, -20°C and -80°C.

samples were stored at different temperatures (Fig. 14B). This pattern appears to be time dependent and was seen at all stages of purification. The addition or elimination of reducing agents did not have any effect on the rate of activity loss. Activity could not be recovered by reconstitution with the other sepharose pools.

To enhance the stability of the P-11 pool before the pool is applied to the sepharose column, complete dehydration of the P-11 pool by speed vacuum drying or lyophilization was suggested. Aliquots of the P-11 pool were either dried using the speed vacuum in our laboratory or shell-frozen in liquid nitrogen and lyophilized at Eastern Virginia Medical School. The aliquots were then stored at -70°C. The samples were reconstituted in RTB-50 and RT activity was measured over a period of one week.

Figures 15 and 16 compare the RT activity of a P-11 pool stored at the same temperatures investigated earlier as well as speed vacuumed and lyophilized samples. The graphs illustrate that P-11 pools are not stable in any particular storage condition. In Fig.16, the P-11 pool dropped to 67% of the original activity when speed vacuumed dry. The sample that was frozen in liquid nitrogen and stored at -70°C dropped similarly to 72% of the original activity. As expected, almost all samples dropped in activity after a week's time had passed. In some cases, activity dropped to almost 30% of the P-11 pool activity. In other preparations, this decrease was not seen at all which adds to the overall inconsistency of storage conditions. Results of this study indicate that perhaps the protein's activity is intrinsically low or that the protein exists in an unfavorable folded state. These speculations will be further addressed in a subsequent section.

TEMPLATE SPECIFICITY

As shown in Figs. 14A and 16, reverse transcriptase activity was also investigated with three synthetic homopolymeric templates, poly(rA)•oligo(dT), poly(rC)•oligo(dG) and poly(dA)•oligo(dT). In earlier studies of template specificity, only poly(rA)•oligo(dT) and poly(rC)•oligo(dG) were analyzed. The recombinant protein does not have a strong preference for either synthetic template (Fig. 14A). However, the recombinant protein in the phosphocellulose pool did prefer the poly(rA)•oligo(dT)

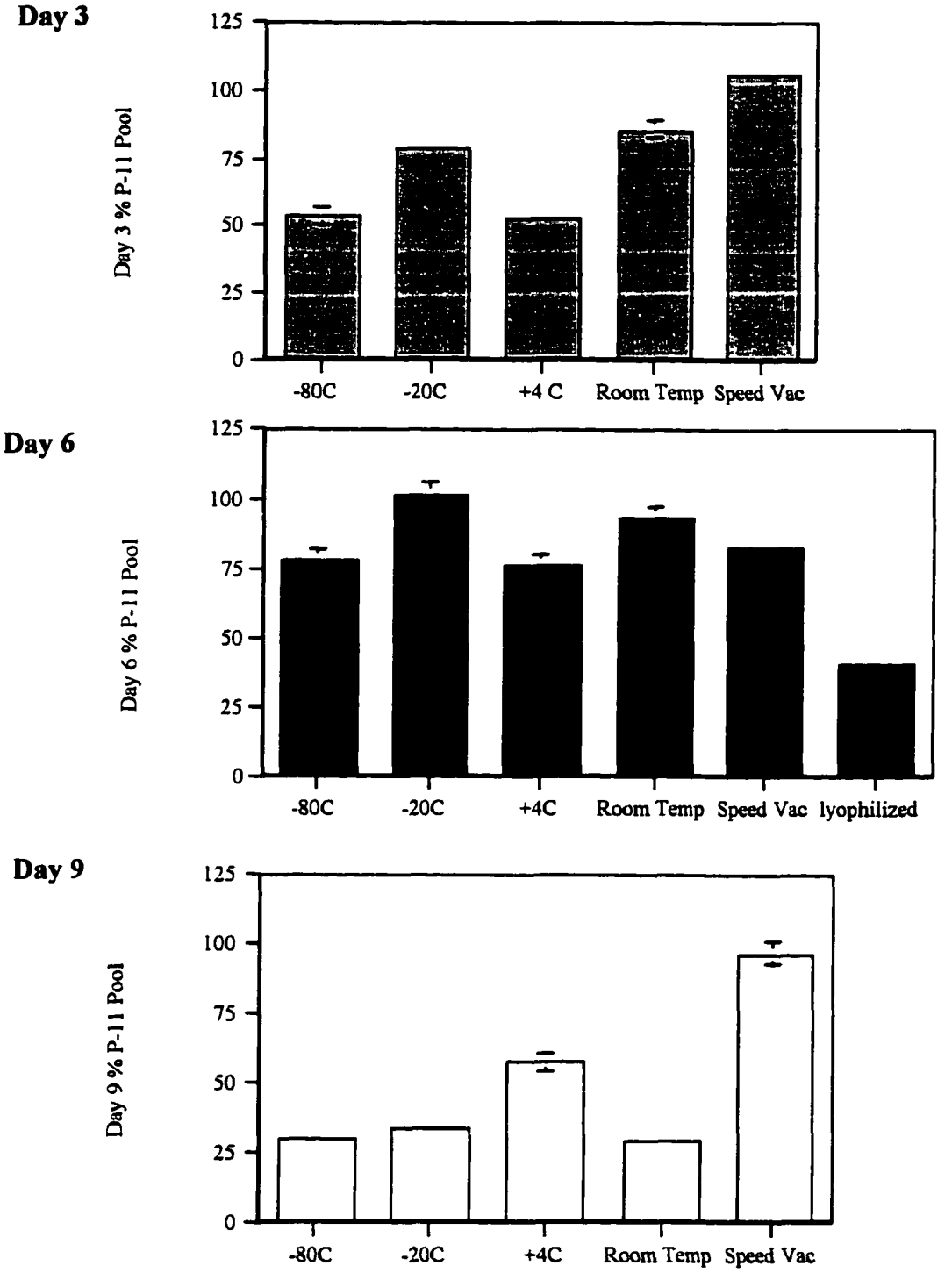
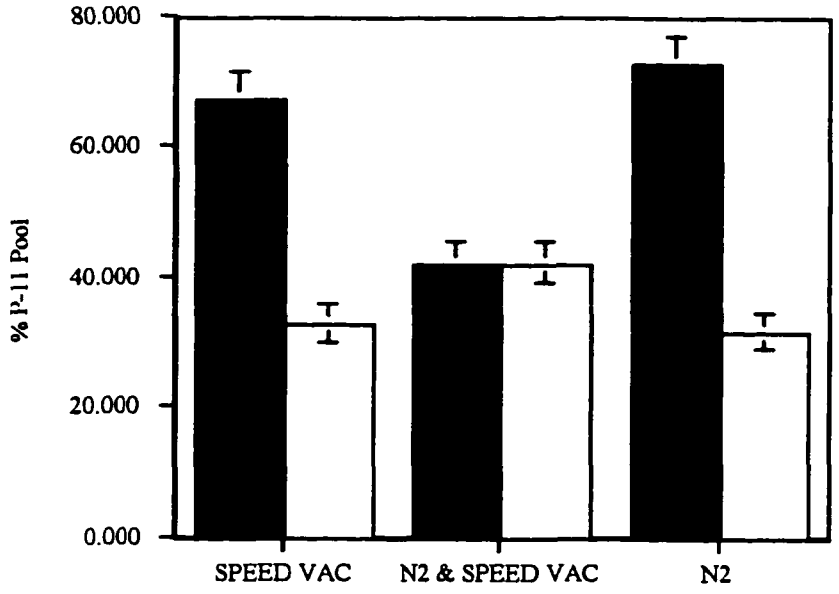


FIG. 15. Stability of P-11 pools. RT activity of P-11 pools stored at different temperatures and conditions was measured. After the P-11 pool was dialyzed, it was stored in 50 µl aliquots at various temperatures.

Next Day



One Week

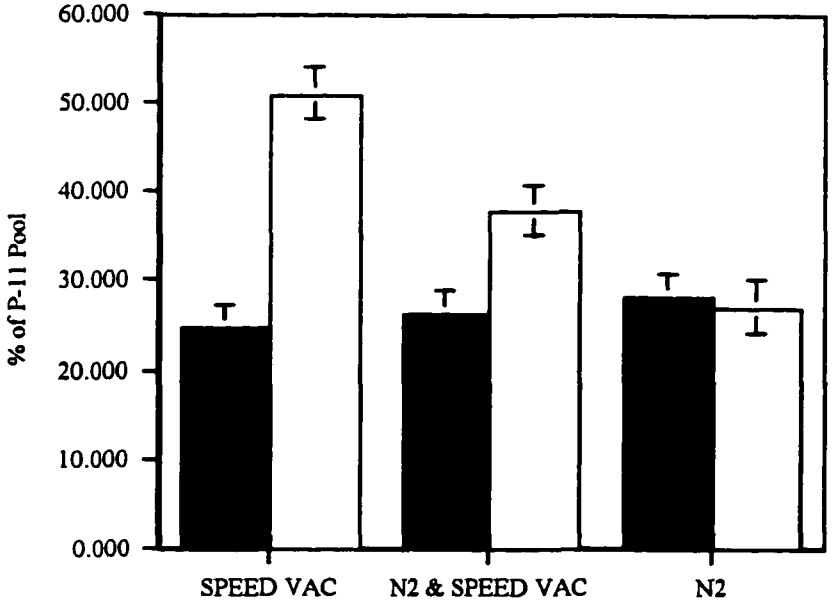


FIG. 16. Re-examination of the stability of the P-11 pool. Another P-11 preparation was further analyzed to assess its stability when stored as a speed vacuumed sample a liquid N_2 frozen sample. Shaded boxes represent poly(C)•oligo(dG) template. Open boxes represent poly(A)•oligo(dT).

template over the poly(rC)•oligo(dG) with a 3-fold increase in nucleotide polymerization (Fig. 14A). Analysis of crude extracts carrying the plasmid by polyacrylamide gel electrophoresis (Fig.12) indicated that many proteins were being synthesized that could contribute to the activity of the protein. The increase in activity was attributed to cellular polymerase interference, particularly *E. coli* Pol I, which has a greater affinity for the poly(rA)•oligo(dT) template.

E. coli DNA POLYMERASE I

Other research reports have confirmed that highly purified DNA polymerase I of *E. coli*, when presented with a poly(rA)•poly(dT) template, can form poly (dT) with a high efficiency (5, 51, 100). Furthermore, poly(rA)•poly(dT) is an equally effective template for both viral and cellular DNA polymerases, and therefore cannot be used to distinguish reverse transcriptase activity from that of other DNA polymerases (101). In contrast, when the same group tested the synthetic oligomer-homopolymer templates, such as poly(rA)•poly(dT) and poly(dA)•poly(dT), the viral reverse transcriptase could clearly be distinguished from the cellular DNA polymerase. The AMV RT and Mason-Pfizer monkey virus (M-PMV) showed a preference for the poly(rA)•poly(dT) while the *E. coli* DNA Pol I showed a preference for the poly(dA)•poly(dT) template.

As previously stated, the recombinant HTLV-1 RT did not show any template specificity when comparing synthetic hybrid templates. The increase in activity seen in the P-11 pool while using the poly(rA)•poly(dT) template may be due to interferences from the *E. coli* DNA Pol I. Moreover, *E. coli* DNA polymerase has an approximate M_r = 109,000 Da. In Fig. 12, the recombinant HTLV-1 RT protein appears between the 87 kDa and 120 kDa molecular weight markers. Since SDS-PAGE provides a relative estimation of molecular weight, it is possible that the *E. Coli DNA* polymerase was purified, not the HTLV-1 RT, which may explain the increase in activity in the P-11 pool and decrease in the sepharose pool. To examine this possibility, the high template specificity of *E. coli* DNA Pol I for a poly(dA)•poly(dT) template was confirmed by making serial dilutions of pure *E. coli* DNA Pol I (Promega). The polymerase activity of

each dilution was measured using three templates: poly(rA)•oligo(dT), poly(rC)•oligo(dG) and poly(dA)•oligo(dT). Figure 17 illustrates that *E. coli* DNA Pol I has a high specificity for a poly(dA)•oligo(dT). The polymerase activity with this template is a 5-50-fold increase of the activity obtained from a poly(rC)•oligo(dG) template. With every subsequent preparation, activity assays were performed using both of these templates. The ratio of activity of poly(dA)•oligo(dT)/ poly(rC)•oligo(dG) of subsequent preparations was approximately 1.1-1.5 indicating that most likely the activity seen in the different purification steps is from the recombinant HTLV-1 RT.

To further confirm that the protein isolated was HTLV-1 RT, the migration pattern of a 1/100 dilution of *E. Coli* DNA Pol I was analyzed on an SDS PAGE gel and compared to the migration pattern of purified HTLV-1 RT. Figure 18 illustrates that the *E.coli* DNA Pol I does indeed migrate at a higher molecular weight than the recombinant HTLV-1 RT protein. Further analysis by Western blotting indicated that there was no cross reactivity with the T7 tag antibody. *The E. coli* DNA Pol I enzyme did not appear in the Western blot.

Still, the stability issue has not been resolved. The protein may not maintain its native, active form due to improper folding of the enzyme. The N-terminus of the reverse transcriptase may play a role in the proper folding of an active enzyme. For example, in HIV-1 RT, the 20 amino-terminal amino acids are absolutely required for its polymerization activity (42). Therefore, by identifying the authentic N-terminus of the HTLV-1 RT, a basis for the structural analysis of the enzyme can be established. Further investigation of the correct N-terminus would allow for a better understanding of folding differences that may explain the low stability of the enzyme.

