Design and characterization of HIV-1 gp120 mimetic peptides
by Alex Charles Johnson

A thesis submitted in partial fulfillment of the requirements for the degree of' Master of Science in Biochemistry
Montana State University
© Copyright by Alex Charles Johnson (1999)

Abstract:
The goal of this research project is to design constrained peptides which mimic the CD4 binding site of HIV-1 gpl20 for use as vaccine antigens. Using the recently elucidated crystal structure of gpl20 in complex with CD4, two peptides were designed using molecular modeling software. The first peptide, Peptide A, is an 18-mer mimicking the gpl20 β20/β21 hairpin loop (from residues 420-437) in which amino acids 420 and 437 have been replaced by cysteine residues and oxidized to form a disulfide bond. The second peptide, Peptide B, is a 27-mer representative of a discontinuous epitope of gpl20 and consists of residues 365-373 and 422-436 joined by three glycine residues. Residues 422 and 436 have been replaced with cysteines and oxidized to form a disulfide bond.

Peptide A was synthesized on an automated peptide synthesizer using standard F-moc chemistry, then N-acetylated, air-oxidized to form the intramolecular disulfide bond, and purified by preparative HPLC. Peptide B was produced by GeneMed Synthesis, Inc. using similar methods. Both peptides were further characterized by MALDI-TOF mass spectrometry, analytical HPLC, and circular dichroism. Peptide A was tested for binding to CD4 via a competitive ELISA assay with gpl20. Peptide B was tested for binding to a CD4-IgG2 chimera via a competitive ELISA assay with gpl20 and tested for HIV neutralization activity using a focal infectivity assay with CD4+ HeLa cells.

It was found that Peptide A had 47% β strand and 48% random coil by CD, indicating that it was likely mimicking the intended gpl20 structure of the β20/β21 hairpin loop. However, it did not bind CD4 in ELISA assays with any greater affinity than an irrelevant peptide. This may be due to insufficient binding interactions between Peptide A and CD4.

Peptide B was found to have 47% β strand and 48% random coil by CD, consistent with the intended portion of gp120's CD4 binding site. This peptide did not inhibit binding of CD4-IgG2 to ELISA plates coated with gpl20, but instead appeared to bind specifically to the D1/D2 domains of CD4-IgG2, resulting in non-specific binding of CD4-IgG2 to the ELISA plate wells, whether or not gpl20 was present. An assay for the HIV-1 p24 protein on HeLa CD4+ cells used in the focal infectivity assay suggested that Peptide B blocked HTV infection at micromolar concentrations.

DESIGN AND CHARACTERIZATION OF HIV-1 GP120 MIMETIC PEPTIDES

by

Alex Charles Johnson

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Biochemistry

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

August 1999

N378
J63

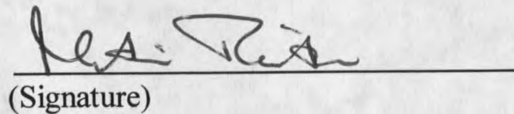
APPROVAL

of a thesis submitted by

Alex Charles Johnson

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

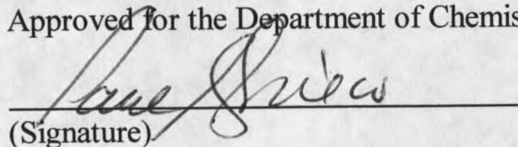
Martin Teintze



(Signature) 7/29/99 Date

Approved for the Department of Chemistry and Biochemistry

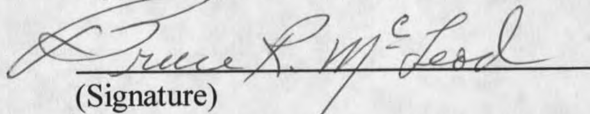
Paul Grieco



(Signature) 7-29-99 Date

Approved for the College of Graduate Studies

Bruce McLeod

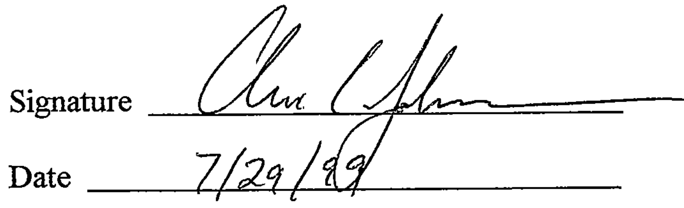


(Signature) 7-30-99 Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.



