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.

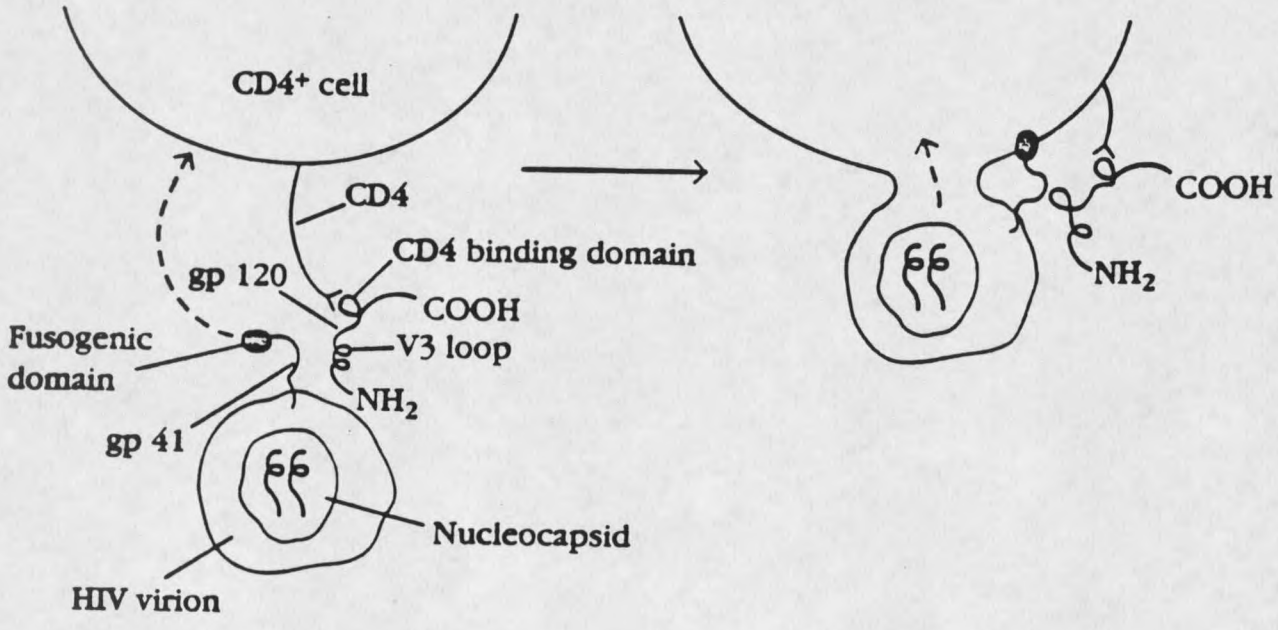


Figure 1. Proposed model of HIV fusion with a target $CD4^+$ cell. [From Kuby (1)]

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

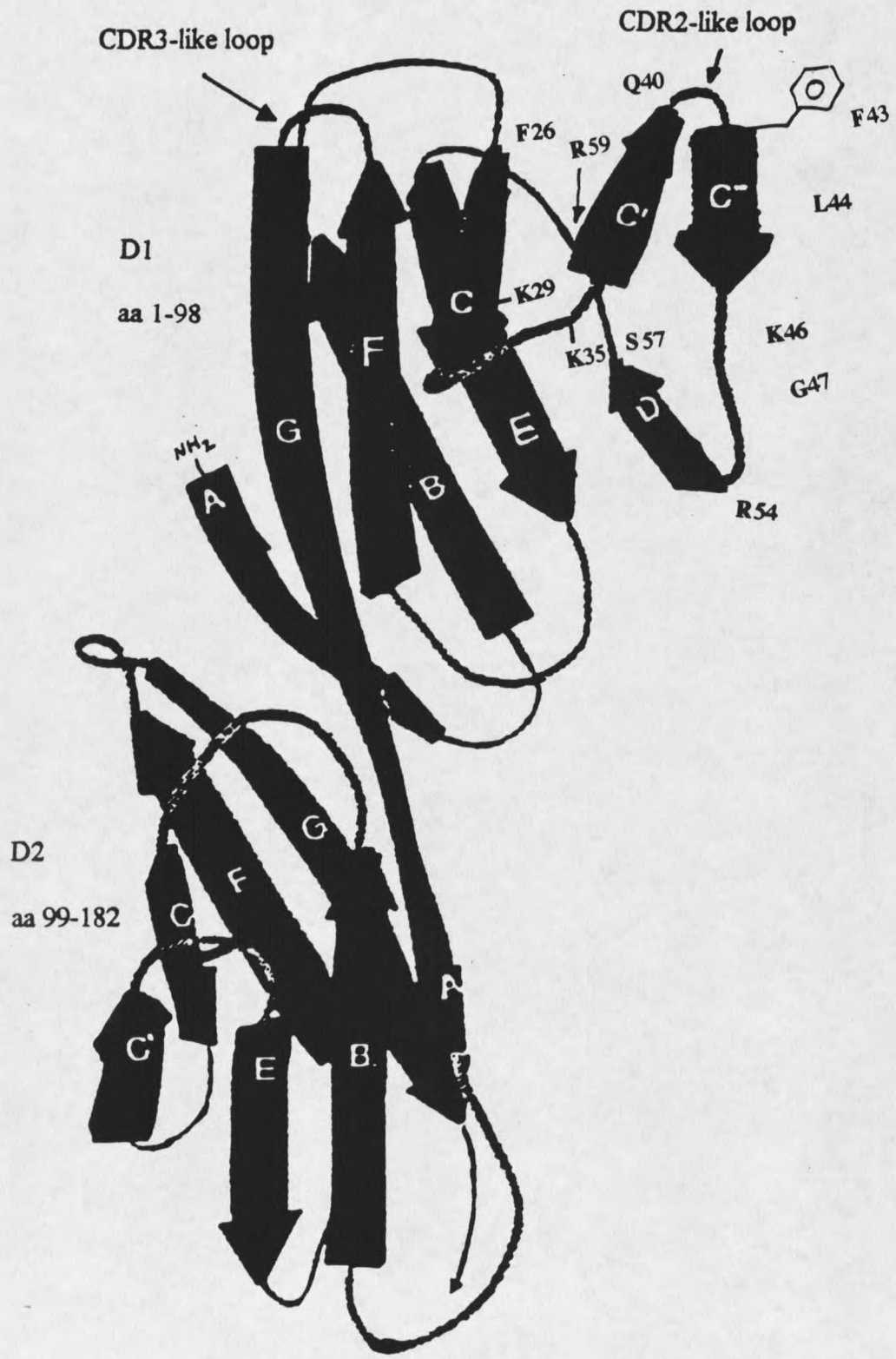


Figure 2. Backbone representation of CD4 (1-182). [From Wang et al. (6)]

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.

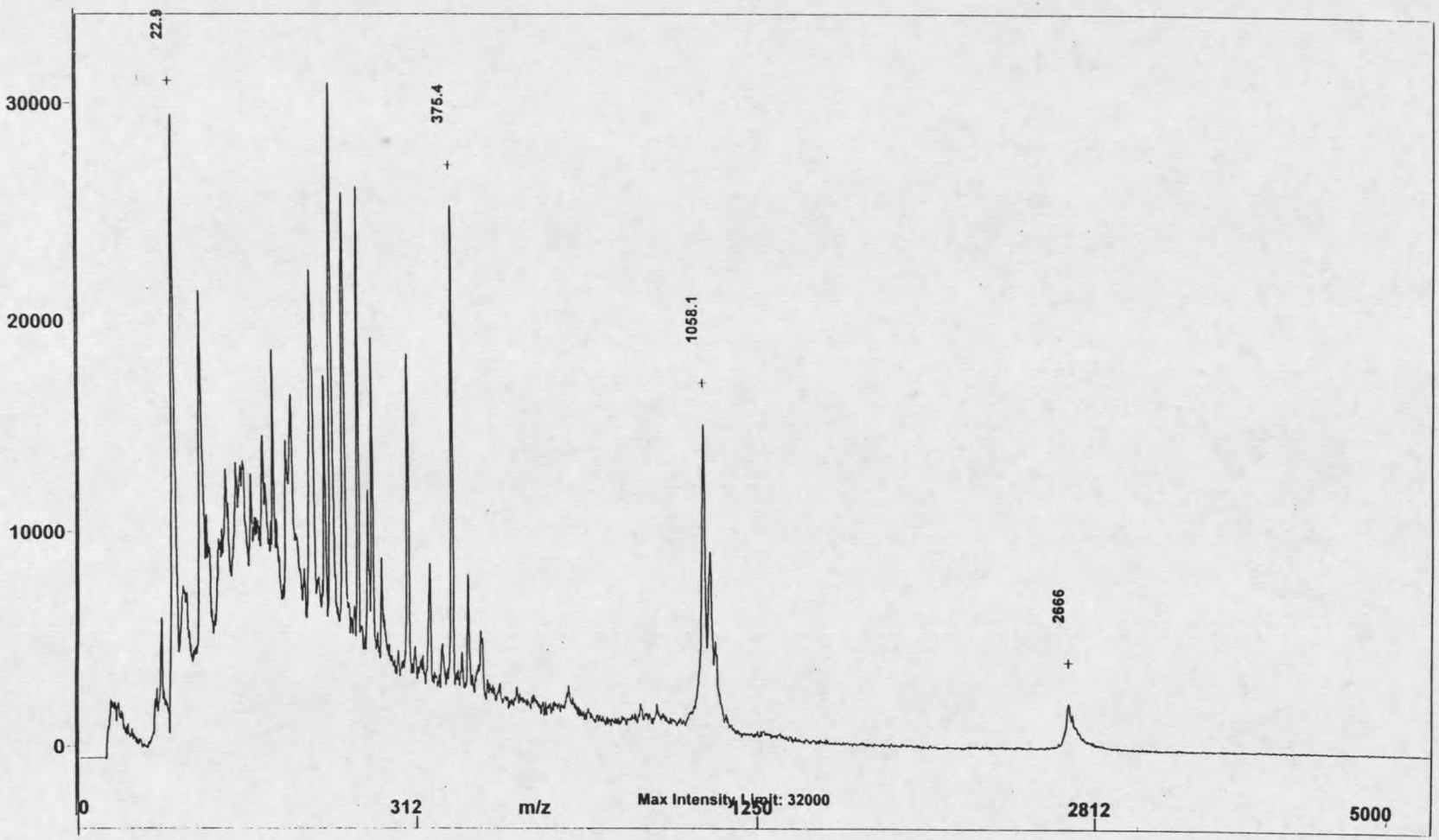


Figure 3. MALDI-TOF of Peptide 3. Bradykinin was used as an internal standard and is seen at 1058.1 and Peptide 3 shows a mass of 2666.

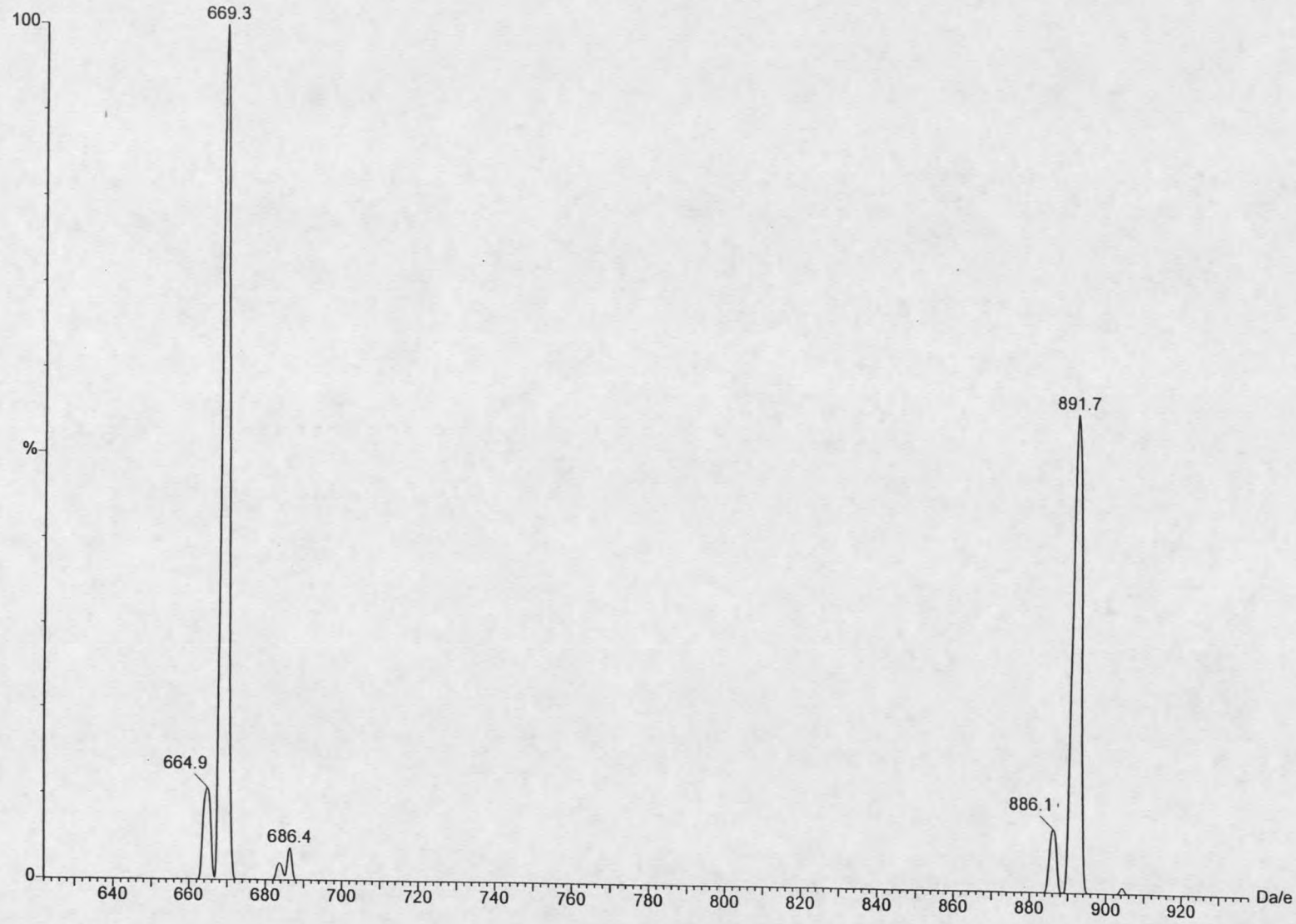


Figure 4. ES-MS of Peptide 3. The peaks shown correspond to PH_4^{+4} and PH_5^{+5}.

References Cited

1. Peterson,A. and Seed,B. (1988) Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4, *Cell* **54**, 66-72.

2. Clayton,L.K., Hussey,R.E., Steinbrich,R., Ramachandran,H., Husain,Y. and Reinherz,E.L. (1988) Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding, *Nature* **335**, 363-366.

3. Mizukami,T., Fuerst,T.R., Berger,E.A. and Moss,B. (1988) Binding region for human immunodeficiency virus (HIV) and epitopes for HIV-blocking monoclonal antibodies of the CD4 molecule defined by site-directed mutagenesis, *Proc. Natl. Acad. Sci. USA* **85**, 9273-9277.

4. Maddon,P.J., Molineaux,S.M., Maddon,D.E., Zimmerman,K.A., Godfrey,M., Alt,F.W., Chess,L. and Axel,R. (1987) Structure and expression of the human and mouse T4 genes, *Proc. Natl .Acad. Sci .USA* **84**, 9155-9159.

5. McClure,M.O., Sattentau,Q.J., Beverley,P.C., Hearn,J.P., Fitzgerald,A.K., Zuckerman,A.J. and Weiss,R.A. (1987) HIV infection of primate lymphocytes and conservation of the CD4 receptor, *Nature* **330**, 487-489.

6. 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.

7. 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.

8. 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.

9. Lifson,J.D., Hwang,K.M., Nara,P.L., Fraser,B., Padgett,M., Dunlop,N.M. and Eiden,L.E. (1988) Synthetic CD4 peptide derivatives that inhibit HIV infection and cytopathicity, *Science* **241**, 712-716.

10. 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.

11. 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.

12. 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.

13. 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.

14. Chen,S., Chrusciel,R.A., Nakanishi,H., Raktabutr,A., Johnson,M.E., Sato,A., Weiner,D., Hoxie,J., Saragovi,H.U., Greene,M.I. and Kahn,M. (1992) Design and synthesis of a CD4 beta-turn mimetic that inhibits human immunodeficiency virus envelope glycoprotein gp120 binding and infection of human lymphocytes, *Proc. Natl. Acad. Sci. USA* **89**, 5872-5876.

15. Weber,P.C., Pantoliano,M.W. and Thompson,L.D. (1992) Crystal structure and ligand-binding studies of a screened peptide complexed with streptavidin, *Biochemistry* **31**, 9350-9354.