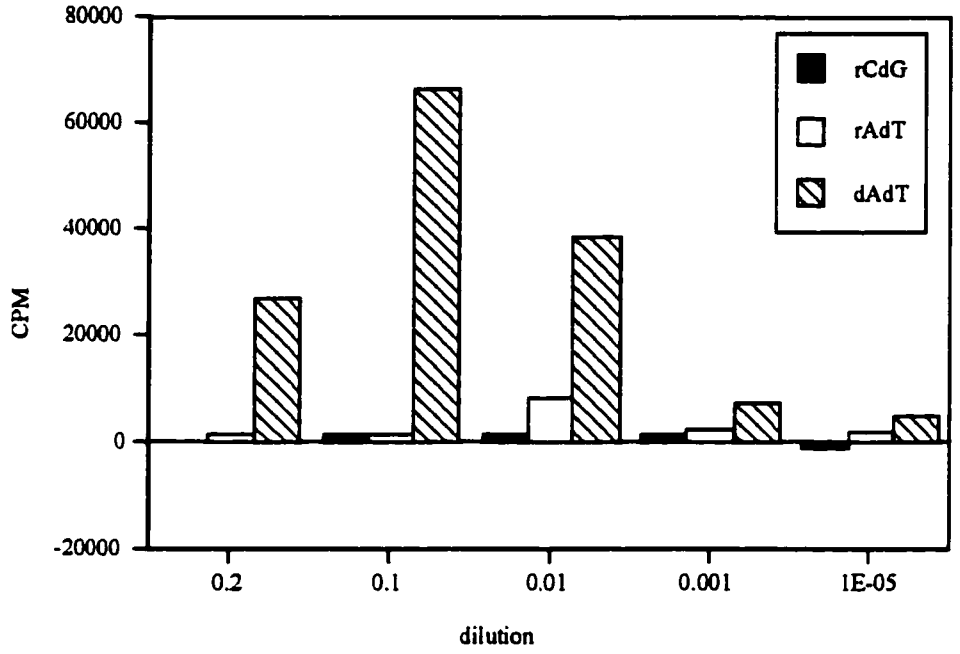


FIG. 17. Activity assay of *E. coli* DNA Polymerase I. Pure *E. coli* DNA Pol I (Promega) was diluted in RTB-50: 1:10,000;1:1,000;1:100, 1:10;1:5. The polymerase activity was measured for each diluted sample using three different templates: poly(rA)•oligo(dT), poly(rC)•oligo(dG) and poly(dA)•oligo(dT).

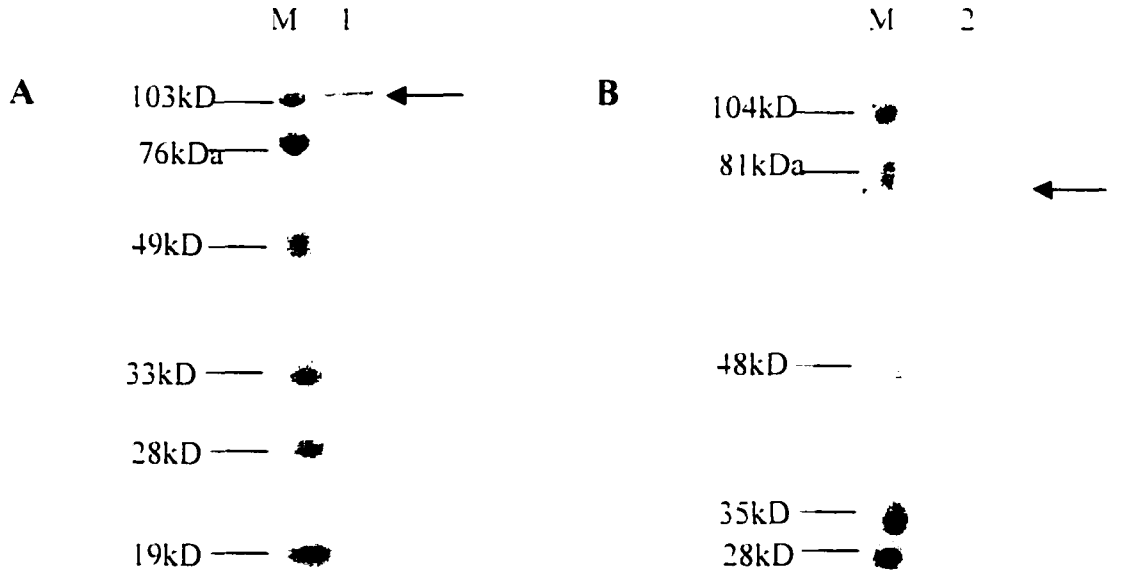


FIG. 18. SDS-PAGE analysis of *E. coli* DNA Polymerase I. Fractions were subjected to electrophoresis on a 16% Tricine gel and stained with Gel Code Blue. (A) Lane M. Markers. Lane 1. DNA Pol I (B) Lane M. Markers. Lane 2. concentrated sepharose pool. The position of marker proteins are indicated at the sides. The black arrow points to *E. coli* DNA Polymerase I and HTLV1-RT. respectively.

CHAPTER IV

RESULTS: IDENTIFICATION OF THE N-TERMINUS OF HTLV-1 RT

Translation of the retroviral reverse transcriptases in the *pol* region is a complex process because of the occurrence of frameshift events. The first reading frame of the genomic RNA encodes for the Gag protein. The first frameshift event results in a Gag-Pro polyprotein precursor. In some retroviruses, such as BLV, a second frameshift occurs resulting in the synthesis of a Gag-Pro-Pol polyprotein precursor. These precursors must be cleaved by viral protease to produce the Pol related enzymes (16, 151).

Little is known about the cleavage sites of HTLV-1 protease, particularly in the *pol* region. HTLV-1 is closely related to HTLV-2 and BLV, and their sequences show a high degree of homology at the DNA and protein level (52, 84). Perach et al. determined the amino terminus of BLV RT matches the last 26 codons of the *pro* gene and is coded for by the *pro* reading frame (92). As discussed earlier, HTLV-1 also uses two successive frameshifts to form its Pol polyprotein. Thus, an analogous cleavage position in the Pro-Pol precursor would create a Pol protein with the first 25 amino acids being derived from the protease reading frame.

Published alignments between the protein sequences of the reverse transcriptases of HTLV-1 and HIV-1 (52) begin the comparison at the first amino acid after the *pro-pol* frameshift in HTLV-1 at a position 27 amino acids downstream from the known amino terminus of HIV-1 RT (25, 61). However, an alignment beginning at the first N-terminal amino acid of HIV-1 RT with the proposed HTLV-1 coding sequence which includes the extra 25 amino acids from the *pro* ORF shows significant similarity. Figure 6 shows this homology comparison and includes equivalent regions from both HTLV-1 and BLV as well. Thus, based on the above evidence, the amino terminus of HTLV-1 RT is believed to be amino acid number 158 at the *pro* ORF (100), 25 amino acids upstream of the frameshift into the *pol* ORF. This cleavage site corresponds to Leu^{157}-Pro^{158} in the *pro* ORF.

To identify the authentic amino terminus of the HTLV-1 RT, Dr. Laura Moen constructed a clone using the same strategy as that used to clone the HTLV-1 RT (Fig. 5) to produce a protein fragment composed of the amino acids from the C-terminal region of the protease up through the amino acids in the frameshift region of the *pol* ORF. The Pro-Pol protein is approximately 11,700 Da and contains the region tentatively identified as the cleavage site based on the sequence alignments (Fig. 19).

PREPARATION OF SAMPLES FOR CLEAVAGE ASSAY

Partial purification of the Pro-Pol protein for the cleavage assay

The protein was partially purified using a Ni^{+2} affinity column and an S-tag agarose resin as described in the methods section. As shown in Fig. 20A, the majority of the Pro-Pol protein eluted in the 50 mM imidazole and the first two 100 mM imidazole washes. These fractions were further purified with S-tag agarose to approximately 70% heterogeneity (Fig. 20B, lane 6).

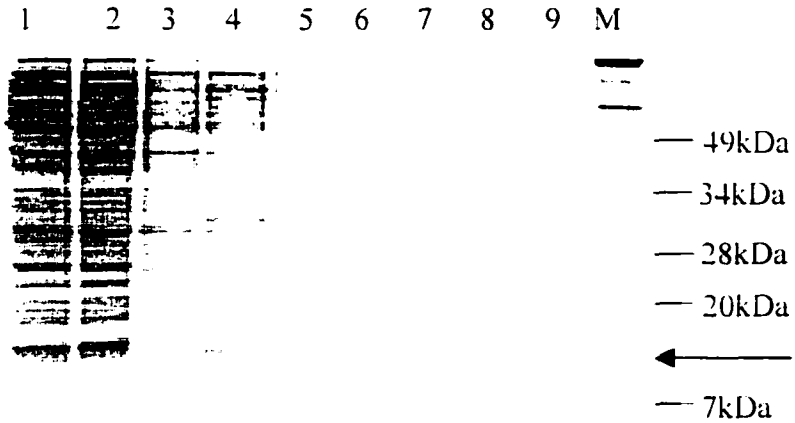
Refolding and activation of protease

Before the HTLV-1 protease clone could be utilized for processing of the Pro-Pol protein, it was first refolded and activated in low pH buffers. The HTLV-1 protease functions as a dimer, composed of two identical monomeric subunits. Each monomeric subunit has a M_r=14,000. *In vivo*, protease activation occurs via excision of the protease monomer from the Gag-Pro precursor followed by dimerization of the monomeric units. Dimerization is essential for its enzymatic activity (39). The protease construct was partially purified following an already published procedure for *in vitro* activation (19). The sequence for the protease construct is depicted in Appendix C. As shown in Fig. 21, the expressed protease construct migrated at a higher molecular weight (M_r=28,000) than the weight predicted for the monomer. This occurrence can be attributed to the histidine tag which carries a positive charge. The mass:charge ratio becomes altered resulting in the slower migration of the protein. Still, the protease was expressed in both the insoluble precipitate in the dialysate and the supernatant (lanes PPT and R). The size of this expression product suggests that the protease did not completely autoprocess (i.e. the histidine tag was not completely processed during dialysis), yet, protease activity

```
10                  20                  30                  40
ATG AAA GAA ACC GCT GCT GCT AAA TTC GAA CGC CAG CAC ATG GAC
TAC TTT CTT TGG CGA CGA CGA TTT AAG CTT GCG GTC GTG TAC CTG
Met  Lys  Glu  Thr  Ala  Ala  Ala  Lys  Phe  Glu  Arg  Gln  His  Met  Asp

      50                  60                  70                  80                  90
AGC CCA GAT CAG GGT ACC CTG GTG CCA CGC GGT TCC ATG GCC ATC
TCG GGT CTA GTC CCA TGG GAC CAC GGT GCG CCA AGG TAC CGG TAG
Ser  Pro  Asp Gln  Gly  Thr  Leu  Val  Pro  Arg  Gly  Ser  Met Ala  Ile

                100                 110                 120                 130
ATA GGT CGT GAT GCC TTA CAA CAA TGC CAA GGC GTC CTG TAC CTC
TAT CCA GCA CTA CGG AAT GTT GTT ACG GTT CCG CAG GAC ATG GAG
Ile  Gly  Arg  Asp  Ala  Leu Gln Gln  Cys Gln  Gly  Val  Leu  Tyr  Leu

     140                 150                 160                 170                 180
CCT GAG GCA AAA GGG CCG CCT GTA ATC TTG:CCA ATA CAG GCG CCA
GGA CTC CGT TTT CCC GGC GGA CAT TAG AAC:GGT TAT GTC CGC GGT
Pro  Glu  Ala Lys  Gly  Pro  Pro  Val  Ile  Leu:Pro  Ile  Gln Ala  Pro

                190                 200                 210                 220
GCC GTC CTT GGG CTA GAA CAC CTC CCA AGG CCC CCC GAA ATC AGC
CGG CAG GAA CCC GAT CTT GTG GAG GGT TCC GGG GGG CTT TAG TCG
Ala  Val  Leu  Gly  Leu Glu  His  Leu  Pro  Arg  Pro  Pro  Glu  Ile  Ser

      230                 240                 250                 260                 270
CAG TTC CCT TTA AAC CCA GAA CGC CTC CAG GCC TTG CAA CAC TTG
GTC AAG GGA AAT TTG GGT CTT GCG GAG GTC CGG AAC GTT GTG AAC
Gln  Phe  Pro  Leu  Asn  Pro  Glu  Arg  Leu  Gln  Ala  Leu  Gln  His  Leu

                280                 290                 300                 310
GTC CGG AAG GCC CTG GAG GCA GGC CAT CTC GAG CAC CAC CAC CAC
CAG GCC TTC CGG GAC CTC CGT CCG GTA GAG CTC GTG GTG GTG GTG
Val  Arg  Lys  Ala  Leu  Glu  Ala  Gly  His  Leu  Glu  His  His  His  His

      320
CAC CAC TGA
GTG GTG ACT
His  His  ***
```

FIG. 19. Amino acid sequences of the HTLV-1 Pro-Pol protein. The Pro-Pol protein sequence spans a region where the HTLV-1 protease is known to terminate. Iso^{132}-Leu^{157} (86), and continues past the pro-pol frameshift site. (-----) dotted lines represent the putative cleavage site. Boxed amino acid sequence represents the region of the second frameshift.

A



B

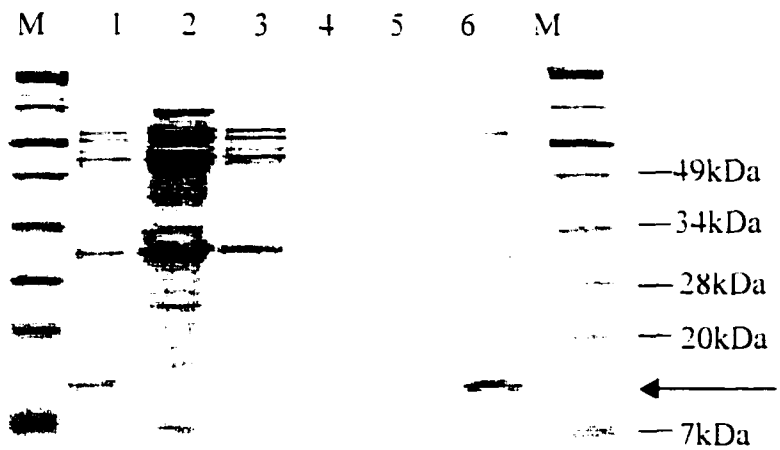


FIG. 20. Analysis of the purification steps of the Pro-Pol protein by SDS-PAGE. (A) Ni^{+2}-NTA affinity fractions of the Pro-Pol protein purification were subjected to electrophoresis through a 16% Tricine gel and the proteins were detected by staining with Gelcode blue. Lane 1, crude extract. Lane 2, pass. Lane 3, 10 mM imidazole wash. Lane 4, 50 mM imidazole wash. Lanes 5-8, 4 sequential 100 mM imidazole washes. Lane 9, 500 mM imidazole eluate. (B) The 50 mM and the first two 100 mM imidazole washes of the Pro-Pol protein Ni^{+2} purification were pooled, dialyzed and incubated with S-tag agarose. Fractions were subjected to electrophoresis through a 16% Tricine gel and the proteins were detected by staining with Gelcode blue. Lane 1, Ni^{+2} dialysate. Lane 2, pass. Lanes 3-5, sequential low salt washes. Lane 6, concentrated $MgCl_2$ eluate. Lane M, molecular weight markers. The position of marker proteins are indicated at the sides. The arrow points to the band at the correct size of the Pro-Pol protein.

