Molecular structure at sites of HIV-1 gp120-CD4 interaction
by Daphne Bryce Moffett

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biochemistry
Montana State University

Abstract:
The goal of this research project is to investigate the three-dimensional structure of a segment of CD4 when it is bound to HIV-1 gpl20. The peptide segment investigated is named Peptide 3 and corresponds to CD4 36-59. This peptide shows biological activity in ELISA assays as a low affinity inhibitor of CD4 binding to gpl20 with ~1 mMkd Conformations of the free peptide were studied by CD and NMR; gp120-bound structures of the peptide were studied by NMR and NMR restrained molecular dynamics methods.

The peptide showed very little secondary structure in solution according to CD analysis.

CD studies suggested the presence of 1% α-helix, 1% |&belta;-sheet, and 98% random coil in the free peptide, and NMR analysis showed evidence of an extended (3-sheet by conformation. The contribution of the free peptide NOESY cross peak intensities was subtracted from the observed NOESY cross peak intensities of the mixture of gp 120-bound and free peptide to obtain NOESY intensities of the gp120-bound peptide which were used as structural constraints in our molecular dynamics studies.

Following several rounds of starting coordinate randomization, simulated annealing, molecular dynamics, and using iterative refinement MARDIGRAS (a program which accurately estimates distances from 2D NOE spectra), we arrived at 2 families of structures for the bound structure. Both families had low total energy values and shared several structural characteristics with the corresponding region of the CD4 crystal structure. While Peptide 3 has provided certain new structural insights, we would like to further elucidate the CD4-gpl20 interactions by studying other peptides with faster koff rates and more biological activity.

Phage display techniques were used to search for a more biologically active peptide candidate. Various methods of phage selection were used and a 9-residue consensus sequence with homology to CD4 emerged. A conformational molecular model of the 9-residue phage peptide was built using the InsightII modeling software, and a comparison to the crystal structure of CD4 suggests that the phage peptide could be mimicking a discontinuous epitope on the surface of CD4.

MOLECULAR STRUCTURE AT SITES OF HIV-1 GP120-CD4 INTERACTION

by

Daphne Bryce Moffett

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

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

May 1997

APPROVAL

of a thesis submitted by

Daphne Bryce Moffett

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.

5/15/97
Date

Chairperson, Graduate Committee

Approved for the Major Department

5/15/97
Date

Head, Major Department

Approved for the College of Graduate Studies

6/3/97
Date

Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature 

Date May 12, 1997

ACKNOWLEDGEMENTS

Several people have been instrumental in the completion of this work. I would like to thank my advisor, Martin Teintze, for his leadership and guidance, as well as his unwaivering belief in my ability to see this project through to the end. I would also like to thank the rest of my committee: Ed Dratz, Jean Starkey, Arnold Craig, Sam Rogers, and Rocky Ross for insightful suggestions and for their support of my endeavors.

Because of the interdisciplinary nature of this project, I have had the pleasure of working with many people in the Chemistry and Biochemistry Department and the Microbiology Department. I appreciate the kind help, support, and friendship I have received from the Teintze, Craig, Dratz, Starkey, Dooley, Pincus, and Jesaitis groups. I am very grateful to Scott Busse and Dawit Gizachew for their friendship and for their close involvement in this project; it has been a real pleasure working and playing with them. Scott and Dawit were invaluable for their NMR and Molecular Dynamics expertise, respectively. I would also like to thank Tami Peters for sequencing the phage clones and Ed Dratz for running several high-field NMR experiments at NMRFAM.

I would like to thank all of the people who have provided moral support during my research years at Montana State University. I have many treasured friendships for which I am very grateful.

Finally, I would like to thank my family for continued encouragement, patience, and love. A very special thanks goes to my husband Mof and my daughter India who gave up their wife and "Mama" for 6 months, so I could finish what was started. Their understanding, support, and love have truly been my strength.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

ABSTRACT

The goal of this research project is to investigate the three-dimensional structure of a segment of CD4 when it is bound to HIV-1 gp120. The peptide segment investigated is named Peptide 3 and corresponds to CD4 36-59. This peptide shows biological activity in ELISA assays as a low affinity inhibitor of CD4 binding to gp120 with ~1 mM K_d. Conformations of the free peptide were studied by CD and NMR; gp120-bound structures of the peptide were studied by NMR and NMR restrained molecular dynamics methods.

The peptide showed very little secondary structure in solution according to CD analysis. CD studies suggested the presence of 1% α-helix, 1% β-sheet , and 98% random coil in the free peptide, and NMR analysis showed evidence of an extended β-sheet by conformation. The contribution of the free peptide NOESY cross peak intensities was subtracted from the observed NOESY cross peak intensities of the mixture of gp120-bound and free peptide to obtain NOESY intensities of the gp120-bound peptide which were used as structural constraints in our molecular dynamics studies.

Following several rounds of starting coordinate randomization, simulated annealing, molecular dynamics, and using iterative refinement MARDIGRAS (a program which accurately estimates distances from 2D NOE spectra), we arrived at 2 families of structures for the bound structure. Both families had low total energy values and shared several structural characteristics with the corresponding region of the CD4 crystal structure. While Peptide 3 has provided certain new structural insights, we would like to further elucidate the CD4-gp120 interactions by studying other peptides with faster k_{off} rates and more biological activity.

Phage display techniques were used to search for a more biologically active peptide candidate. Various methods of phage selection were used and a 9-residue consensus sequence with homology to CD4 emerged. A conformational molecular model of the 9-residue phage peptide was built using the InsightII modeling software, and a comparison to the crystal structure of CD4 suggests that the phage peptide could be mimicking a discontinuous epitope on the surface of CD4.