16. Katz, B.A. (1995) Binding to protein targets of peptidic leads discovered by phage display: Crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence, *Biochemistry* **34**, 15421-15429.

CHAPTER THREE

ELISA AND FLUORESCENCE STUDIES OF PEPTIDE 3 WITH HIV-1 GP120

Introduction

Following the selection of the CD4 peptide candidates, ELISA (enzyme linked immunosorbant assay) and gp120 tryptophan fluorescence quenching studies were conducted to determine the gp120 binding affinities of the peptides. In particular, ELISA has been a popular technique for mapping antibody interactions with gp120 and CD4 (1-3), as well as for studying peptide and other small molecule interactions which exhibit inhibitory effects (4-7).

There are many ways to design an ELISA and most of those are reviewed in T.G. Wreghitt et al. (8). Briefly, 96 well plastic plates are coated with a protein. A ligand molecule with an affinity for the protein, which has been adsorbed to the solid phase, is then added and allowed to bind. An inhibitory molecule can also be included to make a competitive assay. After washing off unbound ligand, an enzyme-linked antibody which recognizes the ligand molecule is added, the plate is washed again, an enzyme substrate is added which is converted to a colored product, and absorbance readings are taken. Results are compared to a standard curve which must be generated each time an assay is run. The standard curve is generated by pipetting serial dilutions of the molecule which binds to the adsorbed protein into the wells, thoroughly washing the wells, and assaying. ELISA theory is elegantly simple; however, there are many problems which can be encountered at various stages of the assay. The utility of this assay relies on the physical interactions of all the molecules involved. Thus if the adsorbed protein assumes a

conformation which renders it unable to bind to the target molecule of interest, then the assay is useless. Also, non-specific interactions can cause problems with these assays. Peptides can often be quite "sticky" and bind to the plastic surfaces, or to secondary antibodies, without specificity. Stringent controls, including wells which contain only peptide and wells which contain only adsorbed protein, are tested for their non-specific interactions with primary and/or secondary antibodies and are a prerequisite to proper data analysis.

Tryptophan fluorescence quenching is another method which can be used to obtain binding information. In most cases, molecules which are raised to an excited electronic state by absorption of radiant energy return to the ground state by radiationless transfer of the excitation energy to surrounding molecules in the form of heat. Occasionally, a molecule will lose only part of its energy of excitation by radiationless transfer and will reradiate some of the energy. This gives rise to the phenomenon called fluorescence (9). In a system where the components are protein P and ligand L and only P is a fluorescent species, the electronically excited P may emit a quantum of energy by fluorescence. The fluorescence efficiency of tryptophan depends on its environment, solvent exposure, and flexibility. When a ligand binds to the protein, it is often found that the tryptophan environment changes enough to show detectable fluorescence efficiency change (10, 11). The wavelength of tryptophan's maximum fluorescence emission varies from 300-350 nm depending on the polarity of the binding site, while its excitation wavelength maximum is 285 nm (12). Tryptophanyl fluorescence quenching studies of proteins, including a thorough description of the kinetics of quenching reactions, have been described elsewhere (13, 14). The ideas presented in these references were applied to the system at hand. The tryptophan fluorescence of gp120 was excited at 280 nm and fluorescence emission spectra were then recorded in the presence of increasing Peptide 3 concentrations; the data was analyzed and plotted in the manner described in the Results and Discussion section.

Materials and Methods

ELISA Experiments

These experiments were conducted using HIV-1_{IIIB} gp120 Capture ELISA Kits from Intracel (Cambridge, MA) and peptides which were synthesized as discussed in Chapter 2. The ELISA plates were polystyrene Dynatech Emulon Z plates which were coated with 500 ng of CD4 followed by 1% BSA per well. The plates were used within 30 days of purchase and were stored at -70° C. Since the plates were divided into strips, it was convenient to remove only the wells needed for the experiment at hand. The wells to be used were washed with 300 μl/well of 1x Wash Buffer immediately after removal from the freezer. After 30 seconds the wash was dumped and the plate was patted dry against a stack of paper towels. The wells were then washed 2 more times with 300 μl/well of 1x Wash Buffer. The final wash was left in the wells for 2 minutes. The wash was dumped and again the plate was patted dry against a stack of paper towels, however not to the point of allowing the wells to completely dry out. One hundred microliter serial dilutions ranging from 500 ng/ml to 0.25 ng/ml of Positive Reference gp120 were added to the wells designated for a standard curve determination. Peptide samples ranging in concentration from 23.5 μM to 6.6 mM were prepared. The total volume of sample in the wells was 100 μl. In competition experiments, 50 μl of the peptide was added to 50 μl of either 62 or 250 ng/ml Positive Reference gp120 to test for inhibition of binding. Multiple blanks consisting of the supplied Diluent/Blocking buffer as well as 50 μl of peptide + 50 μl Diluent/Blocking Buffer were run. After all of the samples were added to the wells, the plate was covered with plastic wrap and incubated at room temperature for 1 hour. The contents of the wells were then discarded and the wells were washed 3 times with 300 μl/well of 1x Wash Buffer, as before. One hundred microliters of Anti-gp120-HRP

Detector Antibody were added to each well, the plate was covered with plastic wrap, and incubated at room temperature for 1 hour. The contents of the wells were then discarded and the wells were washed 3 times as previously described. The TMB (3',3',5',5'-tetramethylbenzidene) substrate was prepared per the kit instructions, and then 200 μl of it was added to each well. The plate was left at room temperature for 30 minutes until the blue color developed and then 50 μl of 2N H_2SO_4 was added to each well to stop the reaction. The plate was then read at 450 nm using a BIOTEK EL-311S microplate reader. The absorbance was then plotted as a function of peptide concentration, as is seen in the Results and Discussion section.

Fluorescence Experiments

These experiments were performed with HIV-1_{sf2} gp120 (expressed protein from CHO cells) which was a gift from Chiron (Emeryville, CA) and Peptide 3 (synthesized as described in Chapter 2) on a Photon Technologies Inc. QM1040 spectrafluorometer in a QS Hellma 3x3 mm pathlength cuvette requiring 150 μl samples. Initially an excitation spectrum of Peptide 3 was taken at an emission wavelength of 410 nm, with the photomultiplier voltage at 625 and slit widths at 10 nm. The peptide concentration was 2.7 mM in PBS at pH 6.0. An emission spectrum from 300-480 nm at an excitation wavelength of 280 nm was also measured. Five additional Peptide 3 samples were prepared in PBS at pH 6.0 such that after addition to the gp120 in the cuvette, the peptide concentrations would be: 0.25 mM, 0.5 mM, 1.0 mM, 2.0 mM, and 3.7 mM. These samples were then dried in a Speed-Vac concentrator, so that the volume in the cuvette could be kept constant. Emission spectra from 300-390 nm (excitation at 280 nm, slit widths 10 nm, pm voltage 575) were collected for the 2.87 μM gp120 sample and then after each addition of peptide. The protein was removed from the cuvette and gently mixed

in the microcentrifuge tube which contained the peptide then returned to the cuvette and placed in the fluorometer. This procedure increased the concentration of peptide in the gp120 sample from 0 mM to 3.7 mM. The fluorescence intensity at 340 nm was measured and analyzed as a function of peptide concentration to determine the peptide-gp120 binding affinity.

Results and Discussion

As was discussed in Chapter 2, three CD4 peptides were synthesized and all three were tested for gp120-binding activity by means of the ELISA assay. Peptide 1 did not have as much inhibitory effect on CD4 binding as Peptide 3, so it was not further tested. There were solubility problems with Peptide 2, which went into solution only at concentrations of 0.2 mM or less, and no inhibitory effects were seen at this concentration. Thus experimental efforts were concentrated on Peptide 3. The ELISA data for Peptide 3 yielded the graph shown in Figure 5. Only the mean values of data from multiple experiments were plotted, so data from peptide concentrations which were not replicated were not included. The data at 2 mM and 6.6 mM show a decrease in absorbance as the peptide concentration increases, as would be expected from a peptide which inhibits the binding of gp120 being measured in the assay. The wells were coated with CD4 and gp120 binding was assayed by means of an HRP (horseradish peroxidase) conjugated monoclonal antibody to gp120. As more of the peptide binds to the gp120, less is available for binding CD4, which means less gp120 will be immobilized in the wells and the absorbance will decrease. Since Peptide 3 showed inhibitory activity, the ELISA data was further analyzed to estimate the dissociation constant (K_d).

The following equation was used:

$$K_d = [peptide]/[CD4] \times (4x10^{-9})$$

The first term in the equation is the ratio of the concentration of peptide at which it exhibits 50% inhibition of gp120-CD4 binding to the concentration of CD4 in the well. This CD4 concentration was $9x10^{-9}$ molar. The second term is the K_d of CD4 binding to gp120 (15-17). Calculations of the K_d of Peptide 3 from 5 sets of experimental data yielded a range of K_d values from 0.2 mM to 1.0 mM. These values are within the range useful for Tr-NOESY NMR experiments.

Tryptophan fluorescence quenching experiments were run on $gp120_{sf2}$ and Peptide 3. There is only one aromatic residue in Peptide 3, Phe43 which is a very weak fluorescer, so the observation of the protein tryptophan fluorescence should not be affected by any fluorescence from the peptide. However, initial emission and excitation fluorescence experiments were done to see if Peptide 3 exhibited any fluorescence, which is sometimes the case with peptides which have been synthesized from Fmoc amino acids (John Mills, personal communication). The initial fluorescence spectrum of Peptide 3 did not indicate any interfering fluorescence. In the next experiment, excitation spectra were recorded of the gp120 with increasing concentrations of Peptide 3 (Figure 6). As is obvious from the graph, with successive increases in Peptide 3 concentration, the tryptophan fluorescence of the gp120 was quenched. With increasing concentrations of peptide, however, there was also a noticeable increase in fluorescence occuring at ~380 nm. It was determined that the source of this 380 nm fluorescent contamination was plasticizer in the tube in which the peptide stock solution was made for that particular experiment. The peptide solution had been prepared in a 1.7 ml microcentrifuge tube for this experiment, whereas it had been prepared in a 0.5 ml PCR tube for the previous experiment in which this contaminant

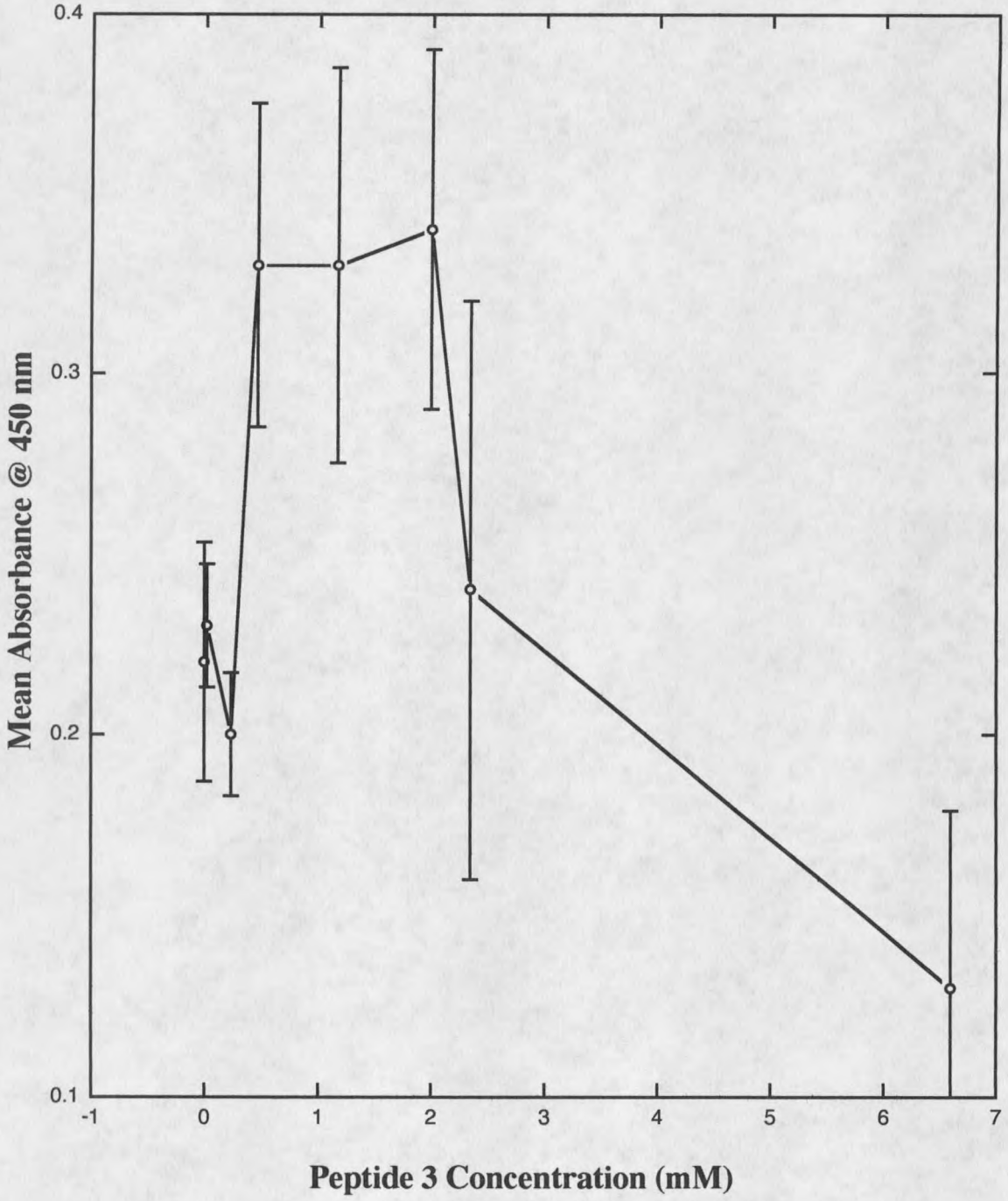


Figure 5. Inhibition of binding of HIV-1_{IIIB} gp120 to CD4 due to increasing concentrations of Peptide 3. Peptide concentrations ranged from 0.0 mM to 6.6 mM. The means of 5 sets of experimental data are plotted with error bars which represent the standard error of the mean.

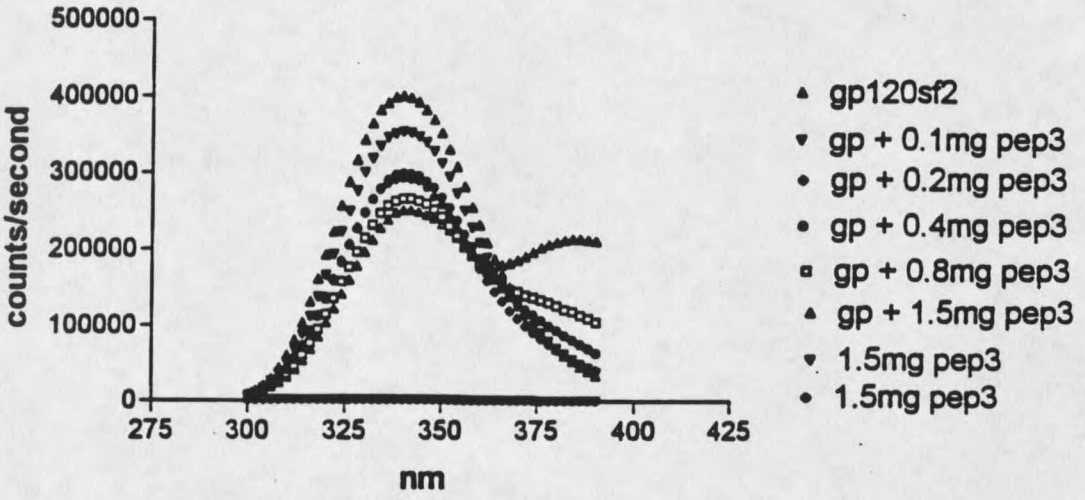


Figure 6. The quenching of fluorescence of HIV-1_{sf2} gp120 by acetylated Peptide 3. Data were obtained at pH 6.0, PBS, 25° C, excitation at 280 nm, emission at 390 nm.