Signature

Date 7/29/99

ACKNOWLEDGMENTS

There are several people to whom I wish to give thanks concerning the completion of this work. First, I would like to thank my advisor, Martin Teintze, for his guidance throughout the project and his unwavering support of my goals. I would also like to thank my other committee members, Ed Dratz and Jean Starkey, for their insightful comments and continual support.

During the course of this project, I have had the courtesy of working with others in the Chemistry and Biochemistry department and the Microbiology department. I would like to thank Steve Arbogast for his help during peptide synthesis, HPLC, and MALDI-TOF; Hua Fang for his guidance through the ELISA techniques; John Mills for his help with HPLC; Tami Peters for taking me under her wing and starting me off in the right direction; and Dave Wooster, Paul Kraft, Brian Bailey, and John Jacobs whose brains I picked endlessly for suggestions and knowledge throughout the project.

Finally, a fond thanks also goes to Dallas and Ann Johnson who have given me a home away from home. I would especially like to thank my wife, Stephanie, for the pillar of support, patience, and encouragement she has given me over the course of this study.

TABLE OF CONTENTS

Page

1. AN INTRODUCTION TO AIDS, HIV, AND VACCINE DEVELOPMENT . . . 1

Discovery of AIDS . . . 1
Overall HIV Structure and Method of Infection . . . 2
Gp120 Structure . . . 6
Gp120 Interaction with CD4 . . . 8
Vaccine Development . . . 10
Statement of Problem . . . 12
References Cited . . . 14

2. PEPTIDE DESIGN, PURIFICATION, AND ANALYSIS . . . 16

Previous Attempts to Create Peptide Vaccines . . . 16
Peptide Design . . . 21
Circular Dichroism . . . 28
Materials and Methods . . . 29
Molecular Modeling . . . 29
Peptide Synthesis . . . 31
MALDI-TOF Mass Spectrometry . . . 32
HPLC Purification . . . 33
Ellman Assay . . . 35
Circular Dichroism . . . 35
Results and Discussion . . . 36
Peptide A . . . 36
Peptide B . . . 36
References Cited . . . 44

3. PEPTIDE ACTIVITY DETERMINATION BY ELISA AND FOCAL INFECTIVITY ASSAY . . . 47

ELISA . . . 47
Focal Infectivity Assay . . . 48
Materials and Methods . . . 49
Peptide A ELISA . . . 49
Peptide B ELISA . . . 51
Focal Infectivity Assay . . . 54

TABLE OF CONTENTS (continued)

Page

Results and Discussion 56
  Peptide A 56
  Peptide B 57
References Cited 60

APPENDIX – Input File for Discover Generated by Insight 62

LIST OF TABLES

Table Page

1. Acetonitrile gradient used for Peptide A. 34

2. Acetonitrile gradient used for Peptide B 34

3. The binding energies of gp120 and Peptide B to CD4. 41

4. Peptide A ELISA at 90 min. 56

5. Peptide B ELISA #1 at 90 min. 57

6. Peptide B ELISA #2 at 60 min. 58

7. Peptide B p24 Assay 59

LIST OF FIGURES

Figure — Page

1. An HIV retrovirus 4
2. The proposed process of infection of a T-cell by HIV 5
3. The amino acid sequence of gp120 7
4. Ribbon diagram of the gp120 core x-ray crystal structure 9
5. Schematic of synthetic, branched, 47-mer peptide 20
6. The crystal structure of gp120's CD4 binding cavity 22
7. Schematic representation of Peptide A 24
8. A molecular modeled structure of Peptide A 25
9. Schematic representation of Peptide B 26
10. A molecular modeled structure of Peptide B bound to CD4 27
11. MALDI-TOF of Peptide A 37
12. CD Spectrum of Peptide A 38
13. MALDI-TOF of Peptide B 42
14. CD Spectrum of Peptide B 43

ABSTRACT

The goal of this research project is to design constrained peptides which mimic the CD4 binding site of HIV-1 gp120 for use as vaccine antigens. Using the recently elucidated crystal structure of gp120 in complex with CD4, two peptides were designed using molecular modeling software. The first peptide, Peptide A, is an 18-mer mimicking the gp120 β20/β21 hairpin loop (from residues 420-437) in which amino acids 420 and 437 have been replaced by cysteine residues and oxidized to form a disulfide bond. The second peptide, Peptide B, is a 27-mer representative of a discontinuous epitope of gp120 and consists of residues 365-373 and 422-436 joined by three glycine residues. Residues 422 and 436 have been replaced with cysteines and oxidized to form a disulfide bond.