CHAPTER ONE
AN INTRODUCTION TO HIV AND AIDS

Discovery of AIDS

In the summer of 1981, five cases of *Pneumocystis carinii* pneumonia, a very rare condition at that time, were reported to the Centers for Disease Control (CDC). Soon thereafter, the center began receiving increasing reports of this form of pneumonia, as well as increasing incidences of a very unusual skin cancer, *Kaposi's sarcoma.* What caught the attention of the CDC was that these diseases had previously been limited to individuals with impaired cell-mediated immunity. In 1982, the resultant disease was named *acquired immune deficiency syndrome* (AIDS). In 1983, once the infectious pathway was determined to be blood-linked as well as being transmittable through body-fluid, it was determined that AIDS was the result of a previously unknown or unrecognized retrovirus, *human immunodeficiency virus* (HIV) (1, 2).

Due to the complexity and the high mutation rate of the virus, it has been difficult to find effective treatments. The exponential growth in the number of HIV affected individuals within the last decade, and the lethal nature of the infection, makes finding a cure critical. The understanding of how this virus works is presumably essential to the establishment of effective therapeutic regimens and to achieving long-term cures.

HIV Infection

HIV is a retrovirus. It is incapable of self-replication without the aid of host cellular machinery. Retroviruses are RNA-containing viruses that replicate through a DNA intermediate by virtue of a viral-coded RNA-dependent DNA polymerase, known as reverse transcriptase (3). Entry of HIV into the primary target cell, $CD4^+$ T-lymphocytes, involves an initial binding of virions to receptors on the target cells. This occurs through the binding of the HIV outer membrane glycoprotein gp120 with CD4. The interaction between these two proteins is one of very high affinity, K_d being approximately 10^{-9} M (4-6). A region known as the V3 hypervariable region of gp120, which spans residues 307 to 330, extends as a loop formed by two disulfide-linked residues at position 303 and 337 (1). The V3 loop of gp120, which is not required for binding to CD4, has been identified as a major determinant of HIV-1 tropism for different cells (7) and represents the main target for neutralizing antibodies (8). When the gp120 binds to CD4, it induces a conformational change that exposes a fusogenic domain on gp41. The fusogenic domain is then thought to mediate fusion of the viral envelope with the target-cell membrane (Figure 1). This study is focused on the mechanism of the initial interaction between HIV gp120 and CD4.

More recently accessory adhesion proteins called fusins have been identified on $CD4^+$ cells that appear to participate in HIV infection (9, 10). This protein is a putative G protein-coupled receptor with seven transmembrane segments. Fusin has been identified as a chemokine receptor which is now called CXCR4 and is a purported co-receptor for HIV strains that appear to dominate during later stages of infection. An analogous co-receptor

CCR5 appears to be most important during early stages of infection. The roles of these proteins are currently being clarified in a number of laboratories.

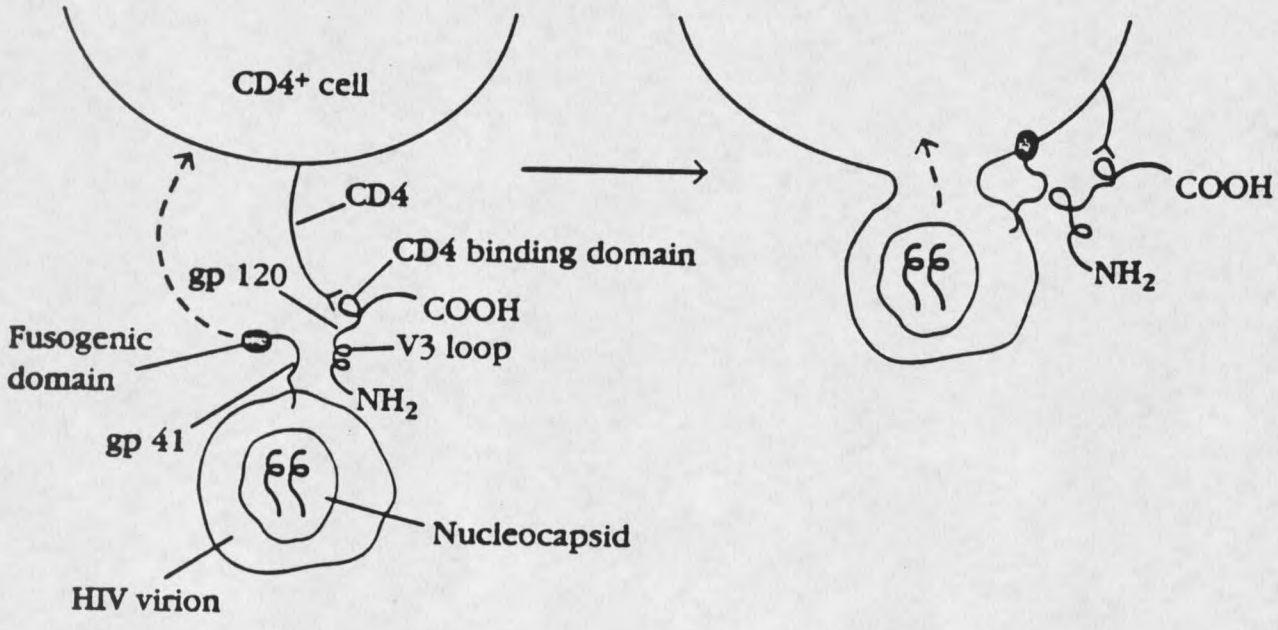

Figure 1. Proposed model of HIV fusion with a target $CD4^+$ cell. [From Kuby (1)]