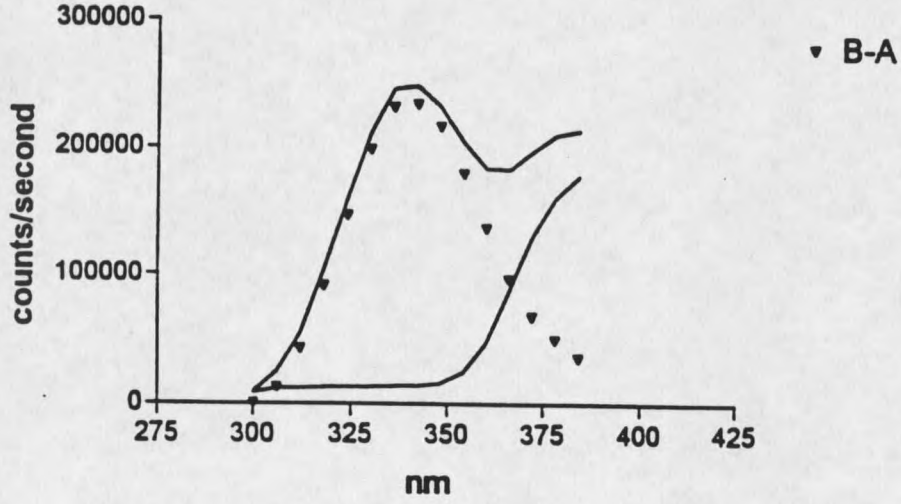


Figure 7. Example of the subtraction of the fluorescence donated by the plasticizer at the highest peptide concentration. Upper trace is gp120 + 3.7 mM Peptide 3; lower trace is 3.7 mM Peptide 3; filled triangles represent the difference spectrum.

fluorescence was not seen. The peptide samples for the quenching experiments were dried in 0.5 ml tubes, so no additional contamination was caused by the transfer of the protein sample to these tubes. The fluorescence of the contaminated peptide solution was measured directly and subtracted from the emission spectra on a percentage contribution basis to correct for this as shown in Figure 7. Thus, at higher peptide concentrations there was a greater contribution to the fluorescence spectrum for the contaminant than at lower peptide concentrations, so a greater percentage was subtracted from those spectra. While the data from the fluorescence experiments looked promising, a control experiment was conducted to determine whether or not the drop in fluorescence was due to the increasing Peptide 3 concentration, or was due to a loss of protein during the transfer of gp120 from microcentrifuge tubes. Unfortunately, the latter was found to be the cause of the decreased fluorescence and thus has invalidated the potential use of stopped-flow fluorescence in the determination of the off rate of Peptide 3.

The K_d value of Peptide 3 obtained from the averaged ELISA experiments was about 1 mM. Because the K_d was sufficiently low to suggest that Peptide 3 might not be a tight-binding peptide, this peptide was chosen for the secondary and tertiary structure analysis of the bound HIV-1 gp120.

References Cited

1. Moore,J.P., McKeating,J.A., Jones,I.M., Stephens,P.E., Clements,G., Thomson,S. and Weiss,R.A. (1990) Characterization of recombinant gp120 and gp160 from HIV-1: binding to monoclonal antibodies and soluble CD4, *AIDS* 4, 307-315.

2. Hinkula,J., Gidlund,M., Persson,C., Osterhaus,A. and Wahren,B. (1994) Enzyme immunoassay (ELISA) for the evaluation of antibodies directed to the CD4 receptor-binding site of the HIV gp120 molecule, *J. Immunol. Methods* 175, 37-46.

3. Gilbert,M., Kirihara,J. and Mills,J. (1991) Enzyme-linked immunoassay for human immunodeficiency virus type 1 envelope glycoprotein 120, *J. of Clinical Microbiology* 29, 142-147.

4. Ghetie,V., Wheeler,T., Scott,D., Uhr,J.W. and Vitetta,E.S. (1992) A CD4-derived peptide carrier blocks acute HIV-1 infection in vitro and binds to gp120 in the presence of Walter-Reed Stage 1-6 HIV+ sera, *AIDS Res. Hum. Retrovir.* 8, 1945-1948.

5. Richalet-Secordel,P.M., Deslandres,A., Plaue,S., You,B., Barre-Sinoussi,F. and Van Regenmortel,M.H.V. (1994) Cross-reactive potential of rabbit antibodies raised against a cyclic peptide representing a chimeric V3 loop of HIV-1 gp120 studied by biosensor technique and ELISA, *FEMS Immunol. Med. Microbiol.* 9, 77-88.

6. Richalet-Secordel,P.M., Zeder-Lutz,G., Plaue,S., Sommermeyer-Leroux,G. and Van Regenmortel,M.H.V. (1994) Cross-reactivity of monoclonal antibodies to a chimeric V3 peptide of HIV-1 with peptide analogues studied by biosensor technology and ELISA, *J. Immunol. Methods* 176, 221-234.

7. Ferrua,B., Tran,T.T., Quaranta,J.F., Kubar,J., Roptin,C., Condom,R., Durant,J. and Guedj,R. (1994) Measurement of the anti-HIV agent 2',3'-didehydro-2',3'-dideoxythymidine (D4T) by competitive ELISA, *J. Immunol. Methods* 176, 103-110.

8. Wood,H.C. and Wreghitt,T.G. (1990) Techniques. In: *ELISA in the Clinical Microbiology Laboratory*, pp. 1-10. Eds. T.G. Wreghitt and P. Morgan-Capner. Public Health Laboratory Services, London.

9. Mathews,C.K. and van Holde,K.E. (1990) *Biochemistry*, Benjamin/Cummings, Redwood City.

10. Connors,K. (1987) *Binding Constants*, John Wiley & Sons, New York.

11. Moore,J.W. and Pearson,R.G. (1981) *Kinetics and Mechanism, 3rd edition.*, John Wiley & Sons, New York.

12. Kyte,J. (1995) *Structure in Protein Chemistry*, Garland Publishing, New York.

13. Eftink,M.R. and Ghiron,C.A. (1981) Flourescence quenching studies with proteins, *Anal. Biochem.* 114, 199-227.

14. Lehrer,S.S. (1971) Solute pertubation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion, *Biochemistry* **10**, 3254-3263.

15. 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.

16. 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.

17. 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.

CHAPTER FOUR

PREPARATION OF HIV-1 GP120 FOR NMR STUDIES

Introduction

In order to conduct the Tr-NOESY NMR studies, it was necessary to have milligram quantities of gp120. Initially, purification of gp120 was accomplished in our lab following a published protocol by Culp et al. (1). However, at the conclusion of two rounds of in-house purification, it was decided that a collaborative effort with Chiron Corporation (Emeryville, CA) would be a more efficient and expedient way of obtaining the needed gp120. A discussion of both the in-house purification and Chiron's purification of gp120, as well as additional preparation of gp120 for the NMR experiments will follow.

Materials and Methods

Materials for Purification of HIV gp120 from Insect Cells

SmithKline Beecham (King of Prussia, PA) kindly supplied a vial of *Drosophila melanogaster* Schneider 2 cells which efficiently express a secreted, truncated form of HIV-1 envelope $gp120_{IIIB}$ protein in a regulated manner (1). The AIDS Research and Reference Reagent Program (Rockville, MD) provided small quantities (~100 ug) of the following: HIV-1_{sf2} gp120, anti-HIV-1_{IIIB} gp120 producing Hybridoma cells (902), and anti-HIV-1_{IIIB} gp120 monoclonal antibody (1501). SP-Sepharose FF was purchased from Pharmacia (Piscataway, NJ); ImmunoPure Immobilized Protein A and a Mouse IgG1 Mild

Elution Buffer Kit were purchased from Pierce (Rockford, IL). An Affi-Gel HZ Immunoaffmity kit from BioRad (Hercules, CA) was also used. Shields and Sang media, gentamycin, and fetal bovine serum (FBS) were purchased from Sigma (St. Louis, MO).

Methods for Purification of HIV gp120 from Insect Cells

Cultures for both the *Drosophila melanogaster* Schneider 2 cells and for the NIH hybridoma line were simultaneously maintained. The hybridoma cells were cultured in large 175 cm^2 Nunc filter cap flasks according to Chesebro et al. (2, 3) and then allowed to overgrow and die for optimal production of the monoclonal antibody to gp120. Three hundred milliliters of supernatant were purified according to the protocol included in the Pierce ImmunoPure Protein A and Mouse Anti-IgG1 kit. Purified fractions of monoclonal antibody to gp120 were then collected to make the immunoaffinity column using the BioRad Affi-Gel Hz Immunoaffinity kit. Throughout the making of the immunoaffinity column, fractions were checked via SDS-PAGE and Western blot analysis utilizing a secondary polyclonal rabbit anti-mouse antibody which was alkaline phosphatase linked to ensure the presence of purified monoclonal antibody.

Production of the gp120 and the monoclonal antibody to gp120 was carried out simultaneously. Scientists at SmithKline Beecham have manipulated the gp120 gene such that the signal sequence and first 31 amino acids of the mature gp120 protein have been removed and replaced with the signal sequence and N-terminus of human tissue plasminogen activator (tPA). Correct cellular processing of the tPA signal sequence results in a secreted evelope protein lacking the first 31 amino acids from the mature form of gp120 and containing 4 amino acids derived from tPA at its amino terminus. Transcription of the gp120 is under the control of the Drosophila metallothionein promoter; thus, addition of $CdCl_2$ or $CuSO_4$ is necessary for induction and protein production. It was determined

that inducing the Schneider cells at a density of 5.5-8 x 10^6 cells/ml for 96 hours with 10 μM $CdCl_2$ yielded optimal gp120 production. The *Drosophila melanogaster* Schneider 2 cell line was maintained in Shields and Sang media supplemented with 10% FBS and 1 mg/ml gentamycin. Purification was conducted on a 1-liter culture scale according to the protocol by Culp et al. (1). Fractions were checked for the presence of purified protein via SDS-PAGE and Western blot analysis utilizing a primary mouse monoclonal antibody to gp120 which was linked to horseradish peroxidase.

Materials for Preparation
of the Chiron gp120 for the NMR Experiments

Purified HIV-1_{sf2} gp120 in the amount of 5 milligrams was provided by the Chiron Corporation (Emeryville, CA). Sodium chloride and sodium phosphate were purchased from Sigma (St. Louis, MO) and sodium azide came from Matheson Coleman and Bell (Norwood, OH). D_2O + TSP was purchased from Aldrich (Milwaukee, WI) and 5 mm NMR tubes which were high-grade 535-PP tubes were bought from Wilmad (Buena, NJ). Amicon (Beverly, MA) Centricon 30 concentrators were also used.

Methods for Preparation
of the Chiron gp120 for the NMR Experiments

The entire gp120 gene (Glu31-Arg509) was cloned, and the signal sequence removed and substituted with that of human tPA. The 5' end of the envelope gene encoding gp160 of HIV-1_{sf2} was engineered for direct expression of gp120 (omitting gp41) by inserting an *Nhe*I restriction endonuclease site 5' to the sequences encoding Glu31, so that the natural signal sequence could be replaced by the human tissue-type plasminogen activator (tPA) signal sequence shown to direct efficient secretion of heterologous gene products. The glycosylated gp120 protein was expressed in CHO cells. A detailed discussion of the

nonaffinity purification of Chiron's gp120 can be found in Scandella et al. (4). Briefly, concentrated cell culture filtrate was applied to a DEAE anion-exchange column and then the unbound fraction was made 40% saturating in ammonium sulfate, centrifuged to remove the precipitate that formed, and then applied to a phenyl hydrophobic interaction column (HIC). A purification of ~10-fold was obtained on the phenyl column. The $rgp120_{sf2}$ peak was again brought to 40% saturation in ammonium sulfate, then applied to an ether HIC column. This step yielded a 2- to 3-fold purification. The remaining lower molecular weight contaminants were removed using high-resolution gel-filtration chromatography on an HPLC. Fractions were analyzed by SDS gel electrophoresis with Coomassie Blue staining. The purity was determined to be 94.8% by SDS-PAGE under reducing conditions; apparent molecular weight was ~120,000. The N-terminus has been sequenced and shown to be Glu31Lys32 (4), a correct match to the previously published sequence for viral gp120 (5). The protein was delivered to us at a concentration of 1.72 mg/ml in 40 mM sodium citrate, 280 mM NaCl, pH 6.0. Before the protein could be used in the NMR experiments, it was necessary to greatly reduce the amount of sodium citrate present and to concentrate the protein into a smaller volume. For this task we used Amicon Centricon 30 concentrators.

The ultrafiltration membranes in all Centricon concentrators contain ~2 μl of glycerin which yields interfering NMR peaks just as the sodium citrate does. It was, therefore, important to thoroughly pre-rinse the membranes. This was done by first adding 1 ml of 0.1 N NaOH to the concentrator and centrifuging at 800 g in an SS-34 rotor at 4° C for ~45 minutes or until most of the sample ran through. The filtrate was discarded and 2.0 ml of PBS (150 mM NaCl, 25 mM NaH_2PO_4-H_2O, pH 6.0) were added and centrifuged as before, except for a longer period of time (approximately 40 minutes). This was repeated a second time. It should be noted that at no time was the membrane allowed to become dry.

This rinsing was found to be sufficient to reduce the amount of glycerin such that its contribution to the NMR spectrum was negligible.

The concentration of the gp120 sample required several rounds of centrifugation. Two milliliters of HIV-1_{sf2} rgp120 was added to the centricon. This was centrifuged down to 350 μl in 3 repetitive rounds in order to reduce the sodium citrate buffer from 40 mM to 0.625 mM and to concentrate the sample volume such that the gp120 concentration would be very high (~82 μM). In the first two centrifugation rounds, the sample was spun in an SS-34 rotor at 800 g, 4° C, for 50 minutes until the volume was reduced to 500 μl. Following each round, the retentate volume was brought back up to 2 mls with PBS. After the third round, the final volume of the retentate was 350 μl. A small aliquot of this was saved for UV-VIS spectrophotometric analysis to more accurately determine the protein concentration. The remainder of the concentrated protein was placed in an acid-washed and heat sterilized high-grade 535-PP Wilmad NMR tube and 1D ^{1}H NMR spectra were collected to determine that indeed the glycerin and sodium citrate had been reduced to non-interfering concentrations. The gp120 was then placed in a sterile eppendorf tube with 1.9 mgs (2 mM), final concentration, N-acetylated Peptide 3. This was mixed gently until the peptide was completely in solution, at which point 30 μl of D_2O +TSP and 0.02% NaN_3 was added to the sample. The pH was determined and adjusted to pH 6.0 using 0.1 N NaOH. The sample was then placed back into the NMR tube. A bottom Vespel susceptibility plug (6) was added and allowed to settle to the bottom of the tube. The sample was purged with cold argon, the top Vespel susceptibility plug was added, and the NMR tube was capped. The cap was wrapped in parafilm, and the tube was stored at 4° C until the experiments could be conducted. It should be noted that upon examination of the actual gp120 concentration by UV-VIS spectrophotometry, it was found that ~ 20% of the

protein had been lost in the centrifugation, thus making the actual gp120 concentration in the NMR tube 66 μM.

Results and Discussion

In the in-house purification, the monoclonal antibody to gp120 was successfully purified and used to make an immunoaffinity column, which was successfully used to purify HIV-1_{IIIB} gp120 as proven by SDS-PAGE and Western blot analysis. However, after two rounds of purification, the in-house production and purification efforts were abandoned. It was decided that a collaborative effort with Chiron Corporation would be a more expedient way of obtaining the necessary milligram quantities of gp120. Maximal production of purified gp120 via the in-house protocol (1) is 1-2 mg/liter as reported by Morikawa et al. (7). The gp120 produced by the *Drosophila melanogaster* Schneider 2 cell line has a glycosylation pattern which is less complex than that of the same protein produced in a mammalian cell line (8-10). Chiron supplied us with 5 mgs of purified HIV-1_{sf2} gp120 from CHO cells. These mammalian Chinese hamster ovary cells glycosylate proteins in a similar manner to human cells; it more closely resembles the glycosylated human protein than the gp120 produced in the *Drosophila* cells. Although comparisons of sCD4 binding for the *Drosophila* produced gp120 versus the viral gp120 demonstrated equivalence of the 2 proteins (1), there is still much controversy surrounding the role of glycosylation of gp120. Bearing in mind that one of the goals of this work is its applicability to human disease, the decision was made to go with a more "human-like" protein.

Another issue which needed to be addressed was the stability of the gp120 once it was placed in the NMR tube and allowed to remain at 4° C for extended periods of time. Due to

the limited quantity of gp120 we had to work with, it was necessary to run as many experiments as efficiently and quickly as possible. At several points in our study, this meant salvaging protein and recycling it for further experiments. At one point, one of our NMR samples was stored for 3 1/2 months at 4° C until it was possible to finish our experiments at Wisconsin's NMR facility (see Chapter 5). The sample was stored at 4° C under Argon with 0.02% NaN_3; as expected the protein lost some of its CD4-binding activity. The loss of the gp120 protein's activity was measured in an ELISA assay which determined the protein's ability to bind sCD4 and a monoclonal antibody to the V3 loop of $gp120_{sf2}$. The gp120 sample retained ~ 60% of its ability to bind the MAb and ~ 40-60% of its ability to bind sCD4.

References Cited

1. Culp,J.S., Johansen,H., Hellmig,B., Beck,J., Matthews,T.J., Delers,A. and Rosenberg,M. (1991) Regulated expression allows high level production and secretion of HIV-1 gp120 envelope glycoprotein in Drosophila Schneider cells, *Bio/Technology* **9**, 173-177.