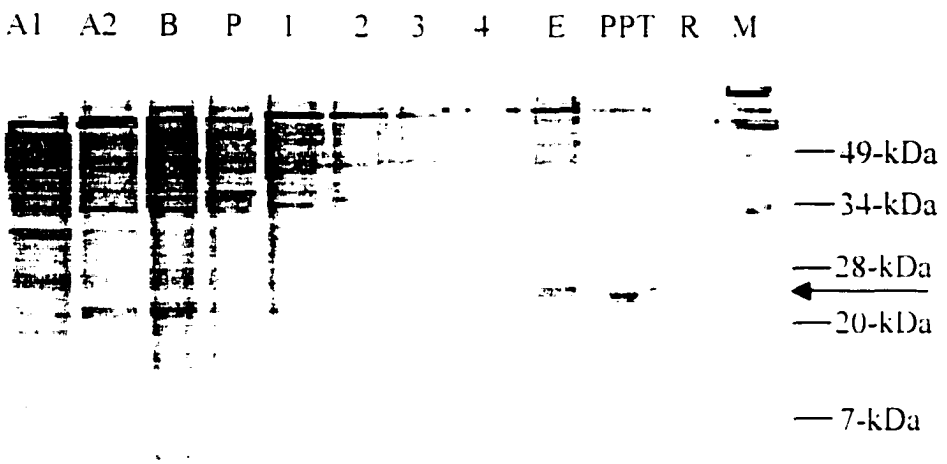


FIG. 21. Purification of histidine-linked HTLV-1 protease on Ni^{2}-NTA agarose. Lanes A1 and A2, 1st and 2nd Buffer A washes, respectively. Lane B, supernatant from Buffer B wash. Lane P, pass. Lanes 1-3, 20 mM imidazole washes. Lane E, 1 M imidazole eluate. Lane PPT, resuspended precipitate formed during dialysis. Lane R, refolded and activated protease. Lane M, molecular weight markers. The position of marker proteins are indicated at the sides. The arrow points to the histidine-linked protease.

measurements of the supernatant dialysate demonstrate that the protease is still active with a specific activity =1.3 ΔAU μL/μg•sec (Fig. 22). Contrary to the findings of Ding et al. (20), most of the activity was found in the supernatant, not the white precipitate (data not shown). However, our data does support the fact that the incomplete processing of the histidine tag from the protease fusion protein does not interfere with the activity of the enzyme (20).

CLEAVAGE PRODUCT ANALYSIS

Months before the partial purification scheme of the Pro-Pol protein was established, HPLC analysis was attempted to separate the Ni^{+2} affinity pool using a reverse phase C18 column. In using HPLC analysis, Pro-Pol protein, protease and cleavage product peaks could be separated and evaluated individually. Regretfully, the protein could not be recovered in an adequate quantity and the investigation was terminated.

Fortunately, cleavage analysis of the Pro-Pol protein does not require the protein to be pure and the Pro-Pol protein from the Ni^{+2} step and the final S-tag purification step were cleaved with active protease. After an overnight incubation of the Pro-Pol protein with activated protease, the samples were analyzed by SDS-PAGE. Figure 23A shows the results of the cleavage of the protein by the protease. The Pro-Pol protein appears as a 12,000 Da band and the active protease appears as a 28,000 Da band. A diffuse protein band at an approximate M_r=5,500, which was suspected to be a fragment of the processed Pro-Pol protein, only appears in the lane corresponding to the sample taken from the cleavage reaction. The molecular weight of the entire Pro-Pol protein is approximately 11,700 Da. Hydrolysis by HTLV-1 protease at the putative cleavage site between Leu^{55} and Pro^{56} (numbered according to the amino acid sequence of the Pro-Pol protein) would produce 5,700 Da and 6,000 Da fragments. As shown in Fig. 23B, these same cleavage samples were analyzed by immunoblotting with an antibody alkaline phosphatase conjugate directed against the N-terminal 15 amino acid S-tag leader sequence fused to the Pro-Pol protein. As shown in lane 2, the Pro-Pol protein construct reacted with the antibody to give a band corresponding to the M_r = 12 kDa expected for the peptide.

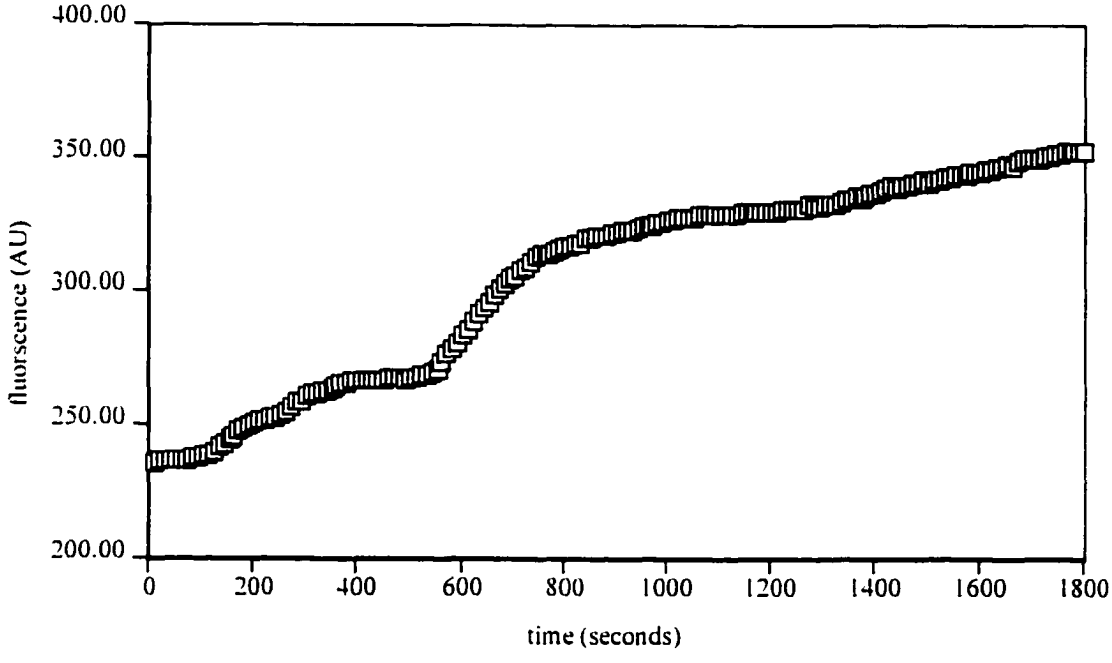
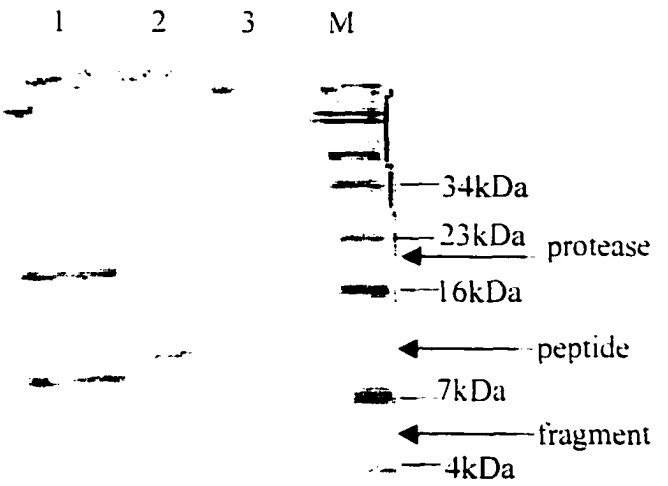


FIG. 22. Activity of protease produced from the insoluble IK19 construct. 200 μL of the dialysate of the IK19 protease construct was incubated in the presence of a fluorescent, synthetic substrate, in 100 mM sodium citrate buffer, pH 5.3. Excitation wavelength was set at 320 nm/15 nm bandpass.

A



B

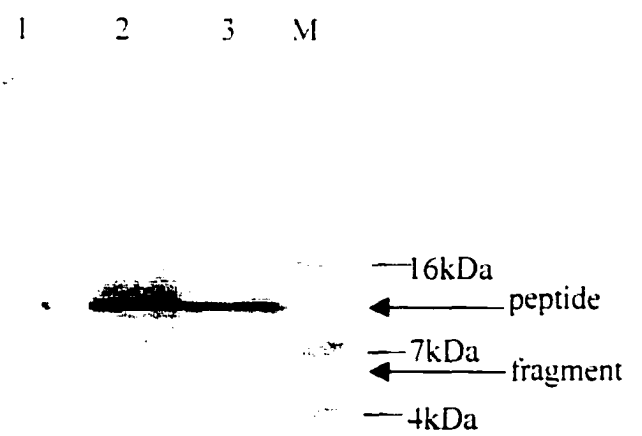


Fig. 23. Processing of the Pro-Pol protein from the concentrated S-tag eluate by HTLV-1 protease. Partially purified peptide samples were incubated with activated protease overnight at 37°C. (A) SDS-PAGE was carried out as described in Materials and Methods using a 10-20 % acrylamide gel which was stained with colloidal Coomassie Blue, or (B) analyzed by Western blotting with an S-tag antibody alkaline phosphatase conjugate. Lane 1, protease only control. Lane 2, Pro-Pol protein only control. Lane 3, cleavage reaction sample. Lane M, markers. The position of marker proteins are indicated at the sides.

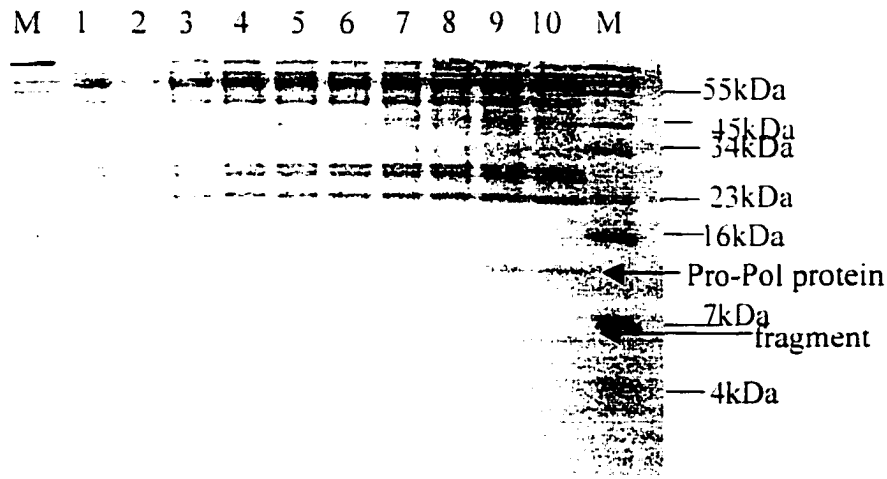
Similarly, both the 12 kDa band and the 5.5-kDa band were detected in the cleavage product sample (Fig. 23B, lane 3). Neither the 12 kDa band nor the 5.5-kDa band was detected in the protease only sample (Fig.23B, lane 1).

PROTEOLYTIC PROCESSING SITE ANALYSIS BY TRYPSIN DIGESTION AND MASS SPECTROMETRY

To identify the location of the proteolytic processing site of the N-terminus of the HTLV-1 RT, several lanes were loaded with cleavage samples and a Pro-Pol protein only sample on an 18% Tricine gel. The gel was stained with Colloidal Coomassie Blue and sent to the Biomolecular Research Facility at the University of Virginia in Charlottesville, Virginia (Fig. 24). Dr. Nicholas Sherman removed the 5.5 kDa band, digested it with trypsin, and the peptides were analyzed by mass spectrometry. Since trypsin cleaves at the C-terminal side of lysine and arginine residues (26), a list of possible peptides was generated from a hypothetical tryptic digestion of the Pro-Pol protein sequence as shown in Table V. A summary of the seven fragments actually recovered from the tryptic digestion of the 5.5 kDa fragment is shown on the right-hand side of Table V. The actual spectra can be found in Appendix D. Five of the seven peptides obtained from the tryptic digest of the cleavage fragment matched those of the peptides calculated from the SEQUEST algorithm. Peptides 1 and 2 correspond to sequences at the N-terminus of the Pro-Pol protein, while peptides 6 and 7 correspond to sequences at the C-terminus.

Isolation of only one fragment of the Pro-Pol protein was expected to produce either C-terminal or N-terminal peptides upon tryptic digestion. The fact that both N-terminal and C-terminal peptides were obtained indicates the presence of two protein bands, which were previously identified as a single Pro-Pol protein fragment migrating at a distance of 5,500 Da. The total molecular weight of the peptides found in the actual tryptic digest is approximately 10,700 Da, almost twice the molecular weight of the band that was digested. Thus, analysis confirms that the diffuse band at M_r =5,500 Da was indeed the two fragments of the Pro-Pol protein predicted by cleavage at or near the proteolytic processing site identified by the sequence alignment.