HIV Gp120

HIV gp120 is a 120 kilodalton (kd) outer membrane glycoprotein (54 kd are attributed to polypeptide while the remaining mass results from oligosaccharide content) which is noncovalently linked to the virus' transmembrane glycoprotein, gp41. It is synthesized from a glycosylated precursor protein, gp160, in the rough endoplasmic reticulum of infected cells. The precursor is cleaved generating the small carboxyl-terminal fragment gp41 which spans the virion membrane and gp120 which is bound to gp41. There are 20-26 complex and high mannose glyscosylation sites on gp120 (11). Over 50% of its mass results from posttranslational modifications with carbohydrates (12), and this may be the reason it has not been possible to crystallize this protein. The gp120 protein consists of 509 amino acid residues and contains five hypervariable regions and 4 constant regions (1). Due to this variability as well as the virus' high mutation frequency, treatments lose their effectiveness over a relatively short period of time. Humans who are infected with the virus do produce antibodies to gp120; however, these antibodies are present in such low titers that they, too, are insufficient combatants of the disease. After an individual is infected with HIV, specific neutralizing antibodies are made to viral proteins and glycoproteins. These antibodies have been shown to block HIV's ability to infect T-cells *in vitro*. However, the virus' high mutation rate, coupled with the high rate of viral replication, enable some viral progeny to become resistant to the antibodies and to continue to infect additional cells and replicate. Eventually, a population of resistant viral particles emerges.

As was mentioned earlier, the V3 hypervariable region is the major antigenic site and plays a unique role in various strains of HIV. Since the V3 loop is the most antigenic and the most variable region of gp120 (it can differ by as much as 50% between HIV-1

isolates), the neutralizing antibody is strain specific (1). A disulfide bond formed between cysteine303 and cysteine337 creates a loop structure which contains a crown region generally extending from amino acid 316 or 317 to 324. Although the entire V3 loop exhibits extensive variation from one viral isolate to another, a significant percentage of isolates have a common crown sequence. Thus HIV isolates can be grouped into a small number of classes based on their gp120 crown sequence (1). An example of this is the MN crown sequence which 30% of HIV isolates in North America have in common (1).

The CD4 binding sequence on gp120 was investigated by researchers at Genentech, who transfected Chinese Hamster Ovary (CHO) cells with the CD4 gene (4). They found that soluble, cloned [^{125}I] labeled gp120 could bind to these transfected cells, but not to untransfected controls. They then cleaved the gp120 molecule into peptide fragments, produced monoclonal antibodies to each fragment, and tested the ability of each monoclonal to inhibit binding of the radiolabeled gp120 to the CD4-transfected CHO cells. By using this procedure, they identified a largely conserved region of amino acids (397-439) near the carboxy terminus of gp120 that appear to be involved in CD4 binding. Further evidence for the role of this sequence in CD4 binding was obtained by synthesizing a peptide with this sequence and showing that it would also block binding of soluble radiolabeled gp120 to the CD4-transfected CHO cells. When regions within the 397-439 sequence were deleted, a substantial reduction in binding to CD4 occurred (4).

The HIV Receptor CD4

The primary cellular target for HIV is the CD4$^+$ T-lymphocyte (13, 15). These CD4$^+$ T-lymphocytes, which comprise the major class of T-helper cells, are involved in cell-mediated immune responses. Although their role in the immune system is not completely

understood, several functions have been proposed. CD4 is an adhesion molecule which binds to Class II Major Histocompatability Complex (MHC) proteins. $CD4^+$ cells secrete cytokines (or lymphokines) which play a central role in the activation of B-cells, cytotoxic T-cells, and other cells. More recently, CD4 has been suggested to play a role in signal transduction since the cytoplasmic domain of CD4 is associated with $p56^{lck}$, a *src*-related protein kinase (15, 16).

Structurally, CD4 is a 55 kd (relative molecular mass) membrane glycoprotein which consists of four immunoglobulin-like domains, a hydrophobic transmembrane region, and a cytoplasmic domain. Two crystal structures of immunoglobulin-like domains 1 and 2 have been refined at 2.3 Å resolution (5, 17) and a structure in a second crystal latttice has also been refined at 2.9 Å resolution (18). Initial identification of the residues in human CD4 critical for the binding of HIV relegated them to the first immunoglobulin-like domain (17, 18). Subsequent mutation/substitution experiments, as well as peptide competition experiments, further localized the binding site to the CD4 29-59 region (17, 18, 21, 22). The residues implicated in direct interaction with gp120 are Lys29, Lys35, Phe43, Leu44, Lys46, Gly47, and Arg59 (18). These putative residue interactions are based on mutational studies of residues in the crystal structures which were highly motivated by solvent accessibility calculations (5, 18, 22). These fractional solvent accessibility calculations determine the degree of exposure for each measured residue as a ratio of its accessibility within the protein to that obtained after reducing the structure to a simple tripeptide consisting of the specific residue flanked by two glycines. The side chain of Phe43 is highly solvent exposed in the crystal structure suggesting the possibility of direct interaction in binding. Replacement of Phe43 with alanine or isoleucine reduces affinity for

gp120 by more than 500-fold (22). Evidence for the roles of other CD4 residues in binding to gp120 is presented in the next chapter.