2. Chesebro,B. and Wehrly,K. (1988) Development of a sensitive quantitative focal assay for human immunodeficiency virus infectivity, *J.Virol.* **62**, 3779-3788.

3. Pincus,S.H., Wehrly,K. and Chesebro,B. (1989) Treatment of HIV tissue culture infection with monoclonal antibody-ricin A chain conjugates, *J. Immunol.* **142**, 3070-3075.

4. Scandella,C.J., Kilpatrick,J., Lidster,W., Parker,C., Moore,J.P., Moore,G.K., Mann,K.A., Brown,P., Coates,S., Chapman,B., Masiarz,F.R., Steimer,K.S. and Haigwood,N.L. (1993) Nonaffinity purification of recombinant gp120 for use in AIDS vaccine development, *AIDS Res. Hum. Retrovir.* **9**, 1233-1244.

5. Robey,W.G., Arthur,L.O., Matthews,T.J., Langlois,A., Copeland,T.D., Lerche,N.W., Oroszlan,S., Bolognesi,D.P., Gilden,R.V. and Fischinger,P.J. (1986) Prospect for prevention of human immunodeficiency virus infection: Purified 120-kDa envelope glycoprotein induces neutralizing antibody, *Proc. Natl. Acad. Sci. USA* **83**, 7023-7027.

6. Bax, A. and Ray, F. (1995) Small samples and Vespel plugs, *36th Experimental NMR Conference* Abstract.

7. Morikawa,Y., Overton, H.A., Moore, J.P., Wilkinson, A.J., Brady, R.L., Lewis, S.J. and Jones, I.M. (1990) Expression of HIV-1 gp120 and human soluble CD4 by recombinant baculoviruses and their interaction in vitro, *AIDS Res. Hum. Retrovir.* **6**, 765-773.

8. Lis,H. and Sharon,N. (1993) Protein glycosylation: Structural and functional aspects, *Eur. J. Biochem.* **218**, 1-27.

9. Davis,T.R., Shuler,M.L., Granados,R.R. and Wood,H.A. (1993) Comparison of oligosaccharide processing among various insect cell lines expressing a secreted glycoprotein, *In Vitro Cell. Dev. Biol.* **29A**, 842-846.

10. Miller,L.K. (1988) Baculovirus as gene expression vectors, *Ann. Rev. Microbiol.* **42**, 177-199.

CHAPTER FIVE

SECONDARY AND TERTIARY STRUCTURE ANALYSIS OF FREE AND GP120-BOUND PEPTIDE 3 BY CD, 2-D ^{1}H NMR, AND MOLECULAR DYNAMICS

Introduction

Circular Dichroism

Circular dichroism (CD) is a chiroptical phenomenon which is sensitive to the detailed structures of proteins and peptides. CD is caused by differential absorption of left- and right-circularly polarized light with chiral molecules. Considerable information concerning the structure of proteins and peptides in solution can be obtained from measurement of their optical activity. In proteins the major optically active groups are the amide bonds of the peptide backbone and the aromatic side chains. The lowest energy electronic absorptions of the amides of the polypeptide backbone occur between wavelengths of 180 and 240 nm. In this region, two electronic transitions account for the absorption of light. The first transition is the $n \rightarrow \pi^*$ in which an electron leaves one of the lone pairs on the acyl oxygen and enters the vacant antibonding π^* molecular orbital of the amide. This $n \rightarrow \pi^*$ transition is responsible for the absorption of light at a wavelength of ~220 nm (1). The other transition is the $\pi^\circ \rightarrow \pi^*$ in which an electron leaves the highest occupied nonbonding π molecular orbital. This transition results in the absorption of light at ~200 nm (1).

Polypeptides and proteins may have regions where the peptide chromophores are in highly ordered arrays, such as α-helices or β-pleated sheets. Depending on the mutual orientation of the peptide bonds, the optical transitions of the amide bond can be split into

multiple transitions, the wavelengths of the transitions can be increased or decreased, and the intensity of the transitions can be enhanced or decreased (2). As a consequence of electronic coupling between the split optical transitions, common secondary structure motifs, such as α-helix, β-sheet, or random coil, result in characteristic CD spectra seen in Figure 8.

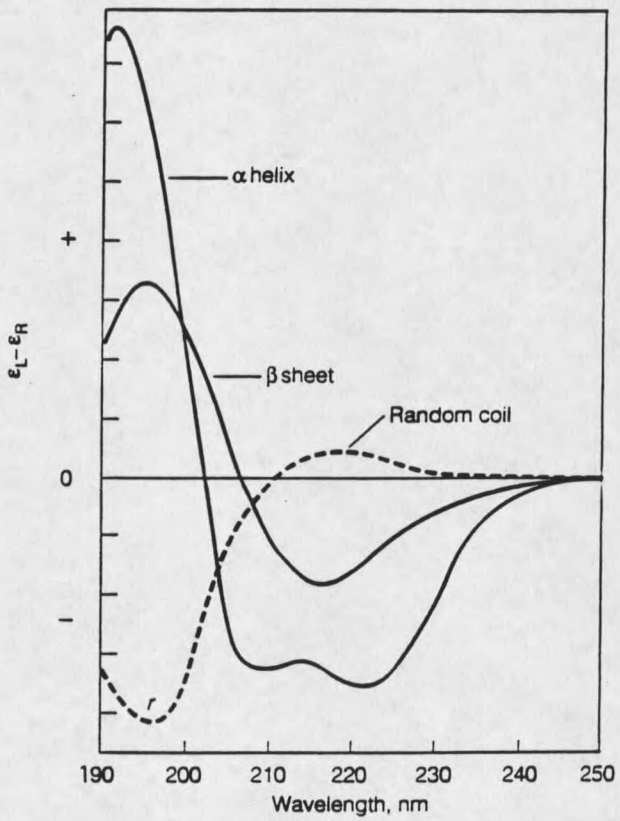


Figure 8. Circular dichroism spectra for polypeptides in various conformations. [From van Holde et al.(3)]

As seen in Figure 8, a polypeptide folded in an α-helix has a CD spectrum distinct from one in a β-sheet. Both of these are distinct from a polypeptide unfolded in a random coil.

In the α-helix CD, the $\pi^{\circ} \rightarrow \pi^*$ transition at 200 nm is split into a positive component which has λ_{max}=191 nm and a negative component which has λ_{max}=205nm. This splitting arises from excitation interactions between the electronic transitions of the peptide amides which interact with a "handedness" relative to the axis of the α-helix (1). The $\pi^{\circ} \rightarrow \pi^*$ transition from a polypeptide in either a β-structure or random coil is unsplit. There is a band at 225 nm which exhibits negative ellipticity from the $n \rightarrow \pi^*$ transition of the peptide bond which, in conjunction with the band of negative ellipticity at 205 nm, gives the CD spectrum of the α-helix its characteristic double minimum (1).

The aromatic residues, tyrosine, phenylalanine, and tryptophan, also absorb light between 180 and 300 nm and have CD spectra if the side chains are in an asymmetric environment with "handedness". To determine the CD spectrum of the polypeptide backbone alone, one can numerically subtract the contributions of the aromatic residues. Circular dichroism is a quantitative spectroscopic technique; consequently, precise knowledge of the polypeptide concentration is necessary for proper CD analysis. Various methods of analysis of protein conformation from circular dichroism are available and have been reviewed by Greenfield (4). A neural network procedure for analyzing CD data based on proteinotopic mapping (5) was employed in the analysis of the data at hand. "Proteinotopic mapping" is a term used to designate the classification of proteins into bidimensional maps. The computer program utilizing this method is named K2d (6). Inspired by the self-organization of the topological maps of the sensorial nervous system during the development of an animal, the Kohonen-type algorithm applied in K2d incorporates the dependence of the CD values of a polypeptide not only on the relative quantities of secondary structure, but also on the length of the chain segments with

different secondary structures. Interactions between amino acids far off in the sequence (tertiary structure interactions) influence CD values, as well, and are also accounted for in this algorithm. K2d consists of a data base of weights and a recall program for determining α-helix and β-sheet structure based on these weights. The program requires data obtained between 240 and 200 nm as input. This is the common recommended wavelength range used in computer programs designed for CD input and analysis of the secondary structure of proteins and peptides in solution.

To determine the structural content of Peptide 3 in solution, a far-ultraviolet CD spectrum was collected. Due to the length of Peptide 3 (24 residues), it had been hypothesized that it might exhibit secondary structure in solution and CD was a useful method by which to test our hypothesis.

Relevant NMR Theory

One of the most important interactions in biology is that of ligands with macromolecules. Ubiquitous interactions include those of antigens with antibodies, substrates with enzymes, hormones with receptors, nucleotides with regulatory proteins, and peptides with phospholipids. The ligands may often be macromolecules themselves. With the exception of examples where the ligand has been crystallized in a complex with the macromolecule, nuclear magnetic resonance (NMR) spectroscopy provides the best method for determining the structure of a bound ligand (7). In this section, the relevant theory of NMR will be presented along with its direct application to the stated research problem.

The One-Dimensional NMR Experiment Numerous books and articles have been written on the principles of NMR which contain treatments at various levels including product

operator and density matrix formalism (8-11). While much of the theory is quite complex, the simplest 1D experiment is relatively easy to picture. The magnetic properties of atomic nuclei are dependent on their spin quantum number (I) and their magnetic moment (μ). The spin quantum number of a nucleus is either an integral or half-integral value or zero. The relationship between the quantum number and magnetic moment is

$$\mu = g\,(I)eh/2mc$$

where e and m are the charge and mass of the proton, respectively, and g is the nuclear g factor (9). Approximately one-third of the different elemental isotopes have nonzero spin quantum numbers. A nucleus with spin 1/2 (e.g. ^{1}H, ^{13}C, ^{15}N, or ^{31}P) is easiest to analyze and will align itself parallel or antiparallel to a static magnetic field ($\mathbf{B}_0$). The nuclear magnetic moments are quite small and the energy difference between these 2 states is only a few millicalories per mole of nuclei in the largest $\mathbf{B}_0$ fields attainable. Only a very few nuclei are present in excess in the lower energy state parallel to $\mathbf{B}_0$ at thermal equilibrium. According to the Boltzmann distribution law ($\exp(2\mu h\pi/kT)$) this excess is less than 1 nucleus in 10,000 for protons (9) but these are the nuclei that give rise to the NMR signal.

To observe the simplest 1D NMR spectrum, a suitable short 90° radiofrequency (rf) pulse is applied to the system along the x axis (B_1) in a laboratory coordinate system which flips the spins into the y axis (z being the designated direction of $\mathbf{B}_0$). The magnetization precesses about the z axis in the xy plane and eventually relaxes back along z. The speed at which the magnetization precesses and relaxes is dependent on the chemical and motional environment of the nuclear spins involved. Receiver coils in a probe surrounding the sample detect the precession frequencies of the protons as a set of oscillating electric currents, induced by the precessing magnetic vectors, which constitute the NMR signal.

Figure 9 shows a vector representation of these processes. The magnitude of the induced voltage decays as a function of time giving rise to a free induction decay (FID) (e in Figure 9). The FID is normally Fourier transformed so that a spectrum in the frequency (f in Figure 9), rather than time, domain is obtained. The difference in the rate of precession of the spins relative to that of a carrier frequency is actually detected, so that the resulting spectrum is designated in terms of the chemical shift from the resonance frequency of some reference compound, typically 3-(Trimethylsilyl)propionic -2,2,3,3-d_4 acid (TSP) for aqueous solutions.

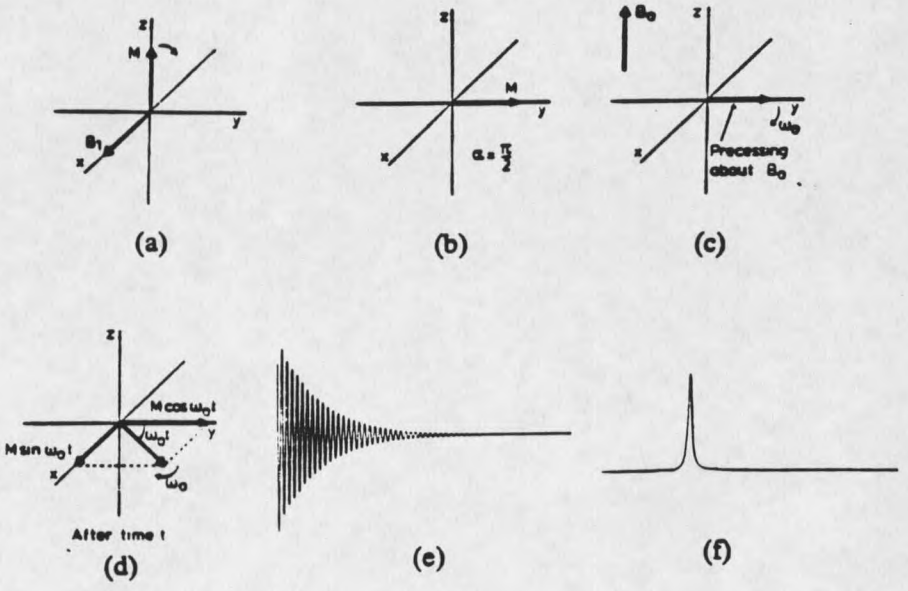


Figure 9. The one-dimensional NMR experiment. (a) Net sample magnetization (**M**) and rf pulse direction ($\mathbf{B}_1$). (b) Effect of a 90°pulse along x. (c) and (d) After the pulse, magnetization begins to precess in the xy plane. (e) FID detected relative to carrier frequency. (f) Result of a Fourier transform on the FID. [From Derome (8)]

Relaxation Experiments There are two primary pathways of relaxation in NMR: longitudinal and transverse. T_1, a time constant, is the common representation for longitudinal, or spin-lattice, relaxation. According to Derome (8): "any time the magnetization is moved away from the z axis, we assume that in the absence of external influences it will return there exponentially with time constant T_1." In response to the various environments that the protons of a specific sample experience, there can be numerous values of T_1. The longest T_1 in the molecule is the one of most interest for our studies; it should be kept in mind that relaxation rates depend on things such as temperature, solution viscosity, molecular size and structure, and in some circumstances the applied magnetic field.

Proton NMR studies carried out on complex spin systems may give rise to significant cross-relaxation effects between the spins. In macromolecules undergoing slow molecular tumbling, (where $\omega\tau_0 > 1$, where τ_0 is the rotational correlation time for overall motion, and ω is the NMR absorption frequency) the nuclear dipolar interaction leads to strong spin diffusion between the different nuclei (12). In these cases, the spin-lattice relaxation times for the strongly dipolar coupled spins are equalized, and the actual rate of relaxation is determined by the group with dynamic behavior which most effectively couples the spin system to the lattice. Multi-pulse experiments typically contain a delay of two to five times the value of T_1 to allow for relaxation of the protons before the pulse sequence is recycled. T_1 values can be measured by an inversion-recovery experiment, such as that diagrammed in Figure 10. T_1 values can then be determined directly from a plot of the signal intensities versus the variable relaxation delay τ.

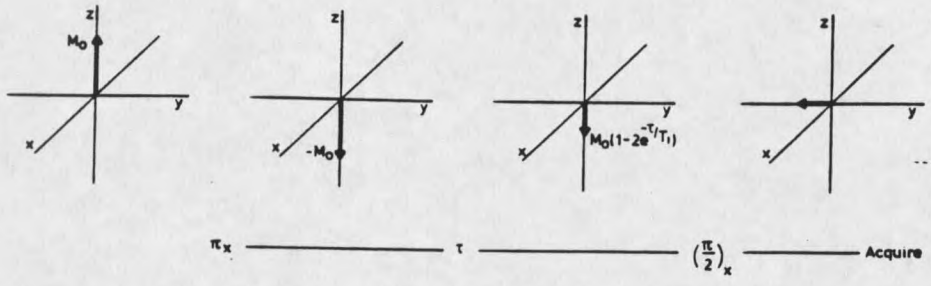


Figure 10. An inversion-recovery pulse sequence for the determination of T_1. [From Derome (8)]

Transverse, or spin-spin, relaxation is a result of magnetization in the xy plane gradually "fanning out" as a result of varying precession frequencies of magnetization within the molecule. Each frequency is know as an isochromat. This process also takes place exponentially and is designated by the time constant T_2. Any process that causes the loss of transverse magnetization contributes to T_2. The collective amplitude of the isochromats can be refocussed and a spin-echo formed (as shown in Figure 11), which can be used to determine the rate of transverse relaxation by plotting the spin echo amplitude as a function of the length of the refocusing pulse train (13). In the intermediate motion region (appropriate for the free peptide) and the slow motion region (appropriate for the bound peptide) T_1 does not effect T_2. T_2 is typically faster than T_1.