A



B

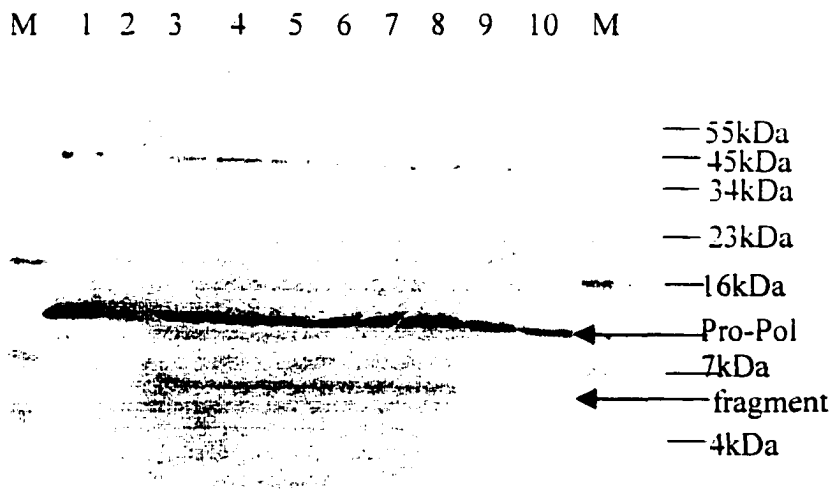


FIG. 24. SDS-PAGE gel sent for MS analysis. A 10-20 % acrylamide gel of cleavage samples either (A) stained with colloidal Coomassie Blue or (B) analyzed by Western blotting with an S-tag antibody alkaline phosphatase conjugate. Lane 1, Pro-Pol protein only control. Lane 2, protease only control. Lanes 3-10, cleavage reaction samples. Lane M, markers. The position of marker proteins are indicated at the sides.

TABLE V

Predicted and Actual Tryptic Peptides of Recombinant Pro-Pol protein

	Predicted		Actual			
Peptide	M_r (Da)	Sequence	Peptide #	M_r (Da) $(M+H)^+$	Charge z	Sequence
Glu^{3}-Lys^{8}	590.7	ETAAAK	1	1581.2	+3	QHMDSPDQGTLVPR
Gln^{12}-Arg^{25}	1581.8	QHMDSPDQGTLVPR	2	820.3	+2	GSM(o)AIIGR
Gln^{12}-Arg^{25}	1597.8	QHM(o)DSPDQGTLVPR	3	1833.1	+2	DALQQCQGVLYLPEAK
Gly^{26}-Arg^{33}	805.0	GSMAIIGR	4	595.3	+1	GPPVIL
Gly^{26}-Arg^{33}	821.0	GSM(o)AIIGR	5	3116.3	+4	PIQAPAVLGLEHLPRPPEISQFPLNPER
Asp^{34}-Lys^{49}	1834.9	DALQQCQGVLYLPEAK	6	1077.6	+2	LQALQHLVR
Asp^{34}-Lys^{49}	1848.1	DALQQCQGVLYLPEAK	7	1661.8	+4	ALEAGHLEHHHHHH
Gly^{50}-Arg^{83}	3694.3	GPPVILPIQAPAVLGLEHLPRPPEISQFPLNPER				
Leu^{84}-Arg^{92}	1078.3	LQALQHLVR				

Note: M(o)denotes oxidized Met. C denotes carboxyamidomethyl Cys

The identity of the cleavage site location was determined by examination of peptides 4 and 5, neither of which matches any of the predicted peptides in Table V. The mass spectra were determined manually. As shown in Figure 19, these peptide sequences are located in the middle of the *pro-pol* amino acid sequence. The C-terminus of peptide is Leu^{55} while the N-terminus of peptide 5 is Pro^{56}. Cleavage at this position could not have occurred by trypsin digestion since no lysine or arginine residues are present.

CHAPTER V

CONCLUSIONS

The collaborative activity of the protease, reverse transcriptase and integrase enzymes during retroviral replication ensures the survival of the virion. The reverse transcriptase catalyzes transcription of viral RNA into double-stranded DNA, which is later incorporated into the host to form the provirus. Structural information obtained about the reverse transcriptase could aid in the development of antiviral therapeutic agents aimed at inhibiting the early onset of replication.

Unlike other retroviruses, little is known about the structure or proteolytic processing of the reverse transcriptase of the human T-cell leukemia virus. In large part, this is due to the fact that little reverse transcriptase is produced from the virus. The availability of methods to express foreign genes in *E. coli* facilitates isolation of large quantities of recombinant reverse transcriptase. Although larger amounts of total protein are recovered by this method, the catalytic amount of product still does not support serious biochemical or biophysical investigations.

Two reports have been published on the expression of recombinant HTLV-1 RT. In the first, Owen et al. expressed an enzymatically active HTLV-1 RT in bacteria, and their clone was created such that translation was initiated at Pro^{33}, which is located within the proposed *pro-pol* frameshift site (91). In the second report, Trentin et al. hypothesized that the N-terminus of the RT should immediately follow the known C-terminus of the protease (131). The resulting protein from their construct also had RT activity (13). However, they used an *in vitro* transcription/translation system that may have had an impact on overall folding and potential subunit structure. Taken together, there are discrepancies between the two published reports on recombinant HTLV-1 RT. Furthermore, Perach et al. (92) found that the amino acid sequence requirements and molecular arrangements of the recombinant BLV RT, the closest relative to HTLV, are substantially different from those of both HTLV-1 RT studies. Here, a recombinant

enzyme was also expressed in *E.coli* to further address the biochemical properties of HTLV-I RT as well as functional aspects of protein-nucleic acid interactions.

By using sequence similarities with other retroviral reverse transcriptases in conjunction with information about the location of the ribosomal frameshift, our laboratory has cloned the correct coding region for the HTLV-1 reverse transcriptase. To our knowledge, this recombinant protein is the first reverse transcriptase activity expressed from the CH strain of HTLV-1.

In summary, a virtually homogenous, recombinant HTLV-1 RT was separated from *E. coli* proteins and enzyme degradation products by using a two-step chromatographic technique. The purification scheme enhanced the presence of the 99kDa band on SDS-PAGE that corresponds to the predicted molecular weight of the recombinant protein. The size and activity of the recombinant enzyme agree with the only published studies on the protein purified from virus-infected human cell lines (101). Although the recombinant protein is not stable for long periods of time, it does maintain reverse transcriptase activity throughout the purification process.

The purified recombinant HTLV-1 RT was further characterized to determine its primary structure. The primary structure of proteins/peptides includes the determination of the linear order of the amino acid residues, plus, any covalently attached groups and the accurate measurement of the protein's molecular weight. To date, the amino terminal sequence of the polymerase gene of HTLV-1, which encodes the reverse transcriptase, had not been determined. The amino terminus of the reverse transcriptase may play a role in the proper folding of an active enzyme. For example, in the HIV-1 RT, the 20 amino-terminal amino acids are absolutely required for its polymerization (42).

Therefore, to confirm that the purified HTLV-1 RT protein has the authentic N-terminus, identification of the proteolytic processing site of the Pro-Pol precursor protein was determined. In the virion, the reverse transcriptase is released from the larger polypeptide precursor by proteolytic cleavage. To mimic this proteolytic processing event *in vitro*, the *pro-pol* cDNA , which spans the region where the HTLV-1 protease is known to terminate and continues 19 amino acids past the *pro-pol* frameshift, was cloned and expressed in *E. coli*. The sequence of the Pro-Pol protein contains a portion of the larger

Pro-Pol precursor and the putative protease cleavage site. Using an active HTLV-1 protease construct (Fig. 22), The protein was cleaved to produce a 5,500 Da cleavage fragment (Fig. 23). LC/MS and MS-MS analysis of the cleavage fragment revealed that the HTLV-1 protease cleavage site which generates the N-terminus of mature HTLV-1 RT is located between Leu^{157} and Pro^{158} in the *pro* ORF (82), 25 amino acids upstream of the *pro-pol* frameshift. This cleavage site does not differ from other retroviral processing sites in the *pol* region (Table VI), which suggests that retroviral proteases preferentially process their substrates at specific cleavage sites (20). The cleavage site is in agreement with the sequence alignment perform by our group and by Trentin et al.(131).

Still, activity of the purified enzyme is low (3.45 U/μg). Further efforts directed towards improving the enrichment and optimal activity of the enzyme were not successful. However, the low levels of activity observed are in agreement with those reported by others (91). The overall inconsistency in activity measurements observed in these different purification preparations may indicate that the RT activity of HTLV-1 in and of itself is low, or perhaps the presence of the T7 tag may have disrupted the favorable native conformation of an active enzyme. Trentin et al. suggest that the structure of the HTLV-1 RT is complex and thus a more detailed structural analysis is necessary (131).

The low RT activity observed in this project provides useful information about the biology of the lymphotropic viruses. Reverse transcriptases lack editing functions and have been shown to be error-prone *in vitro*. Therefore, these enzymes are presumed to contribute to the high mutation rate *in vivo* (52). The frequency of RT errors is dependent on the source of RT. For example, the error rate of HIV-1 RT ranges from 1 in to 4000 7000 (per genome per cycle) while the error rates of ASLV and MuLV are 1 in 9000 to 17,000 and 1 in 30,000, respectively (98, 104). This high level of infidelity may be a common characteristic of RNA tumor virus reverse transcriptases.

High error rates in DNA replication result in genetically diverse populations of viruses. However, HTLV-1 shows minimal strain variation in comparison to a much greater variation seen among different isolates of each lentivirus in the HIV/SIV family (28, 30, 57, 61, 77). Ratner et al.reports a 1.2%-3.3% nucleotide sequence difference among

TABLE VI

Sequence of Retroviral Protease Cleavage Sites in Their Respective PR/RT Junctions

Retrovirus	Sequence
HIV-1	Phe↓Pro
HIV-2	Leu↓Pro
SIV-2	Leu↓Pro
BLV	Gly↓Val
MuLV	Leu↓Thr
ALV	Leu↓Thr
HTLV-1	Leu↓Pro

HTLV-1 isolates (97). The low activity of the HTLV-1 RT and apparent differences in mutation rates may explain the genomic stability and replication mechanism of HTLV-1.

Unlike the HTLV-1 RT, HIV-1 RT is highly stable and has a much higher activity (273.9 pmoles/μg/hr) in the presence of a poly(rC)•oligo(dG) template (12). Chandra et al. further reports the V_{max} for HIV-1 RT is seven-fold higher than the HTLV-1 RT (12). Thus, HIV-1 replicates persistently at a high level and quickly develops extensive within-isolate sequence variation. It can be concluded, then, that there is little persistent replication of HTLV-1 because the RT activity is low. Wodarz et al. suggests that the high viral load that does exist during HTLV-1 infection (0.1-10% of peripheral blood mononuclear lymphocytes), then, is maintained largely by division of provirus-containing cells (145). In short, the virus must rely on another replication mechanism, the mitotic route, to ensure its survival. Wodarz et al. hypothesizes that mitotic division rather than infectious transmission of the virus produces most of the newly infected cells. The mitotically transmitted provirus is subsequently replicated by cellular DNA polymerases. The low mutation rates of HTLV-1 are due to the proofreading capability of cellular DNA polymerases (145).

The results of this research project have provided a basis for future structural characterization of retroviral reverse transcriptases. Tasara, et al. found significant contributions of both the N- and C- termini of the HIV-1 RT p51 subunit to the structural and functional conformation of the p66 subunit, hence, the activity of the HIV-1 RT heterodimer (129). These regions contribute significantly to the maintenance of the optimal conformation between the DNA polymerase and RNaseH catalytic sites of the p66 subunit. The authentic N-terminus of HTLV-1 RT was determined and confirmed in this project. This information will, likewise, lead to a better understanding of the protein-nucleic acid interactions of this class of enzymes. This research has also provided a basis for future studies aimed at comparing the replication pathways addressed previously which would lead to an increased understanding of the pathogenesis of HTLV-1 induced diseases.

REFERENCES

1. Alberts, B. and Herrick, G. (1971) *Meth Enzymol* **21**, 198-217.
2. Andrews, P. (1964) *Biochem J* **91**, 222.
3. Arisawi, K., Soda, M., Akahoshi, M., Matsuo, T., Nakashima, E., Tomonaga, M., and Saito, H. (1998) *Jpn J Cancer Res* **89**, 797-805.
4. Ballaun, C., King-Farrington, G., Dubrovnik, M., Rusche, J., Hauber, J., and Bohnlein, E. (1991) *J Virol* **65**, 4406.
5. Baltimore, D. and Smoller, D. (1971) *Proc Nat Acad SciUSA* **68**, 1507-1511.
6. Barat, C., Lullen, V, Schatz, O., Keith, G., Nugeyre, M.T., Grunininger-Leitch, T., Barre-Sinouiss, F., LeGrice, S.F., and Darlik, J.L. (1998) *EMBO J* **8**, 3279-3285.
7. Bartholomew, C., Cleghorn, F., Jack, N., Edwards, J., and Blattner, W. (1997) *Ann Neurol* **41**, 806-809.
8. Black, A.C., Rutland, C.T., Yip, M.T., Luo, J., Tran, G., Kalsi, A., Quan, E., Aboud, M., Chen, I.S., and Rosenblatt, J.D. (1991) *J Virol* **65**, 6645-6653.
9. Blanco, L., Bernard, M., Blasco, A., and Salas, M. (1991) *Gene* **100**, 27.
10. Blattner, W.A., Kalayanaraman, V.S., Robert-Guroff, M., Lister, T.A., Galton, D.A., Sarin, P.S., Crawford, M.H., Greaves, M., and Gallo, R.C. (1982) *Int J Cancer* **30**, 257-264.
11. Bogerd, H.P., Tiley, L.S., and Cullen, B.R. (1992) *J Virol* **66**, 7572.
12. Chandra, A., Gerber, T and Chandra, P. (1986) *FEBS Lett* **197**, 84-88.
13. Charneau, P., Alizon, M., and Clavel, F. (1992) *J Virol* **66**, 2814-2820.
14. Charneau, P. and Clavel F. (1991) *J Virol* **65**, 2415-2421.
15. Chu, M.R. and Vigna, R.A. (1997) *Pierce Previews* 1:18-21.
16. Coffin, J.M., Hughes, S.H., and Varmus, H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
17. (1989) Current Protocols in Molecular Biology, Vol 1 (Ausebel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J., Struhl, K. Eds.), p.1.8.1, Greene Publishing and Wiley Interscience, New York.
18. Dahlberg, J.E. (1988) *Adv Vet Sci Comp Med.* **32**, 1-35.
19. Delaprote, E., Dupont, A., Peeters, M., Josse, R., Merlin, M., Schnjvers, D., Hamono, B., Bejabaga, L., Cheringou, H., and Boyer, F. (1988) *Int J Cancer* **42**, 687-689.
20. Ding, Y.S., Owen, S.H., Lal, R.B., and Ikeda, R.A. (1998) *J Virol* **72**, 3383-3386.

21. Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek, Jr., T.A., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Minnich, M., Petteway, Jr., S.R., Metcalf, B.W., and Lewis, M. (1992) *Biochemistry* **31**, 6646.

22. Ferriera, O.C., Jr., Planelles, V., and Rosenblatt, J.D. (1997) *Blood Rev* **11**, 91-104.

23. Fitzgerald, P.M.P. and Springer, J.P. (1991) *Ann Rev Biophys Biophys Chem* **20**, 299-320.

24. Franchini, G. (1995) *Blood* **86**, 3619-3639.

25. Furfine, E.S. and Reardon, J.E. (1991) *J Biol Chem* **266**, 406-412.

26. Garnick, R.L., Solli, N.J., and Papa, P.A. (1988) *Anal Chem* **60**, 2546-2557.

27. Gerard, G.F. and Grandgenett, D.P. (1980) *in* Molecular Biology of RNA Tumor Viruses (Stephenson, J.R. Ed.), pp 345-367, Academic Press, San Diego.

28. Gessain, A., Gallo, R., and Franchini, G. (1992) *J Virol* **66**, 2288-2295.

29. Gessain, A., Barin, F., Vernant, J.C., Gout, O., Maurs, L., Calender A., and de The G. (1985) *Lancet* **2**, 407-410.

30. Gessain, A., Yanagihara, R., and Franchini, G. (1991) *Proc Natl Acad Sci USA* **88**, 7694-7698.

31. Gibbs, J.S., Chiou, H.C., Bastow, K.F., Cheng, Y.C., and Coen, D.M. (1988) *Proc Natl Acad Sci USA* **85**, 6672.

32. Goff, S.P. (1990) *J Acq Immu Def Synd* **3**, 817-831.

33. Goldblum, A. (1990) *FEBS Lett* **261**, 241-244.

34. Gopalakrishnan, V., Peliska, J.A., and Benkovic, S.J. (1992) *Proc Natl Acad Sci USA* **89**, 10763-10767.

35. Gout, O., Boulac, M., Gessain, A., Semah, F., Saal, F., Peries, J., Cabrol, C., Foucault-Fretz, C., Paplane, D., and Sigaux, F. (1990) *N Engl J Med* **322**, 383-387.

36. Grossman, R., Berchtold, S., Acpinus, C., Ballaun, C., Boehnlein, E., and Fleckenstein, F. (1991) *J Virol* **65**, 3721.

37. Hanchard, B. (1996) *J Acquir Immune Defic Syndr Hum Retrovirol* **13**(Suppl 1), S20-S25.

38. Harrington, Jr.,W.J., Sheremata, W., Hjelle, B., Dube, D.K., Bradshaw, P., Foung, S.K., Snodgrass, S., Toedter, G., Cabral, L., and Poiesz, B.J. (1993) *Ann. Neurol.* **33**, 411-414.

39. Haynes, B.F. and Palker, T.J. (1988) *in* Virology (Joklik, W. K. Ed.) Prentice Hall, East Norwalk.

40. Hinuma Y.M., Nagata K., Hanoaka, M., Nakai, M., and Miyoshi, I. (1981) *Proc Natl Acad Sci USA* **78**, 6476-6480.

41. Hizi, A. and Joklik, W.K. (1977) *J Biol Chem* **252**, 2281-2289.

42. Hizi, A., McGill, C., and Hughes, S.H. (1988) *Proc Natl Acad Sci USA* **85**, 1218-1222.

43. Hoffman, A.D., Banapour, B., and Levy, J.A. (1985) *Virology* **147**, 326-335

44. Hollsberg, P. (1999) *Microbiol Mol Biol Rev* **63**, 308-333.

45. Huber, H.E. and Richardson, C.C. (1990) *J Biol Chem* **265**, 10565-10573.

46. Hyland, L.J., Tomaszek, T.A., and Meek, T.D. (1991) *Biochemistry* **30**, 8454-8463.

47. Ido, E., Han, H., Kezdy, F.J., and Tang, J. (1991) *J Biol Chem* **266**, 24359-24366.

48. Ijichi, S. and Osame, M. (1995) *Int Med* **34**, 713-721.

49. Inoue, J., Yoshida, M., and Seiki, M. (1987) *Proc Natl Acad Sci USA* **84**, 3653-3657.

50. Isel, C, Lanchy, J.M., LeGrice, S.F., Ehresmann, C., Ehresmann, B., and Marquett, R. (1996) *EMBO J* **15**, 917-924.

51. Jacks, T., Townsley, K., Varmus, H.E., and Majors, J. (1987) *Proc Natl Acad Sci USA* **84**, 4298-4302.

52. Johnson, M.S., McClure, M.A., Feng, D.F., Gray, J., and Doolittle, R. F. (1986) *Proc Natl Acad Sci USA* **83**, 7648-7652.

53. Karkas, J.D., Stravrianopoulos, J.G., and Chargaff, E. (1971) *Proc. Natl. Acad. Sci. USA* **69**, 398-402.

54. Katz, R.A. and Skalka, A.M. (1994) *Annu Rev Biochem* **63**, 133-173.

55. Kiyokawa, T., Seiki, M., Iwashita, S., Imagawa, K., Shimizu, F., and Yoshida, M. (1985) *Proc Natl Acad Sci USA* **82**, 8359-8363.

56. Klarmann, G.J., Yu, H., Chen, S., Dougherty, J.P., and Preston, B.D. (1997) *J Virol* **71**, 9529-9269.

57. Komurian, F., Pelloquin, F., and de The G. (1991) *J Virol* **65**, 3770-3778.

58. Kotler, M.R., Katz, A., and Skalka, A.M. (1988) *J Virol* **62**, 2696-2700.

59. Laemmli, U.K. (1970) *Nature* **227**, 680-685.

60. La Grenade, L., Hanchard, B., and Fletcher, V. (1990) *Lancet* **336**, 1345-1347.

61. La Grenade, L., Manns, A., Fletcher, V., Derm, D., Carberry, C., Hanchard, B., Moloney, E.M., Cranston, B., Williams, N.P., Wilks, R., Kang, E.C. and Blattner, W.A. (1998) *Arch Dermatol* **134**, 439-444.

62. Larder, B.H., Kemp, S.D., and Darby, G. (1987) *EMBO J* **6**, 169.

63. Lehky, T.J., Flerlage, N., Katz, D., Houff, S., Hall, W.H., Ishii, K. Monken, C., Ohib-Jalbutt, S., McFarland, H.F., and Jacobson, S. (1996) *Ann Neurol* **40**, 714-723.

64. Leis, J., Ashok, A., and Cobrinik, D. (1993) *in* Reverse Transcriptase (Skalka, A.M. and Goff, S.P. Eds.), pp 33-84, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

65. Levin, J.G., Hatfield, D., Oroszlan, S., and Rein, A. (1993) *in* Reverse Transcriptase (Skalka, A.M., and Goff, S. P., Eds.), p. 5, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

66. Lightfoote, M.M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A., and Venkatesan, S. (1986) *J Virol.* **60**, 771-775.

67. Litvak, S., Sarin-Cottin, L., Fourmier, M. Andreola, M., and Tarrago-Litvak, L. *Trends Biochem Sci* **19**, 114-118.

68. Lloyd, R., Low, B., Godson, N., and Birge, E. A. (1974) *Journal of Bacteriology* **120**, 407-415.

69. Loeb, L.A. Tartof, K.D., and Travaglini, E.C. (1971) *Nature New Biol* **242**, 66-69.

70. Madeleine, M.M., Wiktor, S.Z., Goedert, J.J., Manns, A., Levine, P.H., Biggar, R.T., and Blattner, W.H. (1993) *Int J Cancer* **54**, 255-260.

71. Malik, K.T.A., Even, J., and Karpas, A. (1988) *J Gen Virol* **69**, 1695-1710.

72. Mak, J. and Kleiman, L. (1997) *J Virol* **71**, 8087-8095.

73. Manns, A., Hisada, M., and La Grenade L. (1999) *Lancet* **353**, 1951-1958.

74. Manns, A., Wilks, R.J., Murphy, E.L., Haynes, G., Figueroa, J.P., Barnett, M., Hanchard, B., and Blattner, W.A. (1992) *Int J Cancer* **51**, 886-891.

75. Marquet, R., Isel, C., Ehresmann, C., and Ehresmann, B. (1995) *Biochimie* **77**, 113-124.

76. Melcher, H.V., Reddy, S., and Ruley, H.G. (1990) *Proc Natl Acad Sci USA* **87**, 3733-3737.

77. Miura, T., Fukunaga, T., Igarashi, T., Yamashita, M., Ido, E., Funahashi, S., Ishida, T., Washio, K., Ueda, S., and Hashimoto, K. (1994) *Proc Natl Acad Sci USA* **91**, 1124-1127.

78. MMWR Morb Mortal Wkly Rep (1988) **37**, 736-747.

79. Mochizuki, M., Watanabe, T. Yamaguchi, K., Takatsuki, K., Yoshimura, K., Shirao, M., Nakashima, S., Mori, S., Araki, S., and Miyata, N. (1992) *Jpn J Cancer Res* **83**, 236-239.

80. Mochizuki, M., Watanabe, T., Yamaguchi, K., Yoshimura, K., Nakashima, S., Shirao, M., Araki, S., Takatsuki, K., Mori, S., and Miyata, N. (1992) *Am J Ophthamol* **114**, 123-129.

81. Moore, R., Dixon, M., Smith, R., Peters, G., and Dickson, C. (1987) *J Virol* **61**, 480-490.

82. Mous, S.J., Heimer, E.P., and LeGrice, S.T.J. (1988) *J Virol* **62**, 1433-1436.

83. Nam, S.H., Copeland, T.D., Hatanaka, M., and Oroszlan, S. (1993) *J Virol* **67**, 196-203.

84. Nam, S.H., and Hatanaka, M. (1986) *Biochem Biophys Res Commun* **139**,129-135.

85. Nam, S.H., Kidokoro, M., Shida, H., and Hatanaka, M. (1988) *J Virol* **62**, 3718-3728.

86. Nerurkar, L.S. and Gallo, S. (1989) *Int J Cancer* **Suppl 4**, 2-5.

87. Nogiera, C.M., Cavalcanti, M., Schechter, M. and Ferreira, O.C., Jr. (1996) *Vox Sang* **70**, 47-48.

88. Oger, J.J., Werker, D.H., Foti, D.J., and Dekaban, G.A. (1993) *Can J Neurol Sci* **20**, 302-306.

89. Okochi, K., Sato, H., and Hinuma, Y. (1984) *Vox Sang* **46**, 245-253.

90. Osame, M., Usuku, K., Izumo, S., Ijichi, N., Amitani, H, Igata, A., Matsumoto, M. and Tara, M. (1986) *Lancet* **1**, 1031-1032.

91. Owen, S.M., Lal, R.B., and Ikeda, R.A. (1998) *J Virol* **72**, 5279-5284.

92. Perach, M and Hizi, A. (1999) *Virology* **259**, 176-189.

93. Picard, F.J., Coulthart, M.B., Oger, J.J., King, E.E., Kim, S., Arp, J., Rice, G.P., and Dekaban, G.A. (1995) *J Virol* **69**, 7248-7256.

94. Poch, O., Sauvaget, M., Delarue, M., and Tordo, N. (1989) *EMBO J* **8**, 3867.

95. Poiesz, B.J., Ehrlich, G.D., and Byrns, B.C. (1990) *Med Virol* **9**, 47-75.

96. Poiesz, B.J., Ruscetti, F.W., Gazdar, A. F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980) *Proc Natl Acad Sci USA* **77**, 7415-7419.

97. Polesky, A.H., Steitz, T.A., Grindley, N.O.F., and Joyce, C.M. (1990) *J Biol Chem* **267**, 10169.

98. Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.B., Barr, P.J., and Luciw, P.A. (1986) *Science* **231**, 1567-1572.

99. Preston, B.D., Poiesz, B.J., and Loeb, L. A (1988) *Science* **242**, 1168-71.

100. Ratner, L., Philpott, T. and Trowbridge, D.B. (1991) *AIDS Res Hum Retrov* **7**, 923-941.

101. Rho, H.M., Poiesz, B., Ruscetti, F.W. and Gallo, R.C. (1981) *Virology* **112**, 355-360.

102. Rice, N.R., Stephens, R.M., Burny, A., and Gilden, R.V. (1985) *Virology* **142**, 357-377.

103. Richetti, M. and Buc, H. (1993) *EMBO J* **12**, 387-396.

104. Robert, M.S., Smith, R.G., and Gallo, R.C. (1972) *Science* **176**, 798-800.

105. Roberts, J.D., Bebenek, K., and Kunkel, T.A. (1988) *Science* **242**, 1171-1173.

106. Rosenblatt, J.D., Cann, A.J., Slamon, D.J., Smalberg, I.S., Shah, N.P., Fujii, J., Washman, W., and Chen, D.S. (1988) *Science* **240**, 916-919.

107. Rosenblatt, J.D., Zack, J.A., Chen, I.S.Y., and Lee, H. (1990) *Nat Immun Cell Growth Regul* **9**, 143-149.

108. Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K. Ogawa, Y., and Ikawa, Y. (1985) *Proc Natl Acad Sci USA* **82**, 677-681.

109. Sankar, S. and Porter, A.G. (1992) *J Biol Chem* **267**, 10169.

110. Sardana, V and Sardana, M. (1996) *Meth Enzymol* **275**, 3-16.

111. Sargadharan, M.G., Allaudeen, H.S. and Gallo, R.C. (1976) *Methods in Cancer Res* **12**, 3-47.

112. Sarin, P.S. (1988) *Ann. Neurol* **23**, S181-4.

113. Schaegger, H and von Jagow, G. (1987) *Anal Biochem* **166**, 368-379.

114. Schatz, O., Mous, J., and LeGrice, S.F. (1990) *EMBO J* **9**, 1171-1176.

115. Schwartz, D.E., Tizard, R., and Gilbert, W. (1983) *Cell* **32**, 853-869.

116. Scopes, R.K. (1994) Protein Purification, Principles and Practice, Springer-Verlag New York Inc., New York.

117. Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983) *Proc Natl Acad Sci USA* **80**, 3618-3622.