Peptide A was synthesized on an automated peptide synthesizer using standard F-moc chemistry, then N-acetylated, air-oxidized to form the intramolecular disulfide bond, and purified by preparative HPLC. Peptide B was produced by GeneMed Synthesis, Inc. using similar methods. Both peptides were further characterized by MALDI-TOF mass spectrometry, analytical HPLC, and circular dichroism. Peptide A was tested for binding to CD4 via a competitive ELISA assay with gp120. Peptide B was tested for binding to a CD4-IgG2 chimera via a competitive ELISA assay with gp120 and tested for HIV neutralization activity using a focal infectivity assay with $CD4^+$ HeLa cells.

It was found that Peptide A had 47% β strand and 48% random coil by CD, indicating that it was likely mimicking the intended gp120 structure of the β20/β21 hairpin loop. However, it did not bind CD4 in ELISA assays with any greater affinity than an irrelevant peptide. This may be due to insufficient binding interactions between Peptide A and CD4.

Peptide B was found to have 47% β strand and 48% random coil by CD, consistent with the intended portion of gp120's CD4 binding site. This peptide did not inhibit binding of CD4-IgG2 to ELISA plates coated with gp120, but instead appeared to bind specifically to the D1/D2 domains of CD4-IgG2, resulting in non-specific binding of CD4-IgG2 to the ELISA plate wells, whether or not gp120 was present. An assay for the HIV-1 p24 protein on HeLa $CD4^+$ cells used in the focal infectivity assay suggested that Peptide B blocked HIV infection at micromolar concentrations.

CHAPTER ONE

AN INTRODUCTION TO AIDS, HIV, AND VACCINE DEVELOPMENT

Discovery of AIDS

In 1981, a rare form of pneumonia named *Pneumocystis carinii* began to appear primarily in the homosexual male population of California. This infection is an oddity in healthy individuals, but patients with immune deficiencies, commonly caused by lymphatic cancers, become susceptible to the disease. In addition to pneumonia, other opportunistic infections caused by various bacteria, viruses, fungi, and protozoa also began to take hold in the same population, and a rare form of skin cancer named *Kaposi's sarcoma* was also concomitant with many of these cases. This set of symptoms became the basis for a new syndrome called *Acquired Immune Deficiency Syndrome* (AIDS), so named for its destructive impact on the immune system which allows the development of the above listed symptoms (1). Humoral and cell-mediated responses to the virus fail to halt progression of the disease. With the exception of a few relatively rare "long term non-progressors," some of which lack proteins integral to the viral infection mechanism(2), to date the mortality rate of AIDS patients without effective treatment is 100%.

Soon after, the infectious pathway for AIDS was found to be blood-linked as well as transmissible by certain bodily fluids, most notably by semen. By 1983, the pathogenic factor was determined to be a previously uncharacterized retrovirus which, after some

debate, was named the *human immunodeficiency virus* (HIV) in 1986 (2).

HIV's complexity and high mutation rate have hindered research efforts towards effective treatment of AIDS, as well as hindered vaccine development. AIDS has since spread into the heterosexual population, mainly through needle sharing of intravenous drug users, and as of 1993, it was thought that 1 of every 250 persons in the United States was infected with HIV (2). Today, there are between 40,000 to 60,000 new cases of HIV infection in the United States each year, and the number of individuals who die from AIDS is 35,000 to 40,000 annually (1). The United States, however, has suffered relatively lightly compared to the remainder of the world. Worldwide, AIDS now ranks as the third leading cause of death after malaria and tuberculosis (1). There are an estimated 13 million HIV positive individuals, and this number is expected to rise to 30 to 100 million HIV infected persons in the first decade of the next century (1, 2). This exponential growth combined with the lethality of the virus make finding a vaccine for HIV critical.