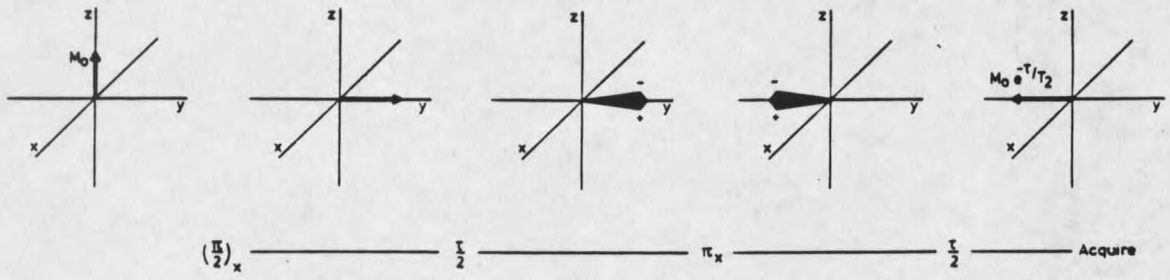


Figure 11. Refocusing magnetic isochromats to form a spin-echo. [From Derome (8)]

T_1 and T_2 are not the only pathways through which relaxation of magnetization occurs. Indeed, even a two-spin system has up to six relaxation paths available to it (8). One of the most important pathways involves dipole-dipole cross relaxation interactions that give rise to the NOE effect, which will be discussed in the following section concerning 2D NMR experiments.

The Two-Dimensional NMR Experiment Jeener first proposed the two-dimensional NMR experiment in 1971. The second dimension of the NMR experiment results from a second time dimension. The 2D experiment can be expressed as the following:

preparation - evolution (t_1) - mixing - detection (t_2)

The preparation time is set so that the sample reaches a steady-state condition in the applied magnetic field. It can involve saturation of some spin states, and it generally ends with the

application of an RF pulse to generate transverse magnetization (9). The detection time t_2 is the same in the 1D experiment and the FT of the signal in this time domain is the ω_2 axis of the spectrum. The evolution time (which is the precession of the magnetization at its Larmor frequency relative to the carrier) is incremented in small steps and a second FT is performed in the t_1 direction obtaining the ω_1 axis. The mixing time allows the spins to interact with each other via scalar coupling or dipolar interactions. Modifications of the mixing time generally distinguish various types of NMR experiments. Two-dimensional NMR experiments are classified as relying on either through bond J (scalar) coupling or through space dipolar interactions.

Two-Dimensional TOCSY TOtal Correlation SpectroscopY (14) is a through bond experiment which can, in favorable cases, reveal the entire spin system of each amino acid residue in one experiment by showing all spin correlations between protons connected by 2, 3, or 4 bonds. This experiment is useful for assignment of the groups of coupled amino acid side chain resonances and associating them with their attached αH and NH protons in a peptide. The basic pulse sequence for a TOCSY experiment is shown in Figure 12.

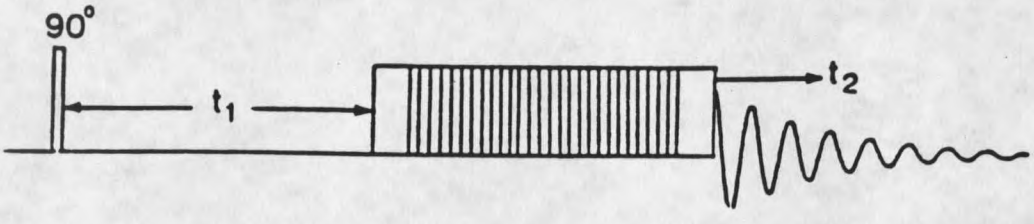


Figure 12. The pulse sequence for the TOCSY experiment.

The first 90° pulse provides the excitation magnetization which evolves during the variable period t_1. Next, a spin-lock is created with a set of short pulses to minimize sample heating. One of the most common spin-lock pulses for TOCSY is the MLEV-17 (15) series which was employed in our experiments. The spin-lock serves to allow flow of magnetization between protons at a rate determined by their J couplings (16) so that most through bond correlations in a system are obtained.

Two-Dimensional NOESY In 2D Nuclear Overhauser and Exchange SpectroscopY, through space interactions can be defined. To understand this particular experiment, it is first necessary to understand NOE's. An NOE is a singular aspect of nuclear relaxation involving a change in intensity of one resonance when the transitions of another are perturbed (8). The NOE is then a manifestation of the attempt of the total system to stay at thermal equilibrium. Figure 13 shows the energy levels and populations of a homonuclear two-spin system prior to irradiation. In the original NOE experiments, the "perturbation" was the saturation of one signal while changes in intensity of the signals of other nuclei were observed. Saturation refers to the elimination of the population differences across one or more transitions. An example would be to saturate both transitions of nucleus *s* and then observe the signals due to *i* after the new equilibrium has been established (Figure 14).

To properly examine NOE's, it is necessary to investigate the relaxation pathways available to nuclear spins in the molecule. The dipole-dipole interaction which gives rise to the NOE signal is in competition with spin-lattice relaxation mechanisms to discharge magnetization. Figure 14 shows a diagram of the energy levels of a simple two-spin system which can exhibit NOE's along different possible relaxation pathways. Spin-lattice interactions occur between the excited spins and other protons in the molecule, the solvent or paramagnetic species in solution.

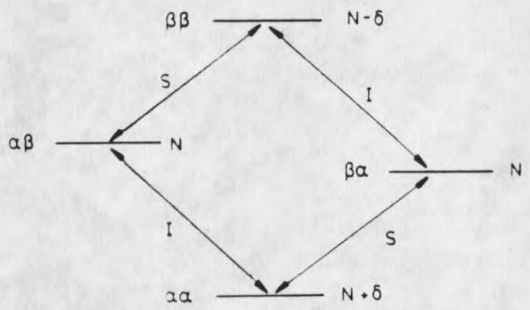


Figure 13. Energy levels and populations of a homonuclear, two-spin system. [From Derome (8)]

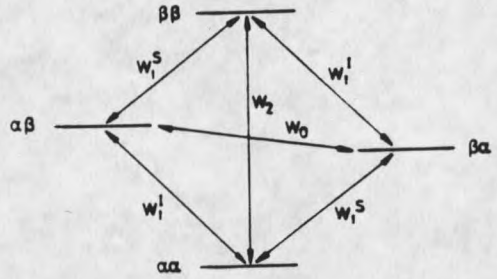


Figure 14. Possible relaxation pathways in a homonuclear, two-spin system. [From Derome (8)]

In a 1D NOE experiment the populations of the levels connected by the S transitions are equalized upon saturation of the S transitions. The W_2 double quantum transition increases the population difference and the intensity of the I transitions to +1/2 +δ/2 (positive NOE), where δ is the population change caused by W_2. The W_0 zero quantum transition decreases the population difference and the intensity of the I transitions to +1/2 - δ/2 (negative NOE), where δ is the population change caused by the W_0 transition. Fast

motions favor W_2 transitions and provide fluctuations with frequency twice the resonance frequency ($\omega_I + \omega_S \approx 2\omega_0$), while slow motions favor the W_0 transition and provide fluctuations with low frequencies ($\omega_I - \omega_S$). With molecular motions near the Larmor frequency (ω_0), the W_0 and W_2 effects cancel and the NOE is zero. Protons in slow motion have negative NOE's, while protons with fast motion have positive NOE's. When spin S is saturated, the NOE signal observed from spin I, $\eta_I(_S)$, is proportional to $1/r^6$, where r is the internuclear distance (8):

$$\eta_{I(S)} \propto \mu\,(\tau_c/\rho^*)(1/r^6)$$

τ_c is the rotational correlation time of the molecule and ρ^* is the nuclear relaxation from mechanisms other than I-S cross-relaxation. The relationship between NOE intensity and rotational correlation time of the molecule is shown in Figure 15.

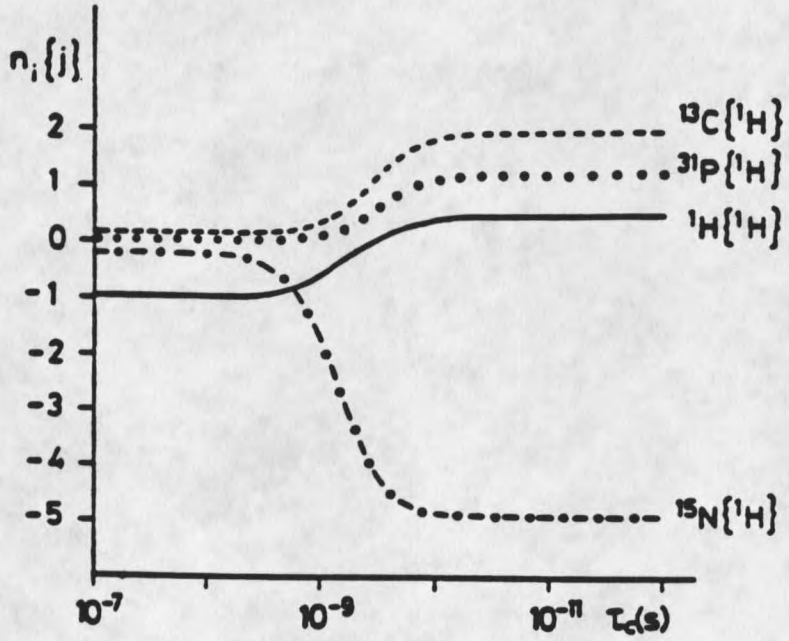


Figure 15. NOE intensity vs. rotational correlation time. [From Wuthrich (17)]

The 2D NOESY experiment uses inversion rather than saturation (used in the simplest 1D NOE) to observe all possible NOE's. The pulse sequence for the experiment is shown in Figure 16.

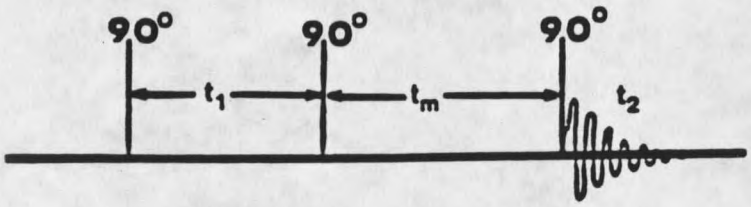


Figure 16. The pulse sequence for the NOESY experiment.

The first two 90° pulses modulate the amplitudes of the different nuclei by their relative chemical shifts. The second pulse generates non-equilibrium spin components inverted along the z axis neglecting T_1. During t_m the z components cross-modulate their amplitudes by NOE cross-relaxation; this leads to off-diagonal cross peaks. By using regularly incremented series of evolution times, each spin obtains an amplitude modulated nonequilibrium population distribution with a frequency offset from the carrier frequency. Thus, all dipole-dipole coupled population transfers occur. The third 90° pulse samples the amplitude of the magnetization transfer, after which detection occurs during t_2.

NOE's can be observed as either positive or negative peaks depending on molecular size, tumbling rates, and the applied field. When studied with higher field NMR spectrometers, small molecules, with short τ_c ($\sim 10^{-11}$ sec.) exhibit positive proton NOE's, and large molecules with long τ_c ($\sim 10^{-8}$ sec.) exhibit negative proton NOE's as can be seen

in Figure 15. The experiment of interest to this study is a variation of the NOESY experiment known as the Tr-NOESY.

Two-Dimensional NMR Experiment Tr-NOESY The 1D version of the transferred nuclear Overhauser enhancement spectroscopy (Tr-NOESY) NMR experiment was first reported by Bothner-By and Gassend (18) as a way to study the conformation of small ligands (L) when they are reversibly bound to large proteins (P):

$$L + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} LP$$

Since then, 2D Tr-NOESY experiments have been used to study the conformations of many different types of small ligands, such as cofactors, nucleotides, peptides, carbohydrates, and drugs when they are bound to large receptors (19-26). A decided advantage of this approach is that, in contrast to full solution structure determination by NMR, Tr-NOESY can provide structural information on proteins that are too large to be studied directly by NMR spectroscopy (21, 27-29). Figure 17 shows a schematic representation of the Tr-NOESY experiment where a linear peptide is in rapid exchange with a macromolecular binding site. The free peptide in solution is typically very flexible and will generally fluctuate between a large number of conformations. When the peptide binds to the macromolecule it is observed to be strongly immobilized and this results in a rapid build-up of negative NOESY peaks between protons on the peptide. The NMR linewidths of protons in the bound peptide (and in the large protein) are too large for them to be observed directly. However, NOESY cross relaxation between spins in the bound peptides is proportional to the bound correlation time and rapidly changes the populations of the nuclear spin levels of protons that are in close proximity to each other in the bound state. When the peptide exchanges off the receptor, the populations of the spin levels characteristic of the bound NOESY's are retained on the free peptide. The cross relaxation

of the free peptide with the solution environment is very slow, since the frequency of the tumbling of the peptide is much too high to couple relaxation to fluctuations in the magnetic environment. If the peptide exchanges rapidly off the protein binding site relative to the bound cross relaxation times, the bound peptide NOESY's can be observed on the sharp lines of the free peptide in the Tr-NOESY experiment.

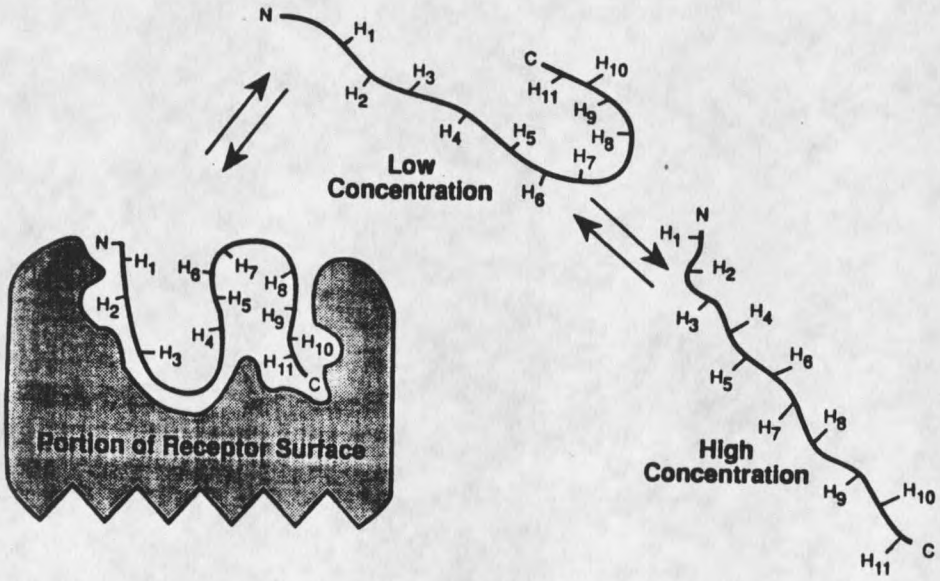


Figure 17. Schematic of the Tr-NOESY experiment with a flexible linear peptide.

The most straightforward treatment of the Tr-NOESY cross peak intensities requires that the ligand of interest must be in fast exchange compared to the fastest cross relaxation rate in the bound state (27, 30-32). It is possible to simulate and analyze Tr-NOESY

experiments if the exchange rate is not at the fast exchange limit (30-32), but the analysis is complicated by a detailed dependence on the exchange kinetics, which causes a lag in the build-up of the NOESY intensities. In the case of sufficiently fast exchange, the observed ligand cross relaxation rates can be simply expressed as:

$$\sigma_{ij}^{obs} = \rho_b \sigma_{ij}^{b} + \rho_f \sigma_{ij}^{f}$$

where ρ_f and ρ_b are the mole fractions of the free and bound ligand and σ_{ij}^{f} and σ_{ij}^{b} are the dipolar cross relaxation rates between proton i and proton j in the free and bound ligand respectively (28, 30, 32). The off rate needed to satisfy the fast exchange condition depends on the size of the protein receptor, since that sets the bound correlation time and the maximum rate of the bound cross relaxation rate. In fast exchange, corrections can be made for any free peptide NOESY cross peak intensities and the cross relaxation rate can be calculated for the bound state:

$$\sigma_{ij}^{b} = (\sigma_{ij}^{obs} - \rho_f \sigma_{ij}^{f})/\rho_b$$

A $T_{1\rho}$ NMR method for determining the ligand dissociation rate (34), under similar conditions as the Tr-NOESY has been employed in our laboratory to determine k_{off}. The requirement for fast exchange means that the ligand must have rather low binding affinity ($K_d \geq 10^{-5}$M). While a straightforward treatment of the Tr-NOESY data relies upon a fast off rate, it is also heavily dependent upon the amount of spin diffusion present in the system.