118. Seiki, M., Hattori, S., and Yoshida, M. (1982) *Proc Natl Acad Sci USA* **79**, 6899-6902.

119. Seiki, M., Inoue, J.L., Hidaka, M., and Yoshida, M. (1988) *Proc Natl Acad Sci USA* **85**, 7124.

120. Sheremata, W.A., Harrington, W.J. Jr., Bradshaw, P.A., Foung, S.K., Raffanti, S.P., Berger, J.R., Snodgrass, S., Resnick, L., and Poiesz, B.J. (1993) *Virus Res* **29**, 71-77.

121. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S., Miwa, M., and Sugimura, T. (1985) *Proc Natl Acad Sci USA* **82**, 3101-3105.

122. Shimoyama, M. (1991) *Br J Haematol* **79**, 428-437.

123. Shinnick, T.M., Lerner, R.A., and Sutcliffe, J.G. (1981) *Nature* **293**, 543-548.

124. Siuzdak, G. (1994) *Proc Natl Acad Sci USA* **91**, 11290.

125. Skalka, A.M. and Goff, S.P. (1993) Reverse Transcriptase, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

126. Smith, C.M., Potts, W.B., Smith, J.S., and Roth, M.J. (1997) *Virology* **229**, 437-446.

127. Sonigo, P., Barker, C., Hunter, E., and Wain-Hobson, S. (1986) *Cell* **45**, 375-385.

128. Terada, K., Katamine, S., Eguchi, K., Moriuchi, R., Kita, M., Shimada, H., Yamshita, I., Iwata, K., Tsuji, Y., and Nagataki, S. (1994) *Lancet* **344**, 1116-1119.

129. Tesara, T. Amacker, M., and Hubscher, U. (1999) *Biochemistry* **38**, 1633-1642.

130. Travaglini, E.C., Duke, D.K., Surrey, S., and Loeb, L.A. (1976) *J Mol Biol* **106**, 605-621.

131. Trentin, B., Rebeyrotte, N., and Mamoun, R.Z. (1998) *J Virol* **72**, 6504-6510.

132. Triglia, T., Peterson, M.G., and Kemp, D.J. (1988) *Nucleic Acids Res* **16**, 8186.

133. Uchiyama, T. (1997) *Ann Rev Immunol* **15**, 15-37.

134. Uchiyama, T., Yodoi, J., Sagawind, H K., Takatsuki, K., and Uchino, H. (1977) *Blood* **50**, 481-492.

135. Veronese, F., Copeland, T.D., DeVico, A.L., Rahman, R., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1986) *Science* **231**, 1289-1291.

136. Weiss, R.A. (1987) *J Clin Pathol* **40**, 1004-1069.

137. Weiss, R. (1984) *in* RNA Tumor Viruses, Vol. 1 (Weiss, R., Teich, N., Varmus, H., and Coffin, J. Eds.), pp 1244-1253, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

138. Weiss, R. (1984) *in* RNA Tumor Viruses, Vol. 2 (Weiss, R., Teich, N., Varmus, H., and Coffin, J. Eds.), pp 406-408, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

139. Whitaker, J.R. (1963) *Anal Chem* **35**, 1950.

140. White, J.D., Zaknoen, S.L., Kastern-Sportes, C., Top, L.E., Navarro-Roman, L., Nelson, D.L., and Waldmann, T.A. (1995) *Cancer* **75**, 1598-1607.

141. Whittaker, S.J., Ng, Y.L., Rustin, M., Levine, G., McGibbon, G.H., and Smith, N.P. (1993) *Br J Dermatol* **128**, 483-492.

142. Wiktor, S.Z., Pate, E.J., Murphy, E.L., Palker, T.J., Champegnie, E., Ramlal, A., Cranston, G., Hanchard, B., and Blanchard, W.A. (1993) *J Acquir Immune Defic Syndr* **6**, 1162-1167.

143. Wiktor, S.Z., Piot, P., Mann, J.M., Nzilambi, N, Francis, H., Vercauteren, G., Blattner, W.A., and Quinn, T.C. (1990) *J Infect Dis* **16**, 1073-1077.

144. Williams, A.E., Fang, C.T. Slamon, D.J., Poiesz, B.J., Sandler, S.G., Darr, W.F., Shulman, G., McGowan, E.I., Douglas, D.K., and Bowman, R.G. (1988) Science **240**, 643-646.

145. Wodarz, D., Nowak, M.A., and Bangham, C.R.M. (1999) *Immunol Today* **20**, 220-227.

146. Wöhrl, B.M. and Moelling, K. (1990) *Biochemistry* **29**, 10141-10147.

147. Yamaguchi, K., Matsuoka, M., Takemoto, S., Tamiya, S., Etoh, K., and Takatsuki, K. (1996) *Intervirology* **39**, 158-164.

148. Yamaguchi, K., Seiki, M., Hoshida, M., Nishimura, H., Kawano, F., and Takatsuki, K. (1984) *Blood* **63**, 1235-1240.

149. Yamashita, I., Iwata, K., Tsuji, Y., and Nagataki, S. (1994) *Lancet* **344**, 1116-1119.

150. Yanagihara, R., Jenkins, C.L., Alexander, S.S., Mora, C.A., and Garruto, R.M. (1990) *J Infect Dis* **162**, 649-654.

151. Yasui, C., Fukaya, T., Koizumi, H., Kobayashi, H., and Ohkanara, A. (1991) *J Am Acad Dermatol* **24**, 633-637.

152. Yip, M.T., Dynan, W.S., Green, P.L., Black, A.C., Arrigo, S.J., Torbati, A., Heaphy, S., Rutland, C., Rosenblatt, J.D., and Chen, I.S. (1991) *J Virol* **65**, 2261.

153. Yoshinaka, Y., Katoh, I., Copeland, T.D., and Oroszlan S. (1985) *Proc Natl Acad Sci USA* **82**, 1618-1622.

154. Yoshinaka, Y., Katoh, I., Copeland, T.D., Smythers, G.W., and Oroszlan S. (1986) *J Virol* **57**, 826-832.

155. Zhao, L.J., and Giam, C.Z. (1991) *Proc Natl Acad Sci USA* **88**, 11445-11449.

APPENDIX A

Sepharose Standard Curve

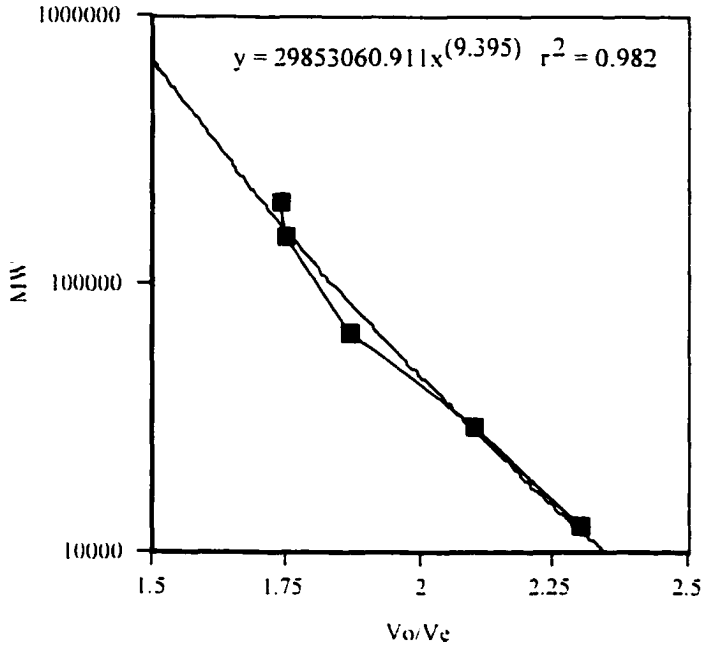


FIG. 25. Standardization of sepharose CL-6B gel filtration column (52 x 5 cm i.d.). The ratio of elution volumes (V_e) to the elution volume (V_o) of blue dextran (M_r = 2,000,000) were plotted against the molecular weights of the standards. Standards: β-amylase (M_r = 200,000), alchohol dehydrogenase (M_r = 150,000), albumin (M_r = 66,000, carbonic anyhydrase (M_r = 29,000 and cytochrome C (M_r = 12,000)

APPENDIX B

TABLE VII

Summary of Preliminary Trials with Chromatographic Resins

Resin	Technique	Problems with Method
DEAE	anion exchanger	did not stick to column; pass fraction is not very clean
heparin sepharose	pseudo-affinity	inconsistency in results; activity drops ≈ 5-fold each preparation;does not consistently appear in Western Blot
phenyl sepharose	hydrophobic interaction	did not stick to column;not much cleaner;never appeared on Western Blot
DNA cellulose	ligand affinity	did not appear on Western Blot; activity dropped 6-fold

APPENDIX C

Sequence of the Protease Construct

```
                10          20          30          40          50
ACC GGG GGA GGT TTA AAC TCC CCC CCC ACA TTA CAG CAA GTC CTT CCT AAC CAA
TCC CCC CCT CCA AAT TGG AGG GGG GGG TGT AAT GTC GTT CAG GAA GGA TTG GTT
Arg Gly Gly Gly Leu Thr Ser Pro Pro Thr Leu Gln Gln Val Leu Pro Asn Gln

      60          70          80          90          100
GAC CTA GCA TCT ATT CTG CCA GTT ATA CCG TTA GAT CCC GCC CGT CGG CCC
CTG GAT CGT AGA TAA GAC GGT CAA TAT GGC AAT CTA GGG CGG GCA GCC GGG
Asp Leu Ala Ser Ile Leu Pro Val Ile Pro Leu Asp Pro Ala Arg Arg Pro

    110         120         130         140         150
GTA ATT AAA GCC CAG GTT GAC ACC CAG ACC AGC CAC CCA AAG ACT ATC GAA GCT
CAT TAA TTT CGG GTC CAA CTG TGG GTC TGG TCG GTG GGT TTC TGA TAG CTT CGA
Val Ile Lys Ala Gln Val Asp Thr Gln Thr Ser His Pro Lys Thr Ile Glu Ala

160         170         180         190         200         210
TTA CTA GAT ACA GGA GCA GAC ATG ACA GTC CTT CCA ATA GCC TTG TTC TCA ATT
AAT GAT CTA TGT CCT CGT CTG TAC TGT CAG GAA GGT TAT CGG AAC AAG AGT TTA
Leu Leu Asp Thr Gly Ala Asp Met Thr Val Leu Pro Ile Ala Leu Phe Ser Asn

      220         230         240         250         260
AAT ACT CCC CTC AAA GAC ACA TCC GTA TTG GGG GCA GGA GGC CAA ACC CAA GAT
TTA TGA GGG GAG TTT CTG TGT AGG CAT AAC CCC CGT CCT CCG GTT TGG GTT CTA
Asn Thr Pro Leu Lys Asp Thr Ser Val Leu Gly Ala Gly Gly Gln Thr Gln Asp

270         280         290         300         310         320
CAC TTT AAG CTC ACC TCC CTT CCT GTG CTA ATA CGC CTC CCT TTC CGG ACA ACG
GTG AAA TTC GAG TGG AGG GAA GGA CAC GAT TAT GCG GAG GGA AAG GCC TGT TGC
His Phe Lys Leu Thr Ser Leu Pro Val Leu Ile Arg Leu Pro Phe Arg Thr Thr

        330         340         350         360         370
CCT ATT GTT TTA ACA TCT TGC CTA GTT GAT ACC AAA AAC AAC TGG GCC ATC ATA
GGA TAA CAA AAT TGT AGA ACG GAT CAA CTA TGG TTT TTG TTG ACC CGG TAG TAT
Pro Ile Val Leu Thr Ser Cys Leu Val Asp Thr Lys Asn Asn Trp Ala Ile Ile

    380         390         400         410         420
GGT CGT GAT GCC TTA CAA CAA TGC CAA GGC GTC CTG TAC CTC CCT GAG GCA AAA
CCA GCA CTA CGG AAT GTT GTT ACG GTT CCG CAG GAC ATG GAG GGA CTC CGT TTT
Gly Arg Asp Ala Leu Gln Gln Cys Gln Gly Val Leu Tyr Leu Pro Glu Ala Lys

430         440         450         460         470         480
AGG CCG CCT GTA ATC TTG CCA ATA CAG GCG CCA GCC GTC CTT GGC CTA GAA
TCC GGC GGA CAT TAG AAC GGT TAT GTC CGC GGT CGG CAG GAA CCG GAT CTT
Arg Pro Pro Val Ile Leu Pro Ile Gln Ala Pro Ala Val Leu Gly Leu Glu
```

FIG. 26. Protease construct created by amplifying the *pro* ORF from pRCH. Primers were chosen to create unique NdeI and Bam HI sites to allow subsequent cloning in frame with the N-terminal His tag coding sequence.