Although the binding sites for CD4 and gp120 have been relegated to specific residue stretches, there seems to be more ambiguity concerning the gp120 binding site than the CD4 binding site. According to the experimental findings of Jameson et al. (21), peptide analogs of the CD4 region 37-53 should be useful in more precise determinations of the fine structure of the HIV-binding site with the potential of leading to the design of small molecule inhibitors of the AIDS virus-receptor interaction (23). The hope is that by studying the interactions of the critical CD4 regions necessary to bind gp120, the structure of the bound gp120 can be elucidated suggesting conformations for blocking agents. Since the publication of Jameson's results in 1988, several groups have expanded the CD4 region to include residues 29-59 (18). The crystal structure indicates that several residues which lie within this region are highly solvent exposed and could be available for interaction with a docking molecule, i.e. gp120.

Statement of Problem

Due to the complexity of HIV, several approaches are being taken to seek directed therapies. An obvious choice, but not necessarily the simplest, is the blocking of the initial infection step so that the gp120 cannot bind to its receptor CD4. Before the discovery of fusins when the present work began, there appeared to be two possible ways to execute this: either the outer membrane glycoprotein gp120 could be blocked, or the cellular CD4 receptor could be blocked. Because the CD4+ T-lymphocyte is essential to the proper functioning of the immune system, and because the MHC II- and gp120-binding sites on CD4 are distinct but overlapping (24), blocking the CD4 receptor would probably not be a

prudent choice. As well, gp120 induces apoptosis in CD4+ cells causing premature cell death and resulting in a compromised immune system (25-27). We have chosen to study CD4 peptides that block the gp120 binding sites for CD4. Thus, we have designed peptide analogs of CD4 to help elucidate the structural basis of the CD4-gp120 interaction by obtaining information on the gp120-bound structure of the interacting segments of the CD4 receptor. It has been hypothesized that CD4 or gp120 or both may undergo conformational changes upon binding (28). If that is the case, the CD4 crystal structure would not show the gp120-bound conformation. In addition, the crystal structures of CD4 domains 1 and 2 in two different crystal lattices (18) show the largest conformational differences on surface loops including the putative gp120 interaction region. Thus, interactions between CD4 proteins in the crystal lattice have a significant effect on the conformation of the surface loops implicated in the gp120 interaction. Given the apparent flexibility of the surface loops, it is anticipated that they may change conformation significantly when they dock with their binding sites on gp120. The goal of the present study is to map structural changes involved in the binding of these two proteins and to suggest higher affinity peptide analogs for further analyses. Studies presented in this thesis focus on a particular CD4 peptide, Peptide 3. Peptide 3 spans CD4 36-59 and was chosen based upon crystal structure analysis and site directed mutagenesis studies. Recent phage display experiments have suggested new peptide candidates for the interaction region with gp120, and those peptides will also be discussed. This thesis has laid the groundwork for many subsequent experiments, some of which will be suggested in the conclusion section.

References Cited

1. Kuby,J. (1992) *Immunology*, Freeman and Company, New York.

2. Watson,J.D., Gilman,M., Witkowski,J. and Zoller,M. (1992) *Recombinant DNA 2nd edition*, Freeman and Company, New York.

3. Vaishnav,Y.N. and Wong-Staal,F. (1991) The Biochemistry of AIDS, *Annu. Rev. Biochem.* **60**, 577-630.

4. Lasky,L.A., Nakamura,G., Smith,D.H., Fennie,C., Shimasake,C., Patzer,E., Berman,P., Gregory,T., and Capon,D.J. (1987) Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor, *Cell* **50**, 975-985.

5. Ryu,S.E., Kwong,P.D., Truneh,A., Porter,T.G., Arthos,J., Rosenberg,M., Dai,X., Xuong,N.H., Axel,R., Sweet,R.W. and Hendrickson,W.A. (1990) Crystal structure of an HIV-binding recombinant fragment of human CD4, *Nature* **348**, 419-426.

6. Moore,J.P., McKeating,J.A., Weiss,R.A. and Sattentau,Q.J. (1990) Dissociation of gp120 from HIV-1 virions induced by soluble CD4, *Science* **250**, 1139-1142.

7. Hwang,S.S., Boyle,T.J., Lyerly,H.D. and Cullen,B.R. (1991) Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1, *Science* **253**, 71-74.

8. Tilley,S.A. and Pinter,A. (1993) Human and chimpanzee monoclonal anitbodies with antiviral activity against HIV-1, *AIDS Res. Rev.* **3**, 255-287.

9. Feng,Y., Broder,C.C., Kennedy,P.E. and Berger,E.A. (1996) HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane G protein-coupled receptor, *Science* **272**, 872-877.

10. Choe,H., Farzan,M., Sun,Y., Sullivan,N., Rollins,B., Ponath,P.D., Wu,L., Mackay,C.R., LaRosa,G., Newman,W., Gerard,N., Gerard,C. And Sodroski,J. (1996) The ß-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates, *Cell* **85**, 1135-1148.

11. Yeh,J., Seals,J.R., Murphy,C.I., Van Halbeek,H. and Cummings,R.D. (1993) Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a baculovirus expression system, *Biochemistry* **32**, 11087-11099.

12. Allan,J.S., Coligan,J.E., Barin,F., McLane,M.F., Sodrowski,J.G., Rosen,C.A., Haseltine,W.A., Lee,T.H. and Essex,M. (1985) Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III, *Science* **228**, 1091-1094.

13. McDougal,J.S., Kennedy,M.S., Sligh,J.M., Cort,S.P., Mawle,A. and Nicholson,J.K. (1986) Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule, *Science* **231**, 382-385.