Campbell and Sykes (7) showed that spin diffusion is present between protons of ligands bound to macromolecules in Tr-NOESY experiments. There is little spin diffusion in the free ligand, because of its small size and correlation time near $1/\omega_0$. However,

when ligands bind to a large macromolecular receptor, the spin diffusion between protons of the ligand and between the ligand and the receptor becomes very similar to the spin diffusion that is present in large proteins. Spin diffusion within the ligand can presumably be accounted for by using the isolated spin pair approximation (ISPA-where two protons are treated as an isolated spin system) to obtain approximate bound structures and full relaxation matrix back calculation of magnetization transfer between all protons in 2D NOESY experiments. Spin diffusion between ligand protons tends to be somewhat less serious in Tr-NOESY than NOESY experiments since each ligand spends only a small fraction of the time bound to the receptor (35, 36). However, a significant amount of magnetization can be exchanged to or from the ligand by spin diffusion involving the large protein or receptor (29, 37, 38). Thus, the spin-echo filtered Tr-NOESY pulse sequence is being developed in the Dratz laboratory at MSU (Figure 18) to reduce the problem of spin diffusion between the ligand and receptor. The spin-echo filter approach takes advantage of the short T_2 of the receptor to dephase the receptor spins during Δ delays, while refocussing the ligand resonances (which have much longer T_2's) after paired Δ delays. The spin-echo filter at the front of the sequence greatly improves the baselines by removing broad features due to the receptor and this improves the ability to measure small NOE's that correspond to the longer range distances in the structures. The spin-echo filter cycles during the mixing time reduce spin diffusion between receptor and ligand by keeping the receptor magnetization almost fully dephased in the transverse plane while the ligand spins are returned to the longitudinal axis for cross relaxation.

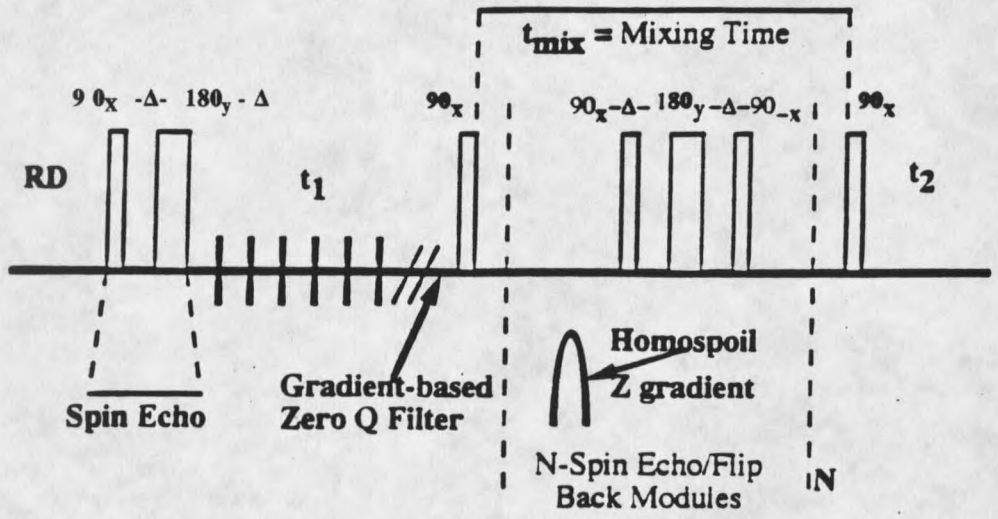


Figure 18. Spin-echo filtered Tr-NOESY pulse sequence.

Computational Methods

Several computer programs were employed, during the course of this research, to refine structures obtained from an ISPA interpretation of the NMR data. These methods are briefly described in the following sections, while specific details are given in the Materials and Methods section.

Minimization Energy minimization moves the atoms in the system to the nearest local minimum - which may not necessarily equal the global minimum. The conformation assumed represents one of many possible conformations that a molecule might assume at 0° K. Forcefields are used to model the potential energy surface of the molecule as a function of distortions of the coordinates of the molecule. The minimizers used in this project were "steepest descents" and "conjugate gradients."

In the steepest descents method, the line search direction is defined along the direction of the local downhill gradient (39). This method is extremely "robust" because it is the method which is most likely to generate a lower-energy structure regardless of what the function is or where it begins, and is particularly useful when structures are far from their global energy minimum.

In a conjugate gradients minimization, each line search deviates somewhat from the ideal direction to the minimum. Successive line search steps correct for this deviation and the path continually oscillates and overcorrects for poor choices of directions in earlier steps (39). The conjugate-gradient minimizer is generally the best choice, as its convergence is much better than steepest descents. However, the radius of convergence of the conjugate - gradient minimizer is not as good as the steepest descents methods and thus it may not converge to the correct minimum if the starting structure is very far from the minimum. "Minimizations are mathematically defined to converge when the point at which the derivatives of the function are zero and the second-derivative matrix is positive definite" (39). A well-minimized final structure of a protein or polypeptide will, generally, have root-mean-square (RMS) values for the atomic derivatives between 0.02 kcal mol^{-1} $Å^{-1}$ and 0.5 kcal mol^{-1} $Å^{-1}$ (39).

Molecular Dynamics Molecular dynamics (MD) solves Newton's equation of motion for a system of atoms. The system contains thermal energy and is allowed to undergo conformational and momentum changes such that the molecules can explore various conformational states. In a dynamics run the temperature of a system is held constant by modeling the system as being in equilibrium with a constant temperature heat bath. Each MD simulation starts with some starting set of coordinants which may be randomized or may come from some other conformations such as an ISPA structure or an extended chain.

MD employs forcefields which have the following contributions:

$$V = V_{\text{van der Waals}} + V_{\text{coulomb}} + V_{\text{dihedral}} + V_{\text{torsion}} + V_{\text{bond}} + V_{\text{angle}} \quad (40).$$

The molecules fluctuate energetically and confomationally using thermal energy to cross energy barriers. Depending on the details of the parameters, MD is less likely than energy minimization to lead to a conformation which is trapped in a local, rather than global, energy minimum.

Simulated Annealing Simulated Annealing (SA) consists of searching for the global minimum of a target function by first substantially reducing the force constants at elevated temperatures to more fully explore conformational space. Low repulsive terms are especially important since they allow atoms to squeeze past one another to explore conformational space and more effectively seek the global minimum. The force constants are gradually and selectively increased while decreasing the temperature until the force constants regain their full values. The target function, for which the global minimum is searched, is the following:

$$F_{\text{tot}} = F_{\text{covalent}} + F_{\text{repel}} + F_{\text{NOE}} + F_{\text{chiral}} + F_{\text{dihedral}} \quad (40).$$

F_{tot} is the effective potential energy in the dynamics calculation. F_{repel} replaces the van der Waals, electrostatic, and hydrogen-bonding potentials of the empirical energy function used in MD. A flow chart which is representative of our SA procedure is seen in Figure 19.

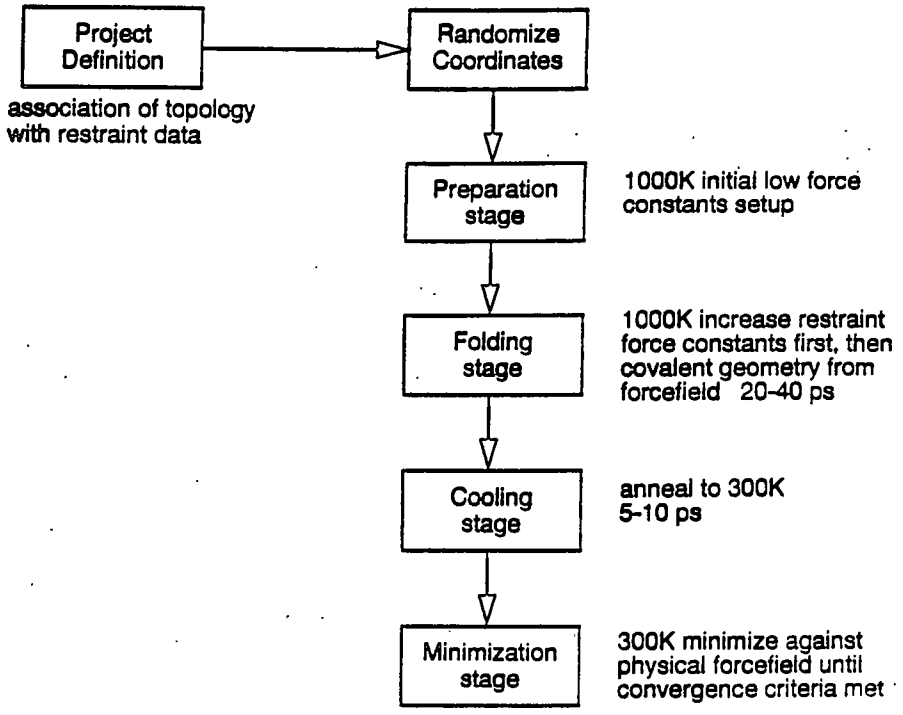


Figure 19. Simulated Annealing Protocol [From (42)]

MARDIGRAS To a first approximation, the intensities of NOE cross peaks are proportional to $1/r_{ij}^{6}$ where r_{ij} is the distance between proton i and proton j (43). This is called the isolated spin-pair approximation (ISPA) which treats the two protons as an isolated spin system and assumes that there are no other cross relaxation pathways to other protons that affect the observed NOESY signals. ISPA ignores spin diffusion, anisotropic tumbling, and segmental motions and thus generates inaccurate or imprecise distance restraints. Two methods that were applied in this project to more completely model the

influence of other protons in the local environment were CORMA (Complete Relaxation Matrix Analysis (44)) and MARDIGRAS (Matrix Analysis of Relaxation for DIscerning the GeometRy of an Aqueous Solution (44, 45)).

CORMA calculates a matrix of cross relaxation mixing coefficients from the experimental data and initial structure data generated by ISPA. "Accurate" intensities which incorporate spin-diffusion and network relaxation effects are calculated by diagonalizing the relaxation matrix (45). The MARDIGRAS program then iteratively uses the matrix generated by CORMA to obtain the best set of distances to fit the observed NOESY constraints and the approximate structure. MARDIGRAS allows distances with observed intensity to change throughout the course of iterations, but keeps constant those distances which are known (i.e. between geminal protons). MARDIGRAS seeks convergence "in a self-consistent set of distances which simultaneously satisfy the experimental intensities and the set of model intensities" (44) which are derived from the approximate structure. Evaluation of the convergence of MARDIGRAS is based upon R and Q factors, which are defined as:

$$R(\tau_m) = \sum_{ij} | E_{ij}(\tau_m) - T_{ij}(\tau_m) | / \sum_{ij} E_{ij}(\tau_m)$$

$$Q(\tau_m) = \sum_{ij} |E_{ij}(\tau_m) - T_{ij}(\tau_m)| / \{\sum_{ij} E_{ij}(\tau_m) + \sum_{ij} T_{ij}(\tau_m)\}$$

$$R^{1/6} = (\sum_{i,j,m} |(A_{ij}^{calc}(\tau_m))^{-1/6} - (A_{ij}^{exp}(\tau_m))^{-1/6}|^2 / \sum_{i,j,m} (A_{ij}^{exp}(\tau_m))^{-1/3})$$

where $E_{ij}(\tau_m)$ is the experimental intensity of the cross peak between protons i and j at mixing time τ_m and $T_{ij}(\tau_m)$ is theoretical intensity (45). Distances calculated from a round of MARDIGRAS are used to generate an improved structure with simulated annealing and/or molecular dynamics resulting in a new starting point for CORMA. The process is iterated until the structures converge.

Materials and Methods

CD Experiments

The CD experiments were performed on a JASCO J-710/720 circular dichrometer. The instrument was calibrated with ammonium d-10-camphorsulfonate as described in the manual. Spectra were recorded from 280 to 190 nm using a Jasco 1 mm pathlength cylindrical QS cell at room temperature. Blank reference spectra of the 50 mM sodium phosphate (NaPP) buffer were collected first in the same cuvette. The concentration of Peptide 3 was 1.5 mM in 50 mM NaPP buffer at pH 6.0. Determination of the peptide's concentration was done via absorbance measurements at 257 nm. Since the peptide had a single aromatic residue, phenylalanine, application of Beer's law ($A=\varepsilon bc$) together with the knowledge of the extinction coefficient of phenylalanine (at $\lambda_{max} = 257$ nm, $\varepsilon = 200$ l/mole·cm) (46, 47) yielded concentration of the peptide.

NMR and MD Experiments

Three samples were prepared for the NMR studies of Peptide 3: the free peptide, the gp120-bound Peptide 3, and a competition experiment which included the addition of rsCD4 to the gp120-bound Peptide 3. The Peptide 3 sample was made 2 mM at pH 6.0 in PBS (150 mM NaCl, 25 mM NaH_2PO_4), 10% D_2O/TSP, and 0.02% NaN_3 (to inhibit microbial growth). The sample was then placed in a sterile high grade NMR tube and purged for 2 minutes with cold Argon, the tube was capped the cap was wrapped in parafilm, and the sample was stored at 4° C until the experiments could be run.

Preparation of the gp120 sample has already been discussed in Chapter 4. The competition experiment with CD4 required the exchange of the PBS buffer in which the

CD4 was originally prepared, so that the addition of CD4 would not change the salt concentration of the Peptide 3-gp120 sample. Concentration and buffer exchange were carried out in a Centricon 10 (Amicon). A thorough pre-rinse was done initially, as described in Chapter 4. The stock CD4 (from UpJohn) consisted of the first two domains (D1 and D2) of CD4 at a concentration of 1 mg/ml in PBS pH 7.4 with 0.015% Tween 80. Six hundred microliters rsCD4 were diluted with 1.4 ml PBS and the concentrator was centrifuged in an SS-34 rotor, at 4° C, 2500 rpm, for 1 hour. Another milliliter of PBS was added and the sample was centrifuged for 3 hours. Centrifugation was done relatively slowly so as to reduce the "pancaking" effect which had been seen with gp120 and had resulted in an approximately 20% protein loss. Absorbance readings at 280 nm were taken to determine the concentration of the concentrated CD4. Twenty nanomoles were lyophylized in a microcentrifuge tube so that upon addition to the gp120-Peptide 3 NMR sample, there would be no increase in volume (thus no dilution of the gp120), and the gp120 and CD4 would both be 66 μM in solution.

NMR spectra were collected on Bruker AM 500, DRX 500, and DMX 750 instruments using 5 mm proton selective probes. The probe used on the DMX 750 at the Nuclear Magnetic Resonance Facility at Madison (NMRFAM) had triple axis gradients. Quadrature detection in t1 was achieved using Time Proportional Phase Incrementation (TPPI) (48). TOCSY spectra (15) were collected at 4° C with an MLEV-17 pulse train. NOESY spectra (49) and Tr-NOESY (33) were collected at 4° C using a 200 ms mixing time and four cycles of spin-echo filtering in the mixing time. For the free Peptide 3, 1024 t1 experiments were recorded consisting of 8192 complex points in t2, while 512 complex t1 experiments were recorded consisting of 8192 complex points in t2 for the bound spectra.

$T_{1\rho}$ relaxation times were measured by setting the spin lock frequency at the peak of interest (ring protons on the the phenylalanine and the glutamine β hydrogens were used

because these showed the largest shifts upon binding) and measuring the peak intensity in the Peptide 3 + gp120 sample as a function of spin lock time and spin lock power levels. The spin lock time was varied over a range of 2-400 ms and the experiment was carried out at spin lock power levels ranging from 20-70 db. The change in $1/T_{1\rho}$ as a function of the log of the spin lock power was then analyzed to obtain k_{off}, which is found from the frequency where there is a breakpoint between between the fast and slow relaxation rates.

Two-dimensional NMR spectra were processed on SGI Indigo^2 workstations using FELIX software from Biosym. Apodization typically applied to the spectra were Gaussian line-broadening window functions in t2 and 70 degree shifted sinebell window functions in t1. Spectral widths of 15000 Hz were used for all spectra collected on the DMX 750. Cubic spline baseline corrections were used in t2 for the NOESY, Tr-NOESY, and the TOCSY. All spectra were collected with the carrier frequency on the water resonance, which was suppressed with soft 90° pulses on the water (hence the term "watergate") surrounding a hard 180 and pulsed field gradient which cleaned up spatial imperfections in the 180s. To reduce spin diffusion effects from the receptor, 4 spin echos of 2 ms were included in the mixing times. Experiments with the free peptide had 1 spin echo in t1 and 4 spin echos in the mixing time, while the Peptide 3 + gp120 experiments had only 1 spin echo in the mixing time. All spectra reported were collected at 4° C and chemical shifts were referenced to TSP.

NOESY cross peak intensities were divided by the volume of a well-resolved single proton diagonal peak volume to correct for cross peak intensity loss due to T_1 relaxation and to provide fractional NOE values. The fractional NOE values were first converted to approximate proton distances using the isolated spin pair approximation (ISPA) and the sidechain β protons of Phe (1.78 Å) as a reference intensity.