APPENDIX D

Mass Spectra of Actual Peptides from Pro-Pol Fragment

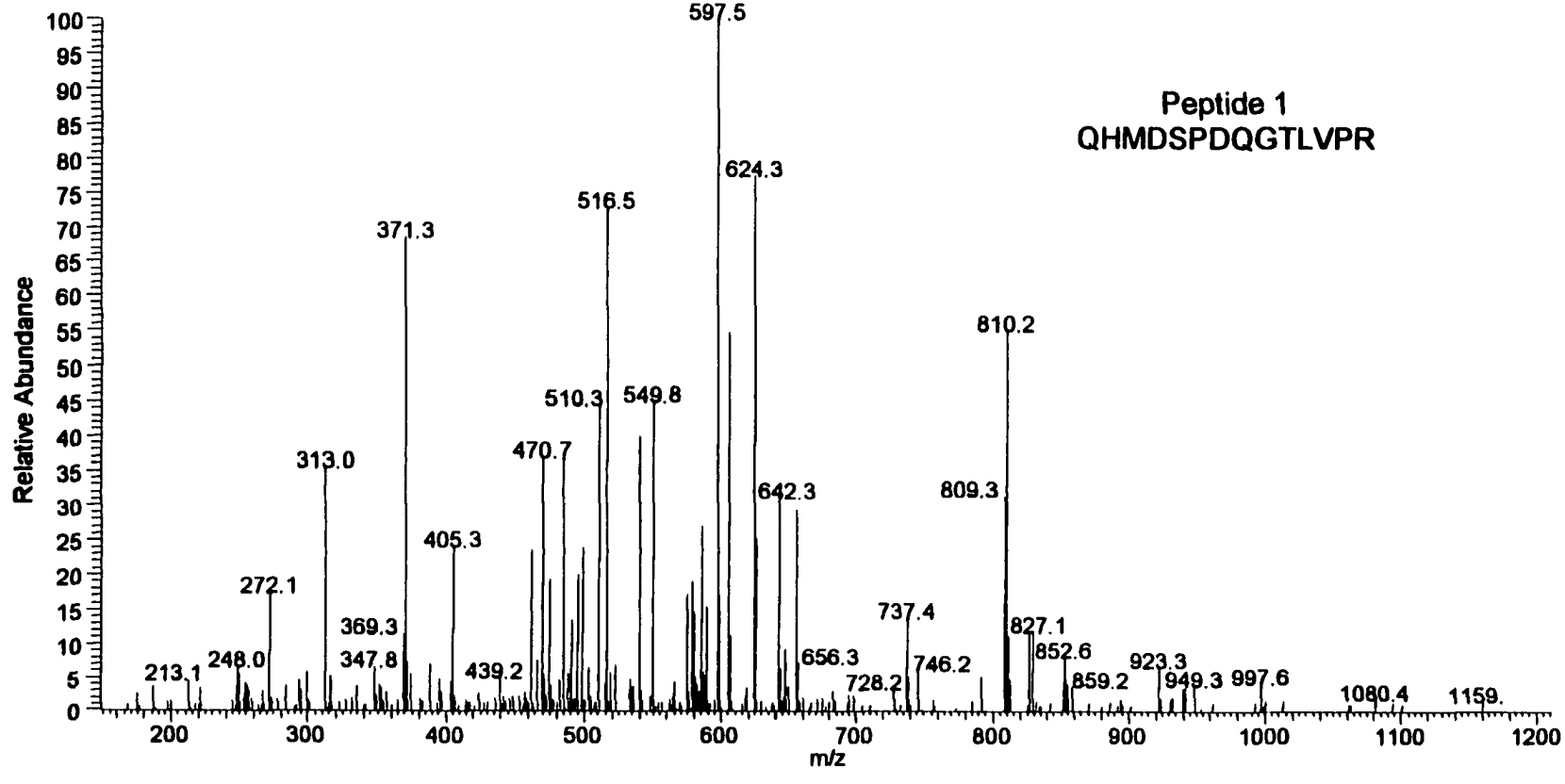


Fig. 27. MS analysis of Peptide 1after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, QHMDSPDQGTLVPR, is 1581.2 daltons.

APPENDIX D, Continued

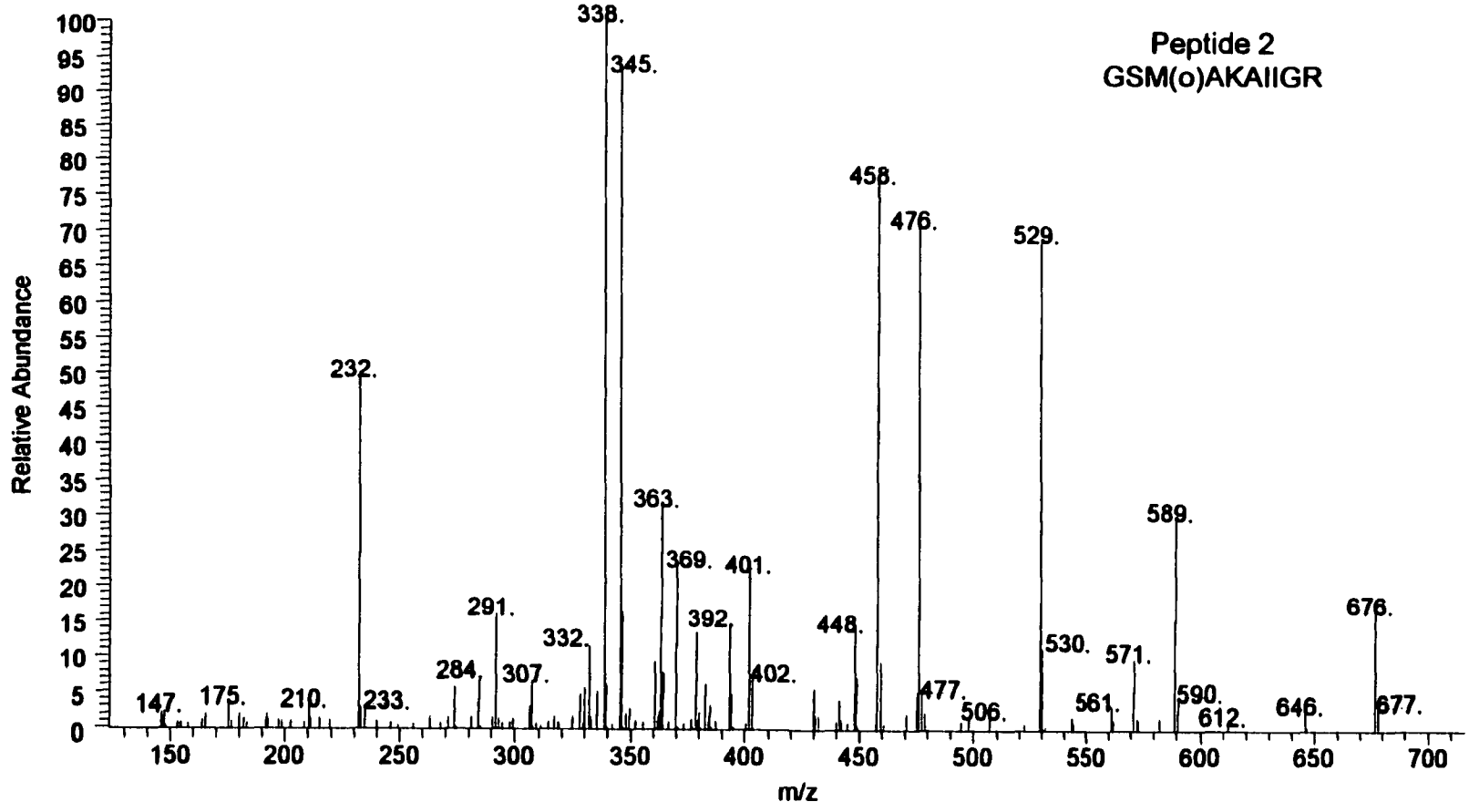


Fig. 28. MS analysis of Peptide 2 after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, GSM(o)AKIIGR, is 820.3 daltons.

APPENDIX D, Continued

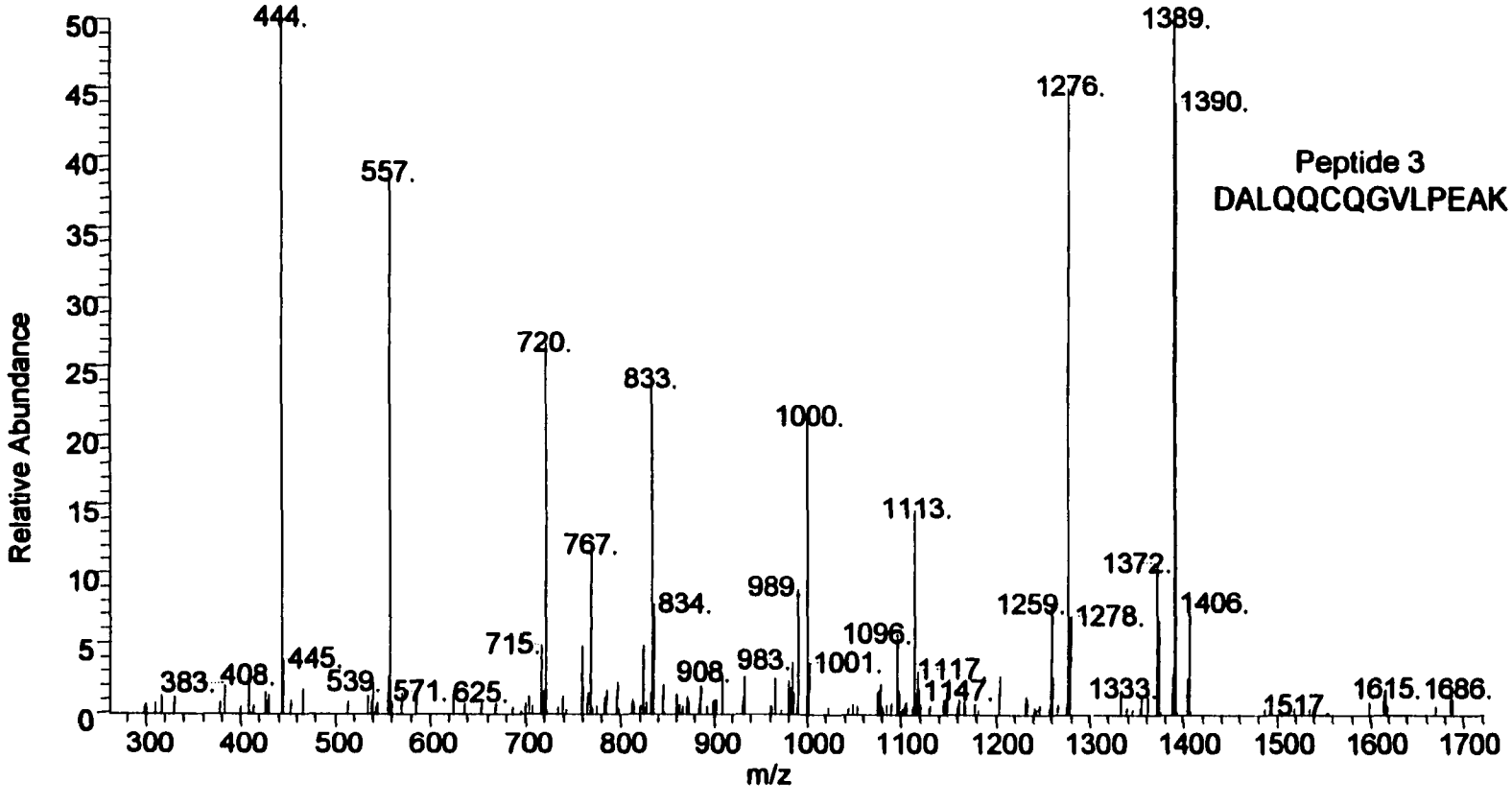


Fig. 29. MS analysis of Peptide 3 after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, DALQQCQGVLYLPEAK, is 1833.1 daltons.

APPENDIX D, Continued

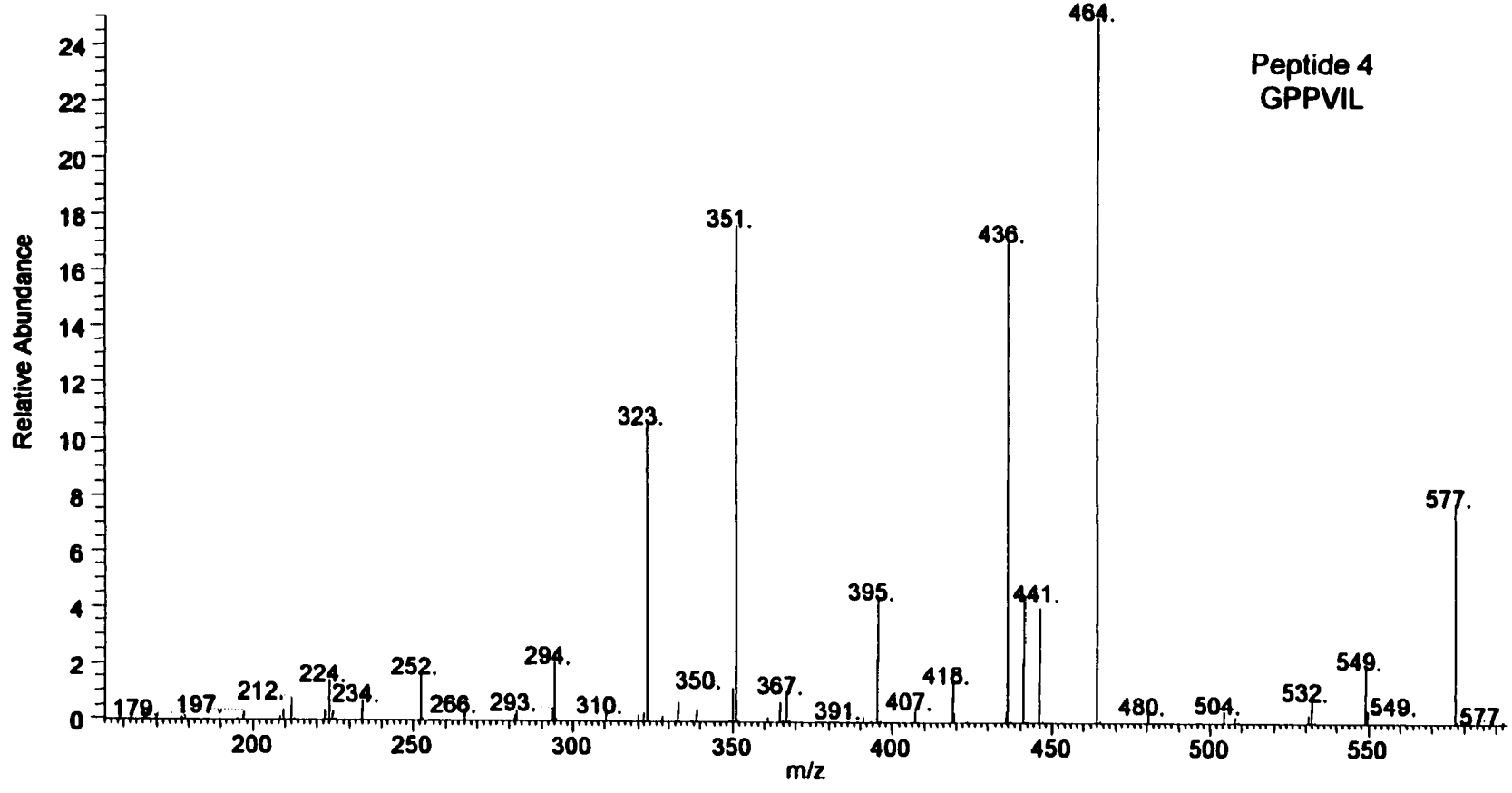


Fig. 30. MS analysis of Peptide 5 after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, GPPVIL, is 595.3 daltons.