14. Sattentau,Q.J. and Weiss,R.A. (1988) The CD4 antigen: Physiological ligand and HIV receptor, *Cell* **52**, 631-633.

15. Tremblay,M., Meloche,S., Gratton,S., Wainberg,M.A. and Sekaly,R. (1994) Association of p56lck with the cytoplasmic domain of CD4 modulates HIV-1 expression. *EMBO J.* **13**, 774-783.

16. Goldman,F., Jensen,W.A., Johnson,G.L., Heasley,L. and Cambier,J.C. (1994) Gp120 ligation of CD4 induces p56lck activation and TCR desensitization independent of TCR tyrosine phosphorylation. *J. Immunol.* **153**, 2905-2917.

17. Wang,J., Yan,Y., Garrett,T.P.J., Liu,J., Rodgers,D.W., Garlick,R.L., Tarr,G.E., Husain,Y., Reinherz,E.L. and Harrison,S.C. (1990) Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains, *Nature* **348**, 411-418.

18. Ryu,S.- E., Truneh,A., Sweet,R.W. and Hendrickson,W.A. (1994) Structures of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices, *Structure* **2**, 59-73.

19. Arthos,J., Deen,K.C., Chaikin,M.A., Fornwald,J.A., Sathe,H., Sattentau,Q.J., Clapham,P.R., Weiss,R.A., McDougal,J.S., Pietropaolo,C., Axel,R., Truneh,A., Maddon,P.J. and Sweet,R.W. (1989) Identification of the residues in human CD4 critical for the binding of HIV, *Cell* **57**, 469-481.

20. Landau,N.R., Warton,M. and Littman,D.R. (1988) The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4, *Nature* **334**, 159-162.

21. Jameson,B.A., Rao,P.E., Kong,L.I., Hahn,B.H., Shaw,G.M., Hood,L.E. and Kent,S.B.H. (1988) Location and chemical synthesis of a binding site for HIV-1 on the CD4 protein, *Science* **240**, 1335-1339.

22. Moebius,U., Clayton,L.K., Abraham,S., Harrison,S.C. and Reinherz,E.L. (1992) The human immunodeficiency virus gp120 binding site on CD4: Delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure. *J. Exp. Med.* **176**, 507-517.

23. Sattentau,Q.J. and Moore,J.P. (1993) The role of CD4 in HIV binding and entry, *Philos. Trans. R. Soc. Lond. [Biol.]* **342**, 59-66.

24. Bour,S., Geleziunas,R. and Wainberg,M.A. (1995) The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiol. Rev.* **59**, 63-93.

25. Ameisen,J.C. and Capron,A. (1991) Cell dysfunction and depletion in AIDS: the programmed cell death hypothesis, *Immunology Today* **12**, 102-105.

26. Terai,C., Kornbluth,R.S., Pauza,C.D., Richman,D.D. and Carson,D.A. (1991) Apoptosis as a mechanism of cell death in cultured T lymphoblasts acutely infected with HIV-1, *J. Clin. Invest.* **87**, 1710-1715.

27. Laurent-Crawford,A.G., Krust,B., Muller,S., Riviere,Y., Rey-Cuille,M.A., Bechet, J.M., Montagnier, L. and Hovanessian,A.G. (1991) The cytopathic effect of HIV is associated with apoptosis, *Virology* **185**, 829-839.

28. Reed,J. and Kinzel,V. (1991) A conformational switch is associated with receptor affinity in peptides derived from the CD4-binding of gp120 from HIV I, *Biochemistry* **30**, 4521-4528.

CHAPTER TWO

SELECTION OF CD4 PEPTIDES FOR STUDY BASED ON MUTATIONAL ANALYSES AND X-RAY CRYSTAL STRUCTURE DATA

Introduction

Two routes for selecting CD4 peptides that bind to the CD4-binding site on gp120 and are suitable for NMR studies will be discussed in this thesis. The first route, which will be the emphasis of this chapter, involved the examination of mutagenesis and the crystal structure data for CD4 in the literature. The second approach, which will be discussed in a later chapter, involved screening a random peptide phage display library. The sequence for Peptide 3 which became the subject of our NMR and molecular dynamics studies was determined using the first approach. Initially, an extensive literature search was conducted in order to determine what information was already available that might aid in the design of our peptide candidates.

In 1988, Peterson and Seed (1) used saturation mutagenesis and complement fixation to obtain CD4 mutants with impaired antibody and human immunodeficiency virus binding. Mutational analysis of antibody epitopes was used to identify key contact residues important for CD4 recognition. Some of the mutant CD4 proteins selected for loss of activity with antibodies known to strongly block HIV binding also displayed impaired virus adsorption and/or syncytium formation. They showed that COS cells transfected with cDNAs for CD4 were unable to bind HIV if they had any of the following mutations: deletion of residues 42-49 or the substitutions Gly47Arg, Thr45Pro and Lys46Asn/Gly47Val.

Similar conclusions emerged from other genetic studies conducted by Clayton et al. (2) and Mizukami et al. (3). The Mizukami group made a series of insertional mutations of Ser-Arg between residues 31 and 57 which impaired binding of gp120, indicating a probable direct role of this region of the first CD4 domain. Specifically, mutants 31, 44, 48, 52, 55, and 57, all of which contained insertions within the disulfide loop of the first domain, interacted weakly or not at all with gp120. The relative extent of impairment varied among the different mutants in this region. Gp120 binding to mutants 31, 44, 48, 55, and 57 was drastically reduced, whereas binding to mutant 52 was only moderately impaired.