Peak volumes were obtained by using the auto peak pick within the FELIX software, and then manually editing the boxes around the assigned cross peaks and using a FELIX utility to measure the volumes (or sum the points in the box). 1-D vector slices were taken of suspect peaks to check for baseline imperfections, overlapping peaks, etc. A total of 116 cross peaks were assigned in the free spectrum, and 122 cross peaks were assigned in the bound. Exact single distance constraints were measured using the Phe β proton scalar peak and upper bounds of 5 Å and lower bounds of 1.8 Å. The fractional NOESY intensities of the free Peptide 3 were subtracted from the bound Tr-NOESY intensities of the gp120-Peptide 3 sample assuming that 1.8% of the peptide was bound to the gp120. This correction is done in order to calculate the Tr-NOE signal of the bound gp120-Peptide 3 since the observed Tr-NOESY intensities are comprised of the sum of contributions from the free and bound species. The NOESY distances were converted to a distance restraint file format using the FELIX database software package (Biosym). Stereospecific assignments were not made and pseudo atom corrections (17) were applied to the appropriate methylene distance constraints.

Molecular dynamics, simulated annealing, MARDIGRAS, and a variety of analysis routines in the Discover , InsightII, and NMRchitect packages (Biosym) were used. Rather than using the initial approximate ISPA distances, an extended chain model of Peptide 3 was built in InsightII and then run through 3 cycles of MARDIGRAS at an effective correlation time of 0.7 ns. Methyl, β, γ, δ protons that were nondistinguishable were treated as pseudoatoms by MARDIGRAS, where the resulting distance was measured from the geometric center of the degenerate protons. All of the amino acids were set to L chirality as part of the restraint file. The torsion angle along the peptide bond (ω) was forced to have a trans conformation by using a very high force constant (300 kcal/mole).

The force constants for the distance constraints were 5 kcal/mole except for the long range NOE's which had values of 10 kcal/mole. This was done because the long range interactions are critical in the determination of the peptide conformation.

We then calculated the structure using simulated annealing (SA) (41). First the extended chain structure coordinates were randomized and energy minimized using very low force constants for the forcefield. The force constants were initially set at 0.1 kcal·mol^{-1}·Å^{-2} and doubled at the beginning of each new cycle to a maximum value of 50 kcal·mol^{-1}·Å^{-2}. This was followed by restrained dynamics calculations at a high temperature (1000° K) with a time step of 1 femtosecond for a total of 30 picoseconds. The restrained dynamics was continued after lowering the temperature to 650° K. The force constant for the NOE forcefield was used at 50% full scale, while the force constant for the covalent forcefield was 70% of full scale for 10 picoseconds with a time step of 1 femtosecond. The restrained dynamics continued by increasing the NOE and covalent force constants to full scale while lowering the temperature from 650° K to 300° K for a total of 8 picoseconds. The structures were then energy minimized resulting in the generation of 15 structures which all fell roughly in the same family. One of the structures was then used in MARDIGRAS and the distance restraints were carefully analyzed (see Table 5 in Results and Discussion).

The distance restraints obtained through MARDIGRAS were employed in the SA procedure, as previously discussed. Out of the 15 structures generated, 4 showed relatively higher total energy and thus exhibited distinct characteristics which did not coincide as well with the remainder of the structures. These all exhibited similar geometries and were thus grouped as one family. Each of the 11 remaining structures from this family were further refined using restrained dynamics. The NOE force constants were doubled from the previous SA calculation to 10 and 20 kcal/mole. Energy minimization followed

by MD at 300° K with a time step of 1 femtosecond followed by another round of energy minimization was conducted until the structures converged (the convergence limit was 0.01 kcal/Å). Two families of structures resulted from the final round of minimization and an analysis of those is discussed in the following section.

Results and Discussion

CD Experiments

The CD spectrum of Peptide 3 can be seen in Figure 20. Upon initial inspection, it suggests that the peptide has very little α-helical character or β-sheet character. Indeed, in order to confirm that the peptide did not have significant secondary structure 41 CD values ranging from 200 nm to 240 nm were submitted to the K2d server which is located on the world wide web at http://www.embl-heidelberg.de/~andrade/k2d.html. The input was CD spectrum values collected in molar ellipticity multiplied by 0.001. The output from the K2d program analyzed Peptide 3 as having 1% α-helical content, 1% β-sheet, and 98% random coil in solution. The K2d neural network arranges a set of proteins based on their ultraviolet circular dichroism spectra in a completely unsupervised learning process (6). A map algorithm was used to obtain protein topological maps. These maps were based on CD data collected from 18 different proteins of known 3D structure. The CD data which is entered into the K2d program is compared to a set of protein topological maps, and matching secondary structural features are weighted and assigned as α-helix, β-sheet, or random coil based on these weights.

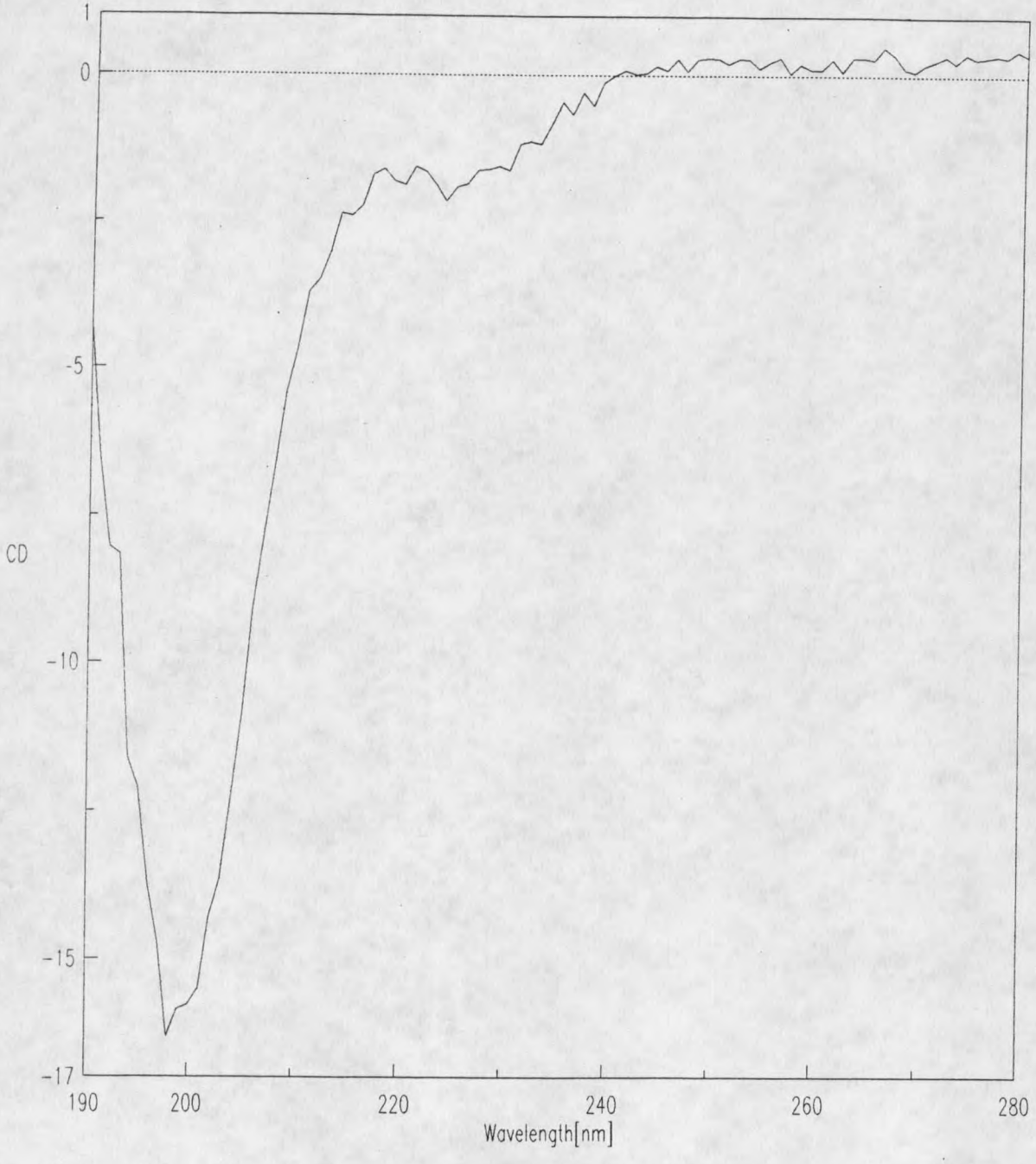


Figure 20. Far-ultraviolet CD spectrum of acetylated Peptide 3 in 50 mM NaPP at pH 6.0. Peptide concentration was 1.5 mM; cell pathlength was 1 mm.

NMR and MD Experiments

The ^{1}H NMR spectra of Peptide 3 was assigned by the combined use of TOCSY and NOESY 2D NMR spectra using standard approaches (17). The TOCSY spectrum leaves some ambiguities in assignments when residues have very similar or overlapping shifts, so NOESY experiments were used to complete the assignments. In a NOESY experiment, the cross peaks are the result of cross relaxation between two protons within an acceptable distance (generally < 5 Å) of each other in space, not because they are part of the same coupled spin system. Most protons in Peptide 3 were successfully assigned with the exception of the the final residue Arg25. Note that the numbering of the peptide started with the Acyl group as residue 1 and ended with Arg25; thus residues 2-25 of Peptide 3 correspond to CD4 36-59. Tables 2 and 3 show the ^{1}H resonance assignments for the free and the gp120-bound Peptide 3. The assignments for the bound were also used in the spectral analysis of the gp120-bound Peptide 3 with CD4.

As was discussed earlier, the straightforward interpretation of the Tr-NOESY experiment relies on the fast off rate of the ligand, which is Peptide 3 in our experiments. The cross relaxation observed σ_{ij}^{obs} of protons (i,s) on a ligand in fast exchange with a receptor system is represented by

$$\sigma_{ij}^{obs} = \rho_b \sigma_{ij}^{b} + \rho_f \sigma_{ij}^{f}$$

where ρ_b and ρ_f are mole fractions of bound and free ligand, and σ_{ij}^{b} and σ_{ij}^{f} are the cross relaxation rates of the free and bound ligand (27). A key to the usefulness of the Tr-NOESY effect is that the σ_{ij}^{obs} can be dominated by the bound ligand cross relaxation even when the free ligand is in large excess, $\rho_b \sigma_{ij}^{b} \gg \rho_f \sigma_{ij}^{f}$. In order for this equation to hold

Table 2. ^{1}H resonance assignments of the free acetylated Peptide 3 at 4°C, pH 6.0.

Residue	NH	αH	βH	Other
Acyl 1				
Ile 2	8.36	4.12	1.85	γCH_2 0.91, 1.20
				δCH_3 0.91
Leu 3	8.56	4.40	1.69	γH 1.59
				δCH_3 0.91
Gly 4	8.54	3.96		
Asn 5	8.53	4.73	2.83	
Gln 6	8.70	4.32	2.04, 2.18	γCH_2 2.41
				δNH_2 7.01, 7.66
Gly 7	8.58	3.95		
Ser 8	8.26	4.42	3.80	
Phe 9	8.38	4.66	3.06, 3.20	2,6H 7.27
				3,5H 7.37
Leu 10	8.29	4.39	1.65	γH 1.57
				δCH_3 0.89, 0.90
Thr 11	8.22	4.33	4.22	γCH_3 1.22
Lys 12	8.50	4.39	1.80, 1.90	γCH_2 1.34, 1.45
Gly 13	8.46	4.06, 4.19		
Pro 14		4.48	2.32	γCH_2 1.96
Ser 15	8.67	4.44	3.89	
Lys 16	8.66	4.35	1.79, 1.91	γCH_2 1.48
Leu 17	8.42	4.30		
Asn 18	8.57	4.71	2.77, 2.87	γNH_2 7.05, 7.77
Asp 19	8.37	4.58	2.70	
Arg 20	8.40	4.33	1.82, 1.94	γCH_2 1.66
				δCH_2 3.23
Ala 21	8.51	4.23	1.42	
Asp 22	8.46	4.60	2.73, 2.76	
Ser 23	8.38			
Arg 24	8.30		1.82, 1.91	
Arg 25				

Table 3. ^{1}H resonance assignments of the HIV-1 gp120-bound acetylated Peptide 3 at 4°C, pH 6.0.

Residue	NH	αH	βH	Other
Acyl 1				
Ile 2	8.36	4.12	1.85	γCH_2 0.91, 1.20
				δCH_3 0.91
Leu 3	8.53	4.42	1.70	δCH_3 0.90
Gly 4	8.58	3.96		
Asn 5	8.51	4.73	2.83	
Gln 6	8.68	4.32	2.04, 2.22	γCH_2 2.41
				δNH_2 7.00, 7.64
Gly 7	8.48	3.95		
Ser 8	8.18	4.42	3.80	
Phe 9	8.36	4.65	3.05, 3.19	2,6H 7.26
				3,5H 7.37
Leu 10	8.27	4.38	1:78	γH 1.57
				δCH_3 0.89, 0.90
Thr 11	8.20	4.33	4.23	γCH_3 1.23
Lys 12	8.48	4.39	1.80, 1.90	γCH_2 1.34, 1.43
Gly 13	8.43	4.04, 4.19		
Pro 14		4.48	2.31	γCH_2 1.97
				δCH_2 3.64
Ser 15	8.64		3.89	
Lys 16	8.63	4.37	1.76, 1.90	γCH_2 1.47
Leu 17	8.40			
Asn 18	8.54	4.71	2.77, 2.87	γNH_2 7.03, 7.75
Asp 19	8.35	4.60	2.70	
Arg 20	8.40	4.33	1.82, 1.91	γCH_2 1.67
				δCH_2 3.23
Ala 21	8.49	4.26	1.42	
Asp 22	8.44	4.60	2.74, 2.76	
Ser 23	8.38		3.89	
Arg 24	8.30		1.82, 1.91	
Arg 25				

true, the exchange rate must be short compared to the reciprocal of the fastest cross relaxation rate of the ligand in the bound state ($1/\sigma_{ij}^{b}$) (27). This condition can be expressed in terms of the exchange constant:

$$k_{-1} \gg \max |\sigma_{ij}^{b}|$$

Failure of this fast exchange condition makes the cross relaxation rates from a Tr-NOESY experiment very difficult to interpret.

$T_{1\rho}$ relaxation experiments are potentially useful methods for measuring exchange rates between proteins and ligands (34). Plots of $1/T_{1\rho}$ vs. $\mathrm{Log}_{10}[\omega_{SL}]$, where ω_{SL} is the applied spin locking RF field, can be analyzed for the break point which indicates that exchange is taking place. $T_{1\rho}$ experiments were conducted on our system and a graph (Figure 21) of the data yielded a sigmoidally-shaped curve allowing us to estimate the off-rate at 200 s^{-1}. Specifically, we examined the shift of the β protons of Gln6. This suggests that k_{off} may be fast enough and that the system investigated may be in fast enough exchange on the cross relaxation timescale to validate our approach and treatment of the data. Ideally, we would like $k_{off} \geq 500\ s^{-1}$.