Overall HIV Structure and Method of Infection

There are two major types of HIV: HIV-1 and HIV-2. Although HIV-2 differs from HIV-1 by more than 55% and is actually closer in primary sequence to *simian immunodeficiency virus* (SIV), it also causes AIDS and appears to infect the host by a similar mechanism as HIV-1 (2). Since HIV-1 is responsible for the vast majority of AIDS infected persons worldwide and was used as the template for these peptide studies, it will be the focus of this discussion.

HIV-1 is a member of one of the five major primate lineages of the lentivirus

family of retroviruses (3). It has a cone-shaped core, composed of the viral p24 Gag capsid protein and the p6 Gag nucleocapsid protein, which contains two identical RNA strands closely associated with reverse transcriptase (RT) and the nucleocapsid Gag protein p7 (Figure 1). A myristolated p17 core Gag protein provides structure for the viral matrix. The RNA strands code for *gag* (Gag capsids proteins), *pol* (RT, protease, and integrase), and *env* (gp160) proteins as well as the unique accessory genes *tat, rev, nef, vif, vpu,* and *vpr* (2). The surface of the protein consists of a lipid bilayer containing up to 72 knobs of trimers of the envelope glycoprotein (2, 4-6). This envelope structure is derived from gp160, a glycosylated 160-kDa precursor, which is cleaved inside the infected cell's Golgi apparatus into a gp120 external surface envelope protein and a gp41 transmembrane protein (6). The virion gp120, which is located on the virus surface and is held to the transmembrane gp41 by noncovalent interactions, contains the binding sites for cellular receptors and the major antibody neutralizing domains (7).

To initiate a new infection, HIV must encounter a potential target cell that expresses the appropriate proteins (Figure 2). In HIV's case, viral entry begins with the recognition of CD4 on T-cells by the viral gp120 envelope protein (8). CD4 is also expressed, although in smaller quantities, on monocytes, dendritic cells, and brain microglia, and these cells are also subject to HIV-1 infection (6). The CD4 binding site on gp120 is transiently occluded by gp120's V1 and V2 loops (see loop descriptions below) (8). Although this may act to help foil the host's immune system, gp120 can still bind to CD4, since the binding site probably fluctuates between an open and closed conformation (8). Following docking to the CD4 receptor, a conformational change in the viral

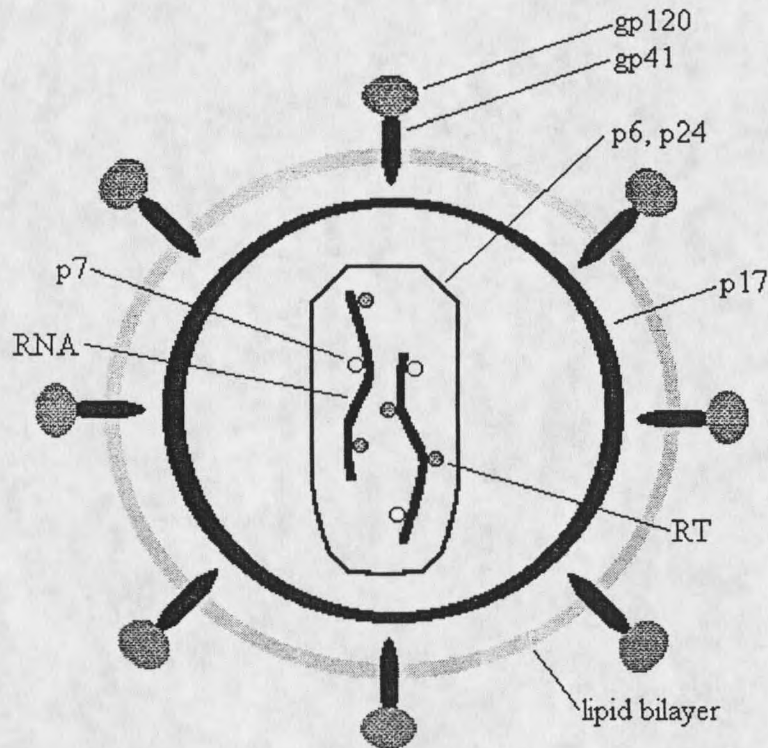


Figure 1. An HIV retrovirus with the structural and other viral proteins identified. Gp120 is shown as a monomer for simplification.