The Clayton group substituted all non-conserved murine for human CD4 residues between amino acid positions 27-167 (2). This extracellular segment of murine CD4 has an overall 50% identity with its human counterpart (4) at the amino acid level, but fails to bind gp120 (5). Oligonucleotide-directed mutagenesis was used to create each of 16 individual mutant human CD4 molecules containing from 1-4 amino acid substitutions. Substitution mutations Pro48Gly, Lys50Pro, and Leu51Ser clearly abrogated the ability of CD4 to bind to gp120. There was evidence, as well, that some amino acids located in D2 (the second domain of CD4) were also necessary for gp120 binding. Modelling studies of their data seemed to suggest that the binding site was localized to the C"ß-strand within CDR2 (the complementary determining region 2 spanning amino acids 42-49), a part of Domain 1 (Figure 2).

Similar experiments were carried out by Landau et al. (7). They prepared chimeric CD4 cDNAs in which the sequence encoding the first two domains of the protein was made up of (1) murine sequence followed by human sequence, (2) human sequence followed by murine sequence, and (3) human sequence followed by murine sequence and then by more human sequence. CHO (Chinese hamster ovary) cells expressing the chimeric CD4 proteins were tested for their ability to bind gp120. Amino acids specific to human CD4 in

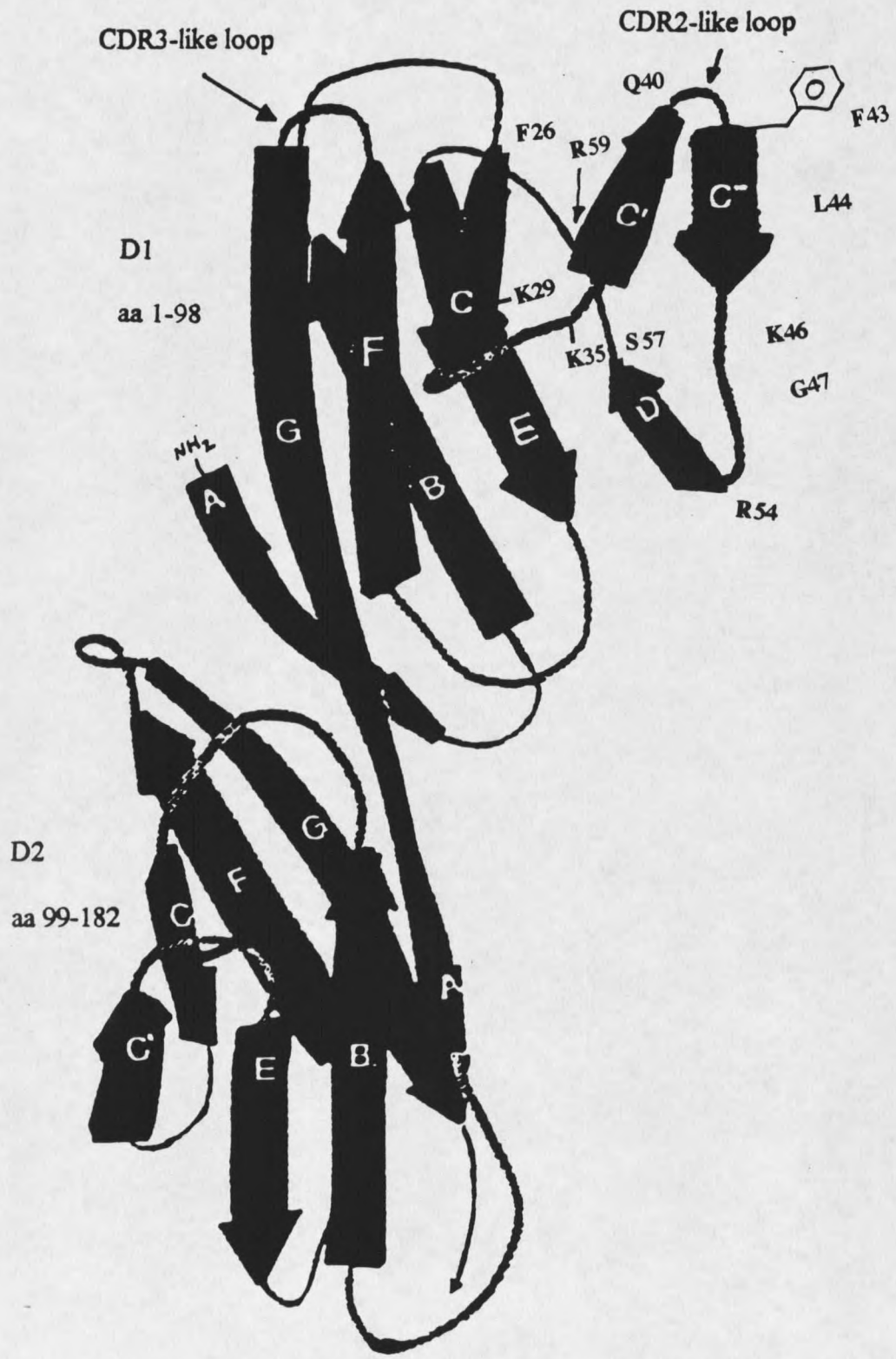

Figure 2. Backbone representation of CD4 (1-182). [From Wang et al. (6)]

the region of amino acids 1-37 were not necessary for gp120 binding, whereas amino acids 38-57 were indispensable. According to the human-mouse and human-mouse-human chimera data, the minimum region of human CD4 necessary for binding of gp120 lies in the region between amino acid 37 and 83.