APPENDIX D, Continued

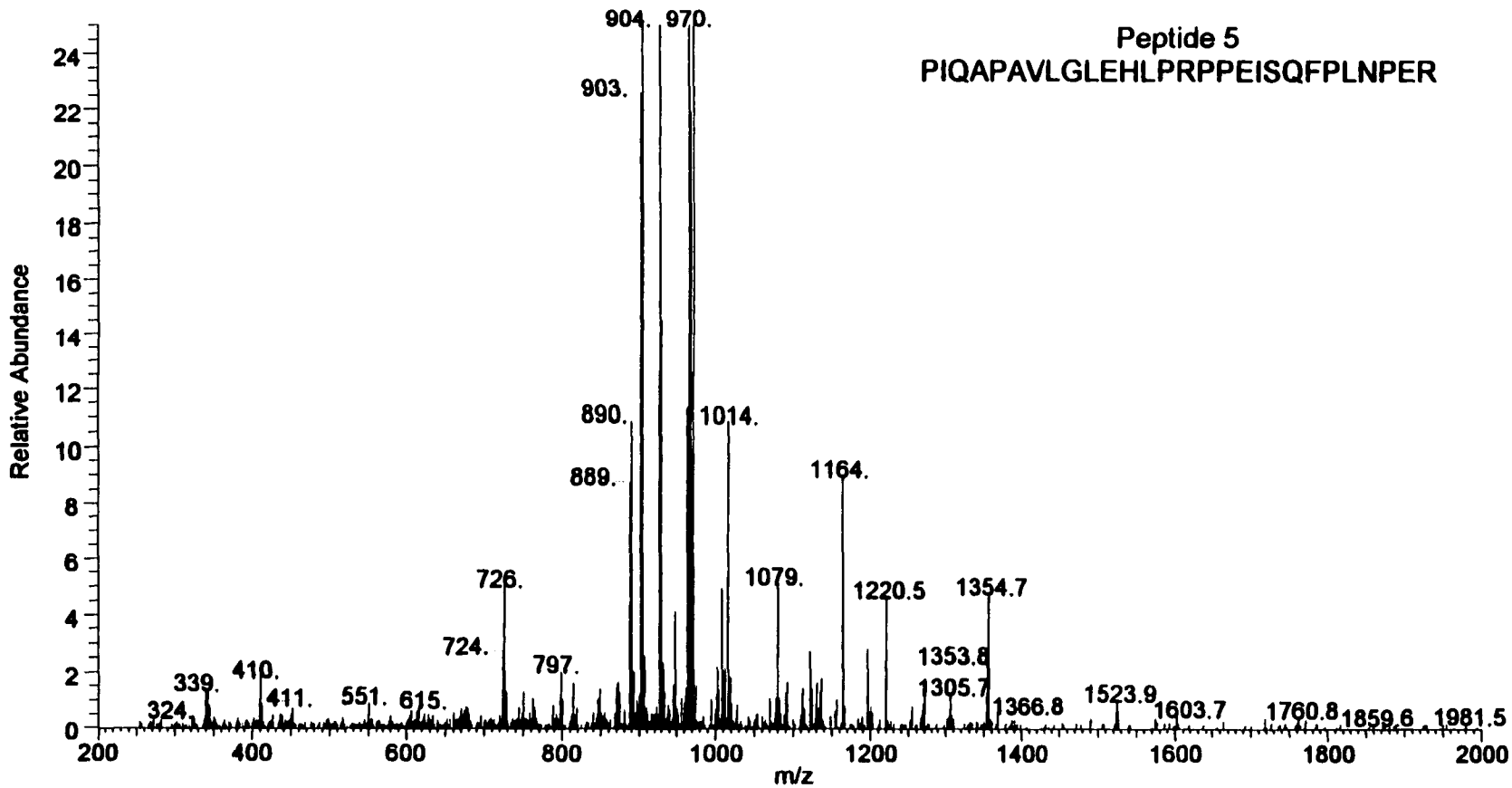


Fig. 31. MS analysis of Peptide 5 after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, PIQAPAVLGLEHLPRPPEISQFPLNPER, is 3116.3 daltons

APPENDIX D, Continued

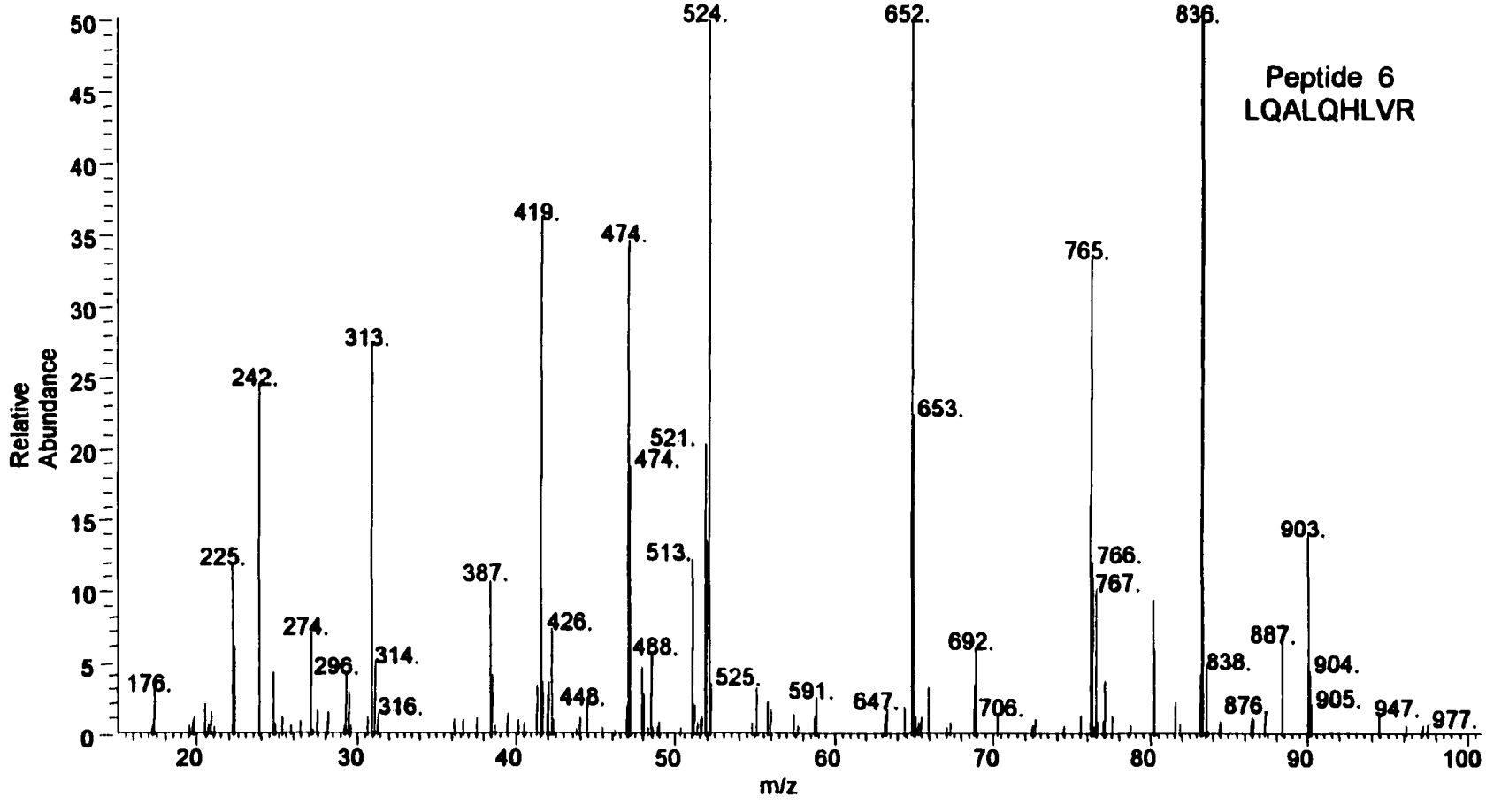


Fig. 32. MS analysis of Peptide 6 after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, LQALGHLVR, is 1077.6.

APPENDIX D, Continued

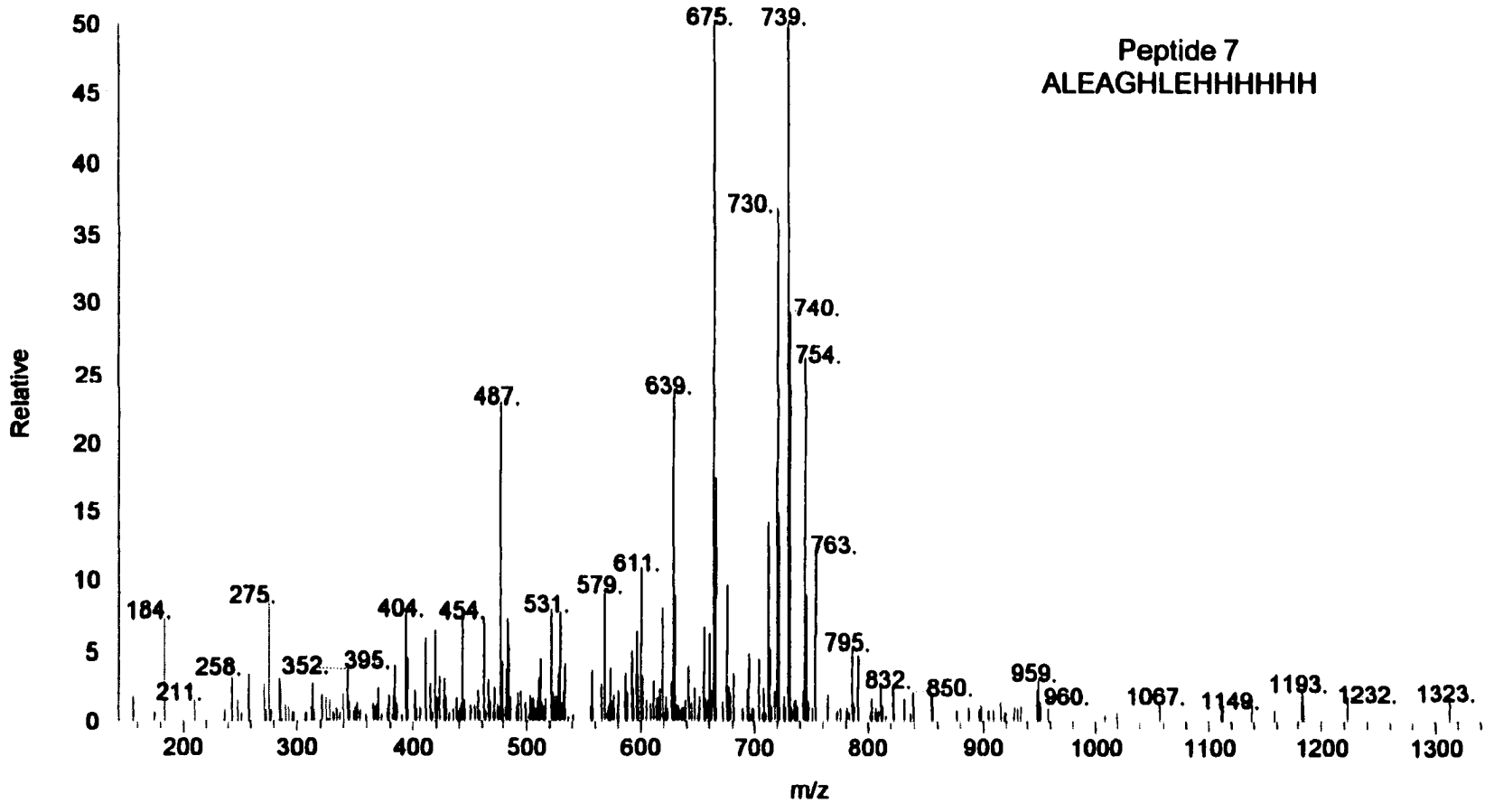


Fig. 33. MS analysis of Peptide 1after tryptic digestion of Pro-Pol protein fragment. The M_r of the sequence, ALEAGHLEHHHHHH is 1661.8 daltons.

VITA

PINKY GUNDAYAO AGBUYA
3927 Seeman Road
Virginia Beach, VA 23452

EDUCATION

Ph.D. Biomedical Sciences, Anticipated May 2000
Dissertation Title: Characterization of Human T-cell Leukemia Virus I Reverse Transcriptase
Old Dominion University and Eastern Virginia Medical School.
Department of Chemistry and Biochemistry, Norfolk, VA

B.S. Pharmacology and Toxicology, May 1992
Philadelphia College of Pharmacy and Science, Philadelphia, PA

EXPERIENCE

Graduate Research Assistant, Department of Chemistry and Biochemistry, Old Dominion University, 1996-2000

Graduate Research Assistant, Pediatric Research Center, Eastern Virginia Medical School, Norfolk, VA, 1994-1995

Graduate Teaching Assistant, Department of Chemistry and Biochemistry, Old Dominion University , 1994-1996, 2000

Toxicology Research Technician, E.I. DuPont de Nemours and Company, Haskell Laboratory for Toxicology and Industrial Medicine, Newark, DE 1992-1993

Undergraduate Research Assistant, Philadelphia College of Pharmacy and Science, Philadelphia, PA 1991-1992

Undergraduate Research Assistant, Department of Pharmacology, Eastern Virginia Medical School, Norfolk, VA, Summer 1991 and December 1991

PUBLICATIONS

Agbuya, P.G., Sherman, N.E. and Moen, L.K. (2000) Proteolytic processing of the HTLV-1 Reverse Transcriptase: identification of the N-terminal cleavage site by mass spectrometry, submitting for publication in Archives of Biochemistry and Biophysics

Agbuya, P.G., and Moen, L.K. (1998) Identification of the Coding Sequence of HTLV-1 Reverse Transcriptase, FASEB Journal **12**:A1447

Agbuya, P.G., Li, L., Zaritsky, A.L., Morris, A.D., and Miles, M.V. (1996) Development of a fluorescence polarization immunoassay for lorazepam quantification, Therapeutic Drug Monitoring **18**: 194-199.