envelope protein's V3 loop allows binding to an obligate second coreceptor, CXCR-4 or CCR-5 (for T-cells and macrophages respectively) (2, 9, 10). Coreceptor binding, in turn, causes a chain of events which allows viral entry into the cell through fusion of the viral membrane with the host cell membrane, a process mediated by gp41, permitting the viral RNA and related machinery to invade the host cell (2, 8). Interestingly, a strain of HIV-2 can fuse efficiently in a CD4-independent, coreceptor-dependent manner (5). This route of entry, however, is quite limited compared with the process mediated by CD4. It is possible that CD4 merely enhances the interaction of gp120 with a coreceptor and this method of viral entry is less efficient in $CD4^-$ cells (2).

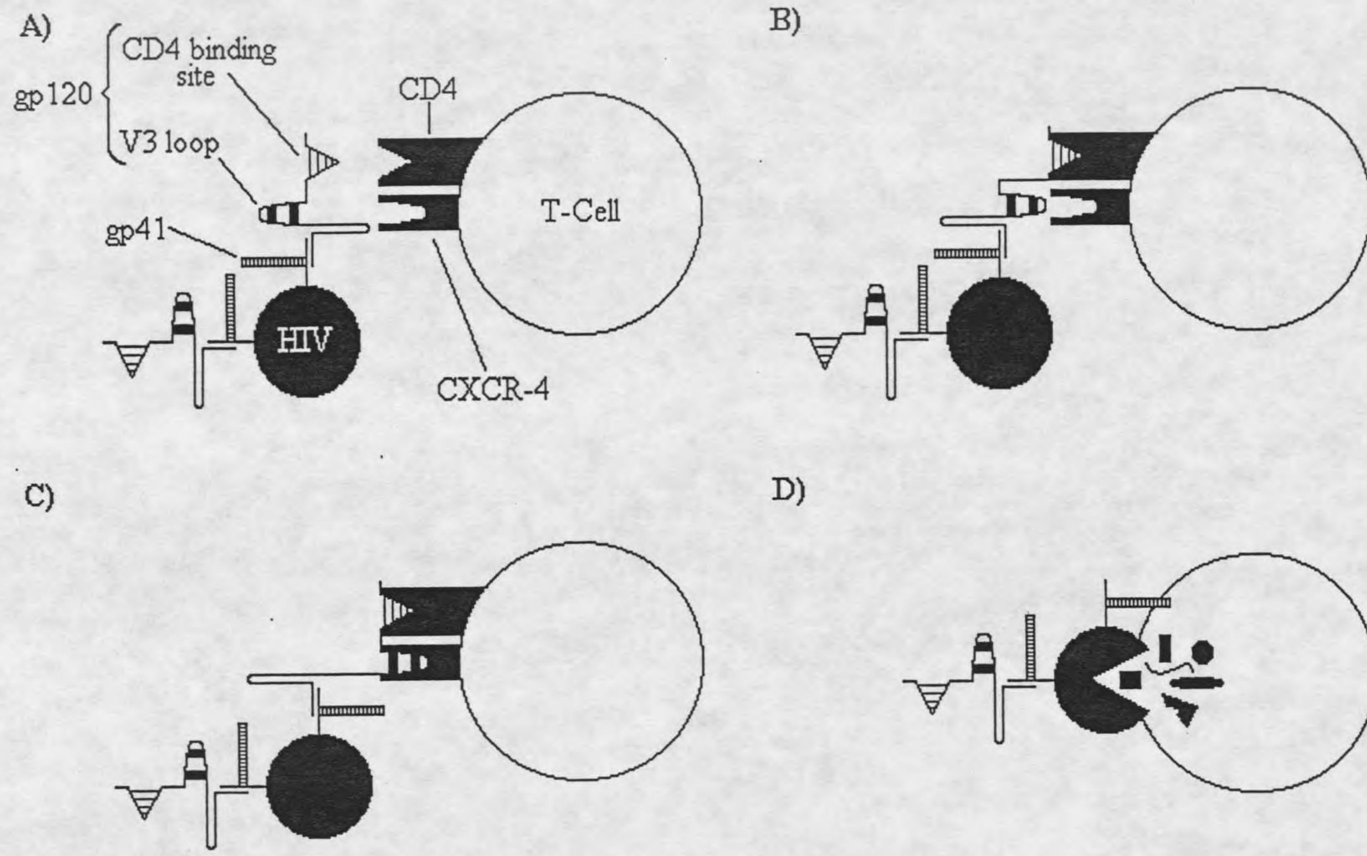


Figure 2. The proposed process of infection of a T-cell by HIV. (A) HIV approaches the cell, and the gp120 protein on the HIV surface interacts with the CD4 receptor on the target cell. (B) Upon gp120 binding to CD4, a conformational change in the viral V3 loop allows binding to the coreceptor CXCR-4. (C, D) Coreceptor binding, in turn, causes a chain of events which allows viral entry into the cell via a gp41 mediated process.

Gp120 Structure

The major protein on the outer surface of HIV-1 is the envelope glycoprotein gp120 (Figure 3). gp120 is a 120-kDa glycoprotein which is non-covalently bound to the viral transmembrane glycoprotein, gp41. There are 20-26 glycosylation sites on gp120 which account for over 50% of gp120's mass. Importantly, the inner domain surface and the receptor binding region are devoid of glycosylation (4). The protein consists of 492 amino acid residues and contains five hypervariable regions (V1-V5) and five constant regions (C1-C5) (6, 8, 9). The five variable regions form loops on the surface of the protein that contain disulfide bonds at their bases, whereas the conserved regions form a central core important for gp41 interaction and interaction with the viral receptors on the target cell (8, 9). Upon initial infection, the V3 region is the major antigenic site. However, due to its high variability (up to 50% difference in this loop alone between HIV isolates) the elicited neutralizing antibody is only specific to one particular strain. Furthermore, variation is manifested at several additional levels: within individuals, between isolates from individuals infected with HIV-1 within a localized geographical area, and between isolates identified in distinct areas of the world (11). Regardless of variability, HIV isolates can be grouped into a small number of classes, or clades, based upon their few conserved residues, and specifically upon a somewhat conserved region within the V3 loop called the crown sequence (11). The major clade found in the United States and Europe is the B clade, from which the crystal structure of gp120 was solved (9, 11). It is important to note that for the crystallization studies, loops V1-V3 were deleted