In another mutational analysis study, Arthos et al. (8) expressed several truncated derivatives of sCD4 (which consists of the entire extracellular domain) and quantitated their affinities for gp120, as well as their ability to inhibit HIV infection in vitro. They then analyzed a series of 26 amino acid substitutions within Domain 1 (CD4 1-106). Their quantitative binding analyses identified a single binding site within this first domain. Substitutions at positions 43 and 55, as well as insertion of a single amino acid between residues 52 and 53, severely disrupted binding. Moderate inhibition of binding resulted from an insertion between residues 41 and 42 and from a substitution at positions 51 and 52. Arthos et al. (8) found that mutations outside this region of Domain 1 had no effect on binding. From these results, they concluded that the determinants for high affinity binding to gp120 resided within a site encompassing residues 41 to 55.

The results from the Arthos experiments were not consistent with conclusions of some studies using synthetic peptides. In one study, Lifson et al. (9) observed inhibition of virus infection with a peptide spanning residues 81-92. However, the peptides exhibiting inhibitory activity were found to be multiply benzylated due to incomplete deprotection following Fmoc synthesis. Particularly cysteine86 and glutamate87 required benzylation for the peptide to exhibit antisyncytial activity. In a synthetic peptide based study, inhibition of virus-mediated cell fusion was observed at high concentrations of a peptide encompassing residues 16-49, which includes the CDR2 homology region (10). Jameson et al. (10) identified a candidate structure for the HIV-1 binding site on the CD4 protein by epitope mapping with a family of eight functionally distinct CD4-specific monoclonal antibodies in conjuction with a panel of large CD4-derived synthetic peptides. At

concentrations down to 200 μg/ml, CD4-derived synthetic peptide 16-49 specifically inhibited syncytium (multinucleate fused cells between infected and uninfected CD4+ cells) formation in a concentration-dependent manner. It was proposed that a loop extending from residues 28-44 of the CD4 protein was a binding site for HIV-1 and that peptide analogs of this region would potentially be useful in designing small molecule inhibitors.

Moebius et al. (11) made quantitative measurements of the contributions made by individual amino acid side chains to gp120 binding. Informed by the atomic structure of CD4 Domains 1 and 2, they substituted exposed and buried residues in Domain 1 of CD4 and then analyzed the mutants to verify that only local perturbations of structure had occured. Twenty-six individual CD4 invariants, substituted at 20 distinct amino acid positions were generated by site-directed mutagenesis in order to define the key determinants of HIV gp120 binding. The mutants spanned a section of the first domain of CD4 from residues 19-89. To exclude the possibility that diminished HIV gp120 binding capacity was a trivial consequence of overall disruption of conformation in the first CD4 domain, each mutant was analyzed with a panel of anti-CD4 monoclonal antibodies directed to the native Domain 1. Binding of CD4 monoclonal antibodies to cells expressing CD4 mutants was determined by immunofluorescence and compared with binding of mAb OKT4, which binds to a membrane proximal domain of CD4. None of the mutants led to loss of the majority of antibody epitopes. The binding of monoclonal antibodies known to interact with the C'C" ridge (as shown in Figure 1) was reduced or eliminated by mutations in this region, but the binding of other monoclonals interacting with different parts of Domain 1 was unaffected. The restricted effect of amino acid substitutions on monoclonal antibody reactivity demonstrated that these mutations exert only local effects on the structure of CD4 and suggest that the overall conformation of Domain 1 is approximately preserved in the mutants.

The affinity of wt-CD4 and CD4 mutants for HIV_{IIIB} gp120 was determined through equilibrium binding experiments and kinetic experiments utilizing immunofluorescence. According to the authors, substitutions at positions 35, 43, and 62 had the most pronounced effects on the affinity of CD4 for gp120. In particular, it seems that Phe43 plays a critical role in HIV gp120 binding. In order to determine if a hydrophobic residue was the only requirement at this position, four other hydrophobic residues were substituted. Substitution of Ala or Ile at position 43 resulted in a ~500-fold loss in gp120 affinity; Trp and Tyr substitutions gave 15-30-fold decreases; and Leu yielded only a two-fold reduction. Thus, the chemical structure of the hydrophobic side chain at position 43 dramatically influences gp120 binding. Trp62, which is a buried residue, was mutated to tyrosine. Most of the tryptophanyl side chain is buried beneath the C'C" turn, and its bulk is probably critical for the way the turn projects from the edge of the CD4 molecule (11). The modest reduction in side chain volume at Trp62Tyr and the repositioning of a hydrogen-bonding group (the phenolic hydroxyl of tyrosine lies ~2 Å farther from the Cß than does the indole nitrogen of tryptophan) lead to a 150-fold loss of affinity for gp120 (11). The reduced affinities of the CD4 mutants for gp120 were found to be primarily due to an increase in the k_{off} rate. The authors concluded that surface topography in the neighborhood of Phe43, including the entire C'C" turn from residue 38 to 45, can influence the gp120 interaction. Also it seems that Trp62 is critical in its bulky character and may be responsible for the way the peptide turn located in this region projects from the edge of the CD4 molecule.