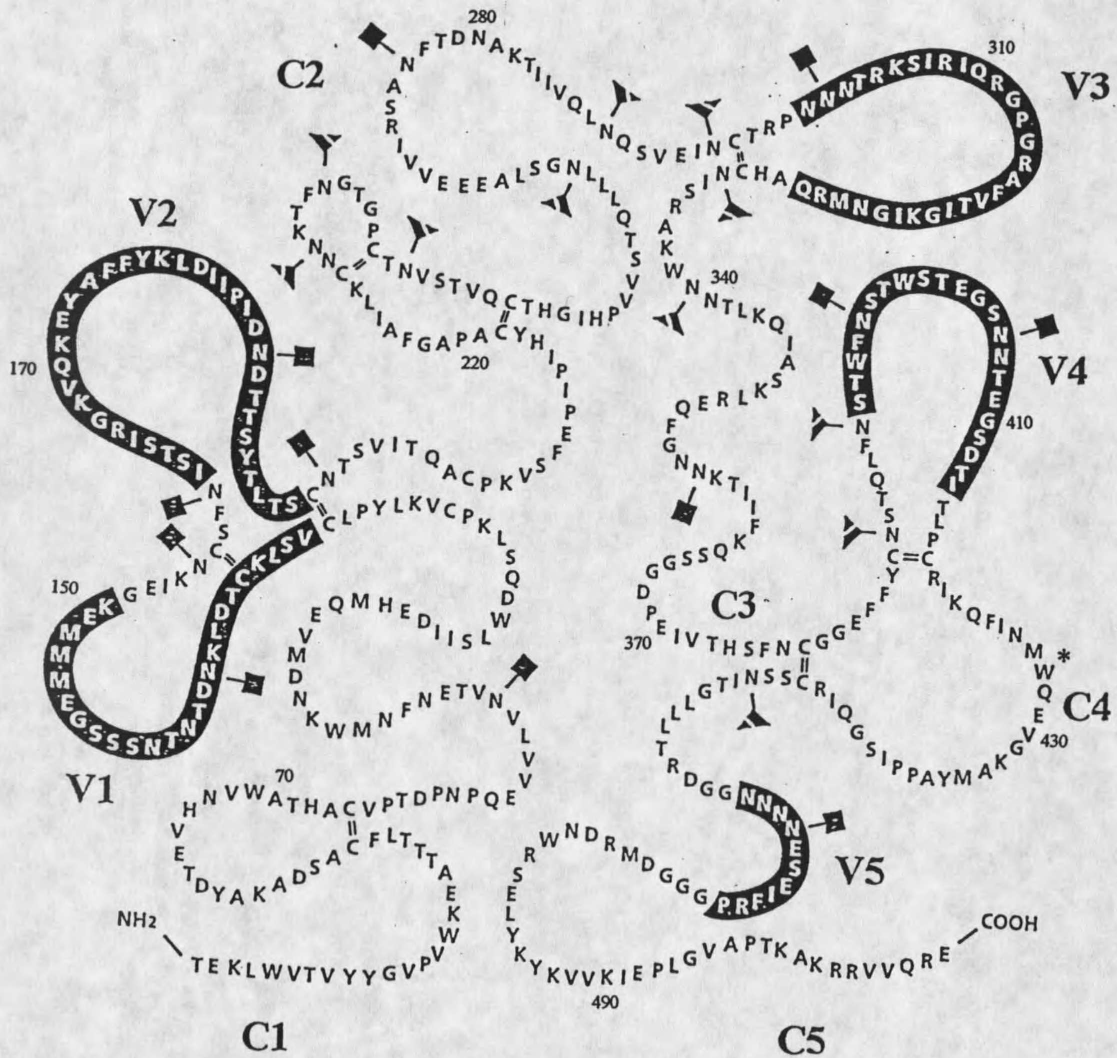


Figure 3. The primary amino acid sequence of gp120. The five hypervariable regions are shown in black (V1-V5) and the five constant regions of the gp120 core are shown in white (C1-C5). Glycosylation sites are shown by black squares (high manose/hybrid-type oligosaccharides) and broken triangles (complex-type oligosaccharides). Naturally occurring disulfide bonds are shown between cysteine residues by a double lines. (Note: this schematic does not include the leader sequence. For reference, Trp427 is indicated by an asterisk.

due to their conformational instability (8, 9); 90% of the attached carbohydrate was also removed (8, 9). Thus, the crystal structure, albeit a major breakthrough in the study of HIV, is not the *exact* structure of native gp120. However, the "core gp120" used for the x-ray structure was still able to bind CD4 with high affinity (8, 9).

Gp120 Interaction with CD4

HIV's first step in gaining access to T-cells is attachment through recognition of CD4 by gp120 (2). The residues on CD4 in contact with gp120 are concentrated in the span from 25 to 64 and are located on a protuberant ridge along one face of CD4's D1 domain. Two key amino acid residues, Phe43 and Arg59, make multiple contacts with residues Asp368, Glu370, and Trp427 in the C4 region of gp120, which are all conserved among primate immunodeficiency viruses (2, 4, 9).

The x-ray crystal structure at 2.5 Å resolution of a HIV-1 gp120 core complexed with a two-domain fragment of the human CD4 and an antigen-binding fragment of a neutralizing antibody revealed much about the CD4 binding surface. A relatively acidic gp120 surface was shown to complement a basic patch on CD4, and two major pockets were shown to be associated with the CD4 binding surface: a hydrophobic cavity (~150 $Å^3$) and a water-filled cavity (~280 $Å^3$) (8, 9). The overall binding surface is ~800 $Å^2$ and is a discontinuous epitope consisting of six segments of gp120: one residue from the V1/V2 stem, loop ζD, the β15-α3 excursion, the β20-β21 hairpin, strand β23, and the β24-α5 connection (Figure 4) (8, 9). Residues within the hydrophobic cavity, which extends roughly 10 Å into the interior of gp120 and is plugged by Phe43 of CD4, are