A study conducted by Ryu et al. (12) describes the structures of Domains 1 and 2 of human CD4 in the context of two different crystal lattices. The human CD4 D1D2 (Domains 1 and 2) protein that gave the type I crystals was produced as a secreted protein in CHO cells (13) and its structure was determined by X-ray crystallography to a resolution

of 2.3 Å. The D1D2 sample that gave the type II crystals was produced in *Streptomyces lividans* and was resolved at 2.9 Å (12). Residues in the CDR2-like loop that contains the crucial Phe43 are well-defined in both crystal structures, and show significant differences in local conformation in this region in the two lattices. The C'C" hairpin loops in both structures form similar slightly distorted type II'ß-turns. The phenyl group of Phe43 adopts a well-defined structural orientation despite its unusual solvent exposure for a hydrophobic group. The comparison of this putative binding region in the two crystal packing environments reveals significant but limited flexibility in conformation. The conclusions drawn from Ryu et al.'s (12) work suggest that the overall fold of CD4 Domain 1 may not change much upon binding to gp120, but they do not address the possible changes which occur in the Phe43 region . The C'C" loop showed the largest difference between the two structures. It appears from this that crystal packing forces can affect the conformation of surface loops on proteins, such as the C'C" loop in CD4 D1; however, it is not clear if the two lattices sampled expressed the full range of flexibility possible in this loop.

A final study is pertinent to this discussion - the interaction of a CD4 ß-turn mimetic with HIV gp120 (14). Chen et al. designed and synthesized a conformationally restricted peptidomimetic which incorporated CD4 residues 40-50. The small molecule mimetic inhibited the binding of human T-lymphotropic virus (HIV) type IIIB gp120 to CD4+ cells at low micromolar levels and reduced syncytium formation 50% at 250 μg/ml. This paper is significant because it shows the utility of a conserved CD4-like secondary structure in the design of an effective peptide inhibitor. It also supports our research direction in seeking to determine the detailed conformation which inhibiting peptides assume upon binding to gp120. Once the conformational study has progressed further, the synthesis of peptide analogs in locked, bound conformations can be completed. It is our hypothesis that such peptides will show higher biological activity than the linear parents. There is considerable evidence to support our hypothesis: Streptavidin mimetic peptides are good examples. It

has been shown that streptavidin mimetic peptides have lower affinities relative to biotin largely due to unfavorable changes of binding of linear peptides (15), while cyclic S-S analogs show 1000x higher affinity (16).

The papers which have thus far been discussed represent various approaches to the determination of CD4 residues that are critical for HIV gp120 binding. Not all of the authors are in agreement; however, there are some strong consistencies. In contemplating the design of peptides for NMR studies, we also had to consider other factors. We needed to choose peptides of a modest length (<30 residues), preferably containing a minimal number of repetitive residues for NMR analysis of the free peptide. Assignment of sidechain protons of residues which are repetitive may be difficult in linear peptides in solution and may be impossible without site-specific ^{13}C or ^{15}N labelling. The peptide also needed to have a k_{off} faster than the fastest cross-relaxation time of the ligand protons in the bound state. In general, that means that the K_d of the peptides should be between 10 μM and 1 mM. Taking all of this into consideration, we chose three peptides with which to start. The peptides and their sequences are compared in Table 1. They have different lengths; however, they all contain the C'C" ridge of CD4 which has been considered the most important region for gp120 binding in the literature (1-3, 6-8, 10-14).

Table 1. Comparison of CD4 Peptide Sequences

Peptide 1 CD4 (38-52)	GNQGSFLTKGPSKLNamide
Peptide 2 CD4 (33-48)	Ac-QIKILGNQGSFLTKGPamide
Peptide 3 CD4 (36-59)	Ac-ILGNQGSFLTKGPSKLNDRADSRRamide

Peptide 1 was selected based primarily on the data of Jameson et al. (10). However, its sequence was based directly on the residue numbers of the sequence published in the paper, which was later determined to be offset from the correct sequence 9 residues in the carboxyl-terminus direction. Peptide 2 was designed to incorporate more of the loop structure interdigitated between the CC' antiparallel ß-sheets. Peptide 3 was based, to a large extent, on the findings of Moebius et al. (11) and Ryu et al. (12). This peptide is considerably longer and contains many more of the key residues involved in binding, as determined by the previously discussed mutational analyses.

Materials and Methods

Peptides were synthesized using a Milligen Model 9050 continuous flow automated peptide synthesizer with Fmoc chemistry. PepSyn-KB resin (MilliGen/Biosearch) was used to initiate the synthesis and the completed peptides were cleaved from the resin using TFA. The crude peptides were isolated by ether precipitation. They were then purified by reverse phase HPLC and analyzed with MALDI-TOF and electrospray mass spectrometry to determine the purity. The amino acid sequence was confirmed by the NMR experiments (see Chapter 5).

Results and Discussion

Although three peptides were synthesized, it will later be shown that only one was an appropriate candidate for the NMR studies; therefore, I will focus my discussion on the results of Peptide 3. Mass spectral analysis, subsequent to the reverse phase HPLC purification, suggested that the peptide was >90% pure, as was also consistent with the presence of a single HPLC peak. The MALDI-TOF spectrum (Figure 3) showed a single

peak at 2666 mass units with a slight shoulder. Bradykinin and insulin ß-chain were used as internal standards and α-cyano-4-hydroxy-cynnamic acid (Sigma, St. Louis) was the matrix. This apparent molecular weight is approximately 0.2% less than the theoretical mass 2672.1 $gmole^{-1}$, but is well within acceptable limits of error for this technique. The electrospray mass spectrum (Figure 4) showed predominantly quadruply and triply charged species for Peptide 3 and suggested an apparent molecular weight 0.1% to 0.2% greater than the theoretical mass.

All three peptides were tested by ELISA (enzyme linked immunosorbent assay) to determine their abilities to inhibit HIV-1 gp120-CD4 interactions and this is discussed in the following chapter.