It has been mentioned that spin echo filters were used in the pulse sequence for the Tr-NOESY experiments. Both spin echo and non-spin echo experiments were performed. A spin echo time of 2 ms was used to refocus the peptide proton spins but not those of the protein. Previous experiments performed on the 500 MHz NMR at MSU showed that the ring protons on Phe43 were affected by the use of the spin echo pulses, suggesting that the Phe sidechain is in close contact with protons on the protein and that intimate interactions with gp120 cause spin diffusion. This data supports the idea that the Phe sidechain fits into

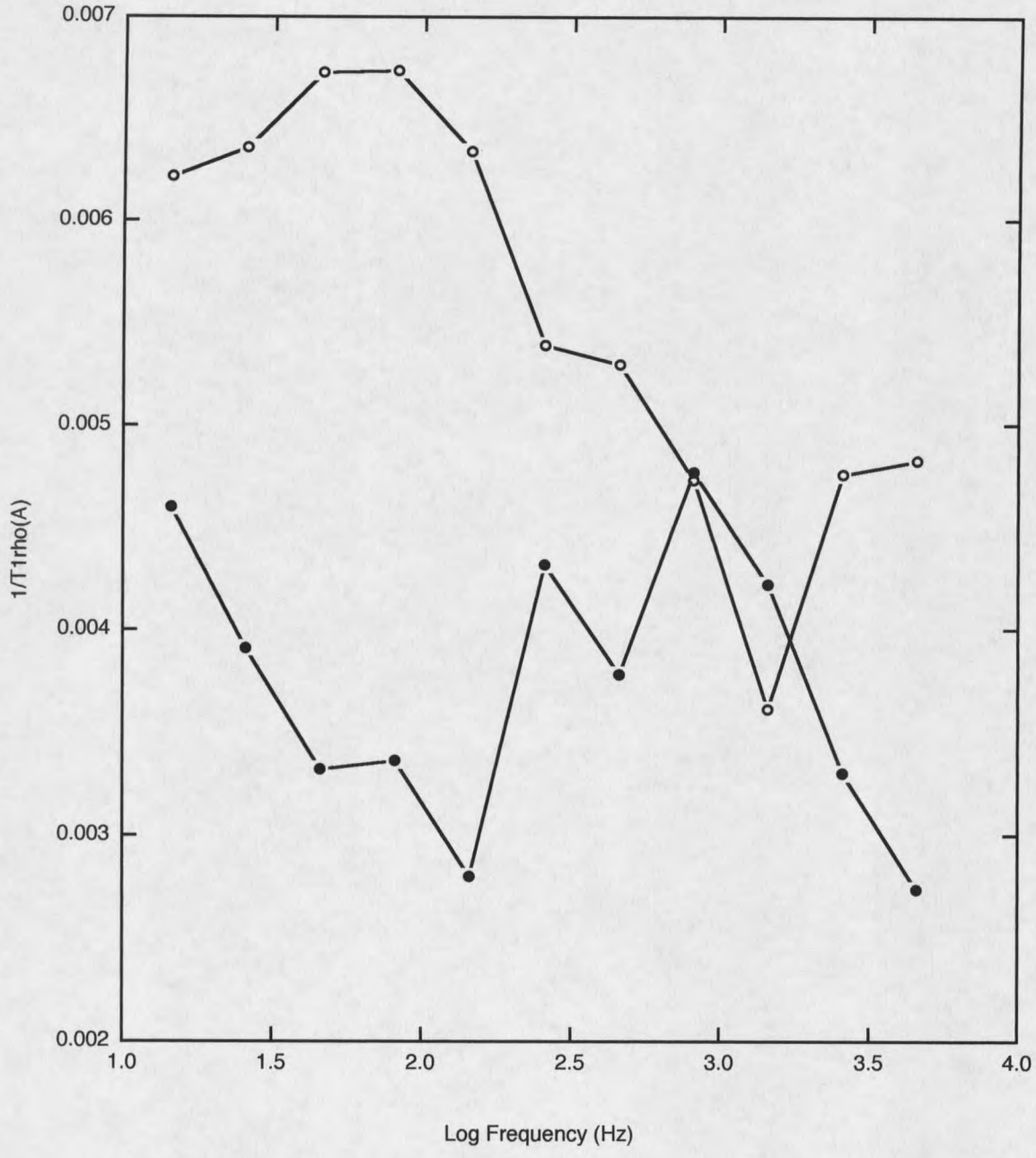


Figure 21. $T_{1\rho}$ area plot of the β protons of Gln6 in the free and gp120-bound Peptide 3. The filled circles represent the free Peptide 3. The data suggests the peptide has an approximate off-rate of 200 s^{-1}. The 90 was 7.7 μs and the power was set at 3 dB. To convert dB to frequency, the following equations were used: y = 32467.5/x (where y is the desired frequency value for the plot) and x = 10^(dB-3dB)/20 (where dB is the value at which data was collected).

a binding pocket on the gp120 molecule. Comparison of spin echo and non-spin echo experiments from the 750 MHz data have been more complex in their analyses due to the introduction of other variables, such as sample degradation. The 750 MHz NMR improved the quality of our data such that we were able to identify twice as many long range NOE's as was seen from the 500 MHz data.

Table 4 shows the long range NOE's which were present in the free and bound NMR experiments.

Table 4. Long range NOE's present in the free and gp120-bound Peptide 3 NMR spectra.

Free	***Bound***
Ile_2:HN *Gln_6:HB1	Ile_2:HN *Gln_6:HB1
*Arg_24:HN *Arg_20:HA	*Arg_24:HN *Arg_20:HA
Leu_3:HN Ser_23:HN	Leu_3:HN Ser_23:HN
*Asp_22:HN Lys_12:HB1	*Asp_22:HN Lys_12:HB1
Present in Free only	***Present in Bound only***
Leu_ 17:HN *Arg_24:HN	Thr_11:HN *Gln_6:HB1
Ser_23:HN Asp_19:HA	Thr_11:HN *Ser_23:QB
Ser_23:HN *Arg_20:HA	Thr_11:HN Pro_14:QD
Leu_3:HN Leu_17:HA	Ile_2:HN *Lys_16:QG
	Lys_16:HN Arg_20:HN

As can be seen from Table 4, five new long range NOE's appeared in the gp120-bound peptide that had not been present in the free; 4 of these were strong (1.8-2.5 Å). Four long range NOE's (2.1-2.6 Å) in the free disappeared completely in the bound. Another 4 were present in both the free and the bound spectra. Use of the spin echo filtering did not change any of these constraints by more than 0.3 Å. These data indicate that Peptide 3 may have some structure when it is free in solution, but it appears to change its conformation significantly upon binding to gp120. Figure 22 shows a comparison of free and bound spectra in regions where cross peaks appear upon binding.

In retrospect and after re-examination of some of the suspect assignments, changes were made to Table 4 and have been denoted by an asterisk (*). The changes which affect the long range NOE assignments include: Arg_24:HN has been changed to Arg_25:HN, Arg_20:HA has been changed to Arg_24:HA, Ser_23:QB has been changed to Ser_15:QB, and Lys_16:QG has been changed to Ile_2:HG1,2. It is also suspected that Gln_6:HB1 may be the N-acetyl group and that the weak cross peak involving Asp_22:HN may be Gly_13:HN. The NH-NH regions of the spectra involved some ambiguous cross peaks due to noise and, as such, the spectra must undergo re-processing.

The addition of sCD4 to the peptide + gp120 sample did not simply cause the spectrum to change back to that of the free peptide without gp120 as might be expected. However, all of the long range NOEs that were present in the gp120-bound peptide, but not in the free peptide, weakened considerably upon addition of sCD4. This data suggests that the sCD4 was at least partially displacing the peptide from the gp120. Additionally, an analysis of the structure of the gp120-bound peptide has been completed and has shown some interesting results.

In place of the ISPA distances generated from the NMR experiments, an extended chain model of Peptide 3 was built in InsightII and used in the MARDIGRAS calculations in determining the structure of the gp120-bound Peptide 3. In order to obtain the structure of

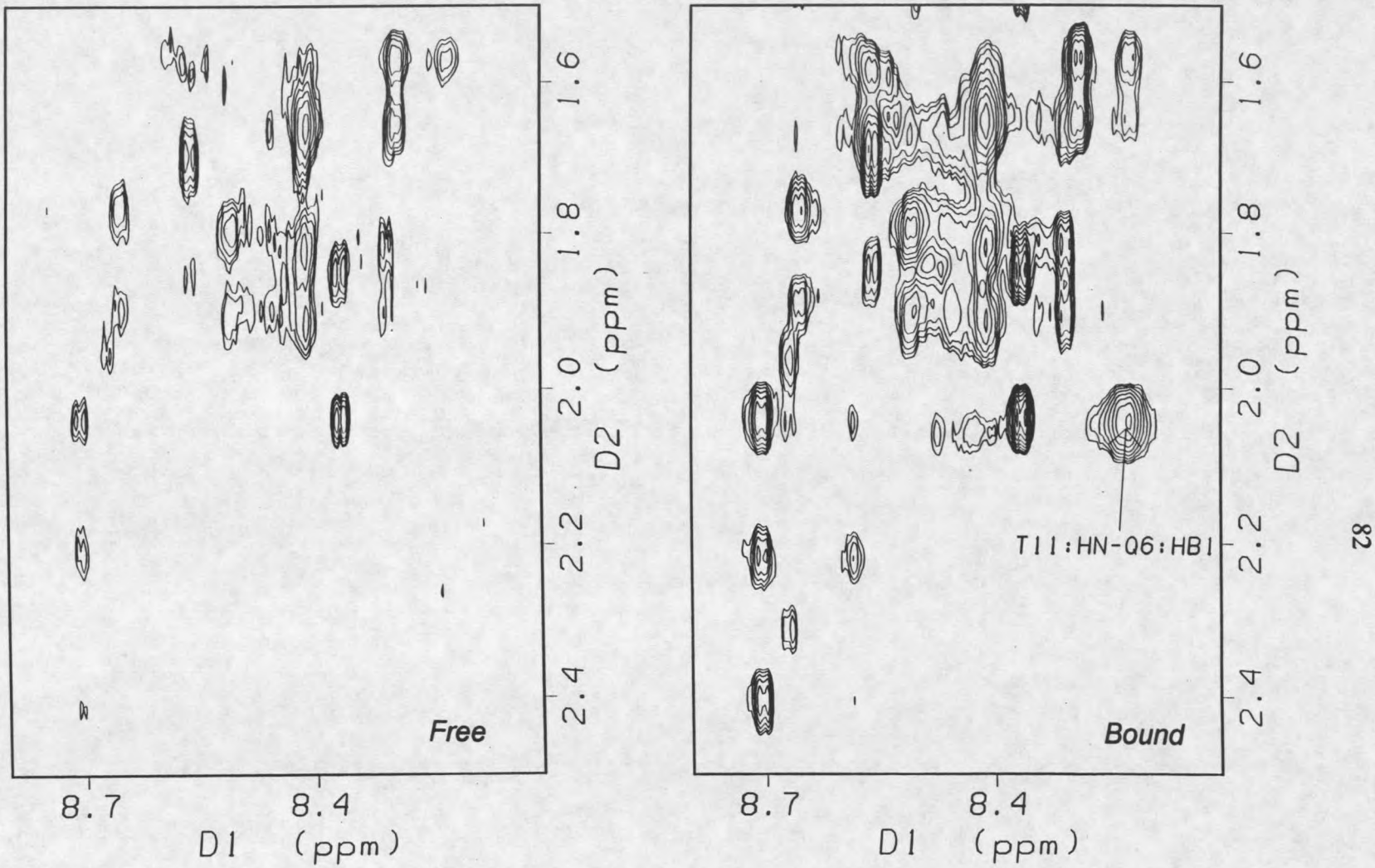


Figure 22. A comparison of gp120-bound and free Peptide 3 NMR spectra. These spectra show the appearance of a new long range cross peak in the NMR spectra upon binding of Peptide 3 to gp120.

the bound peptide, we first subtracted the contribution of the free by subtracting the normalized intensity files. The intensity difference file was converted from a Biosym format to a MARDIGRAS file format using macros written at MSU by Chris Jones and Dawit Gizachew. A constraint file was submitted to MARDIGRAS and 83 of the 122 original constraints were accepted. Some of the constraints are not "useful" to MARDIGRAS, such as geminal proton assignments, so these constraints are thrown out by the program. The number of constraints per residue are represented in Figure 23.

MARDIGRAS requires a correlation time to properly calculate structures. After some initial short runs of MARDIGRAS using various correlation times, it was noticed that the distances for the Gln6 and Asn18 residues were behaving oddly, i.e. they were not converging. These distances were very short and remained so regardless of the distance changes in the rest of the molecule following these rounds of MARDIGRAS. This suggested that the correlation times at those particular residues could be different from the rest of the molecule. To check whether this might be the case, the $T_{1\rho}$ data was examined at high spin lock power. Although this experiment had not been successful at yielding a k_{off} rate, examination of the $T_{1\rho}$ at high power (25dB) for all peaks was a way to compare the relative correlation times for each proton. The data indicated that, indeed, the β hydrogens of Gln6 and Asn18 had shorter and thus faster correlation times than the rest of the peptide (see Table 5).

Fifteen structures were generated after the first round of SA, and all of these fell roughly within the same family. These structures had been generated using 10 sets of randomized coordinates which was thought to be less likely to trap the structure into a local energy minimum than starting with an extended chain conformation. We generated structures using both extended chain and the randomized coordinates and found that while there were some differences, the structures were very much alike. If the extended chain did trap the

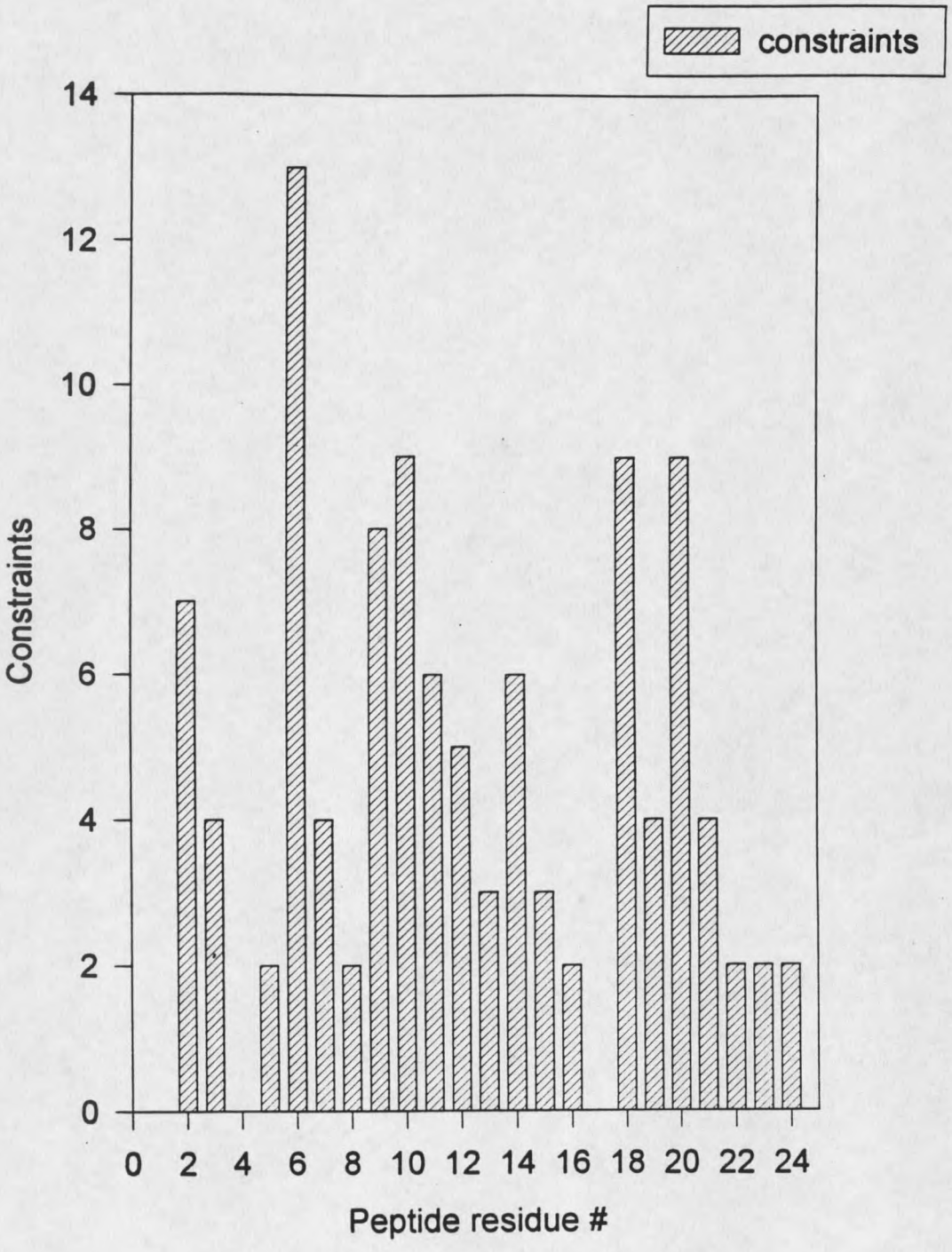


Figure 23. The number of constraints per peptide residue used in the MARDIGRAS calculations.

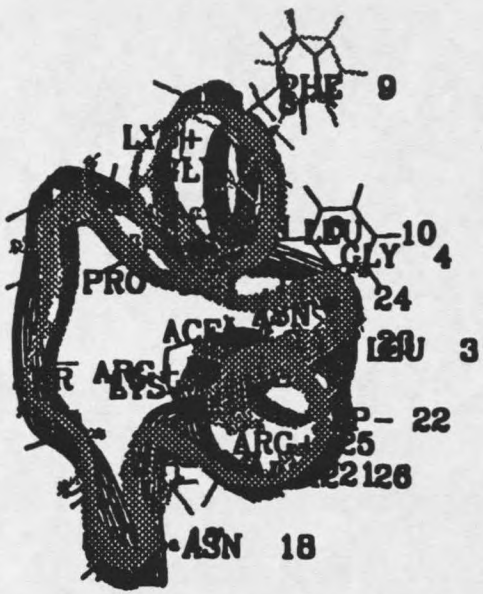
Table 5. R and Q factors in a cross comparison of specified distance constraints after 9 cycles of MARDIGRAS.

τ_e Pep3	τ_e Gln6	R	Q	$R^{1/6}$ x10^{-2}	2-16	2-6	6-11	3-23	11-23	6-6 β1,2	6-6 E1,2	18-18 β1,2
0.45	0.27	0.022	0.011	0.72	2.908	2.297	2.001	3.258	2.371	1.959	2.053	1.737
0.38	0.27	0.022	0.011	0.26	2.906	2.258	1.964	3.258	2.367	1.838	1.996	1.737
0.40	0.27	0.009	0.006	0.13	2.768	2.166	1.866	3.104	2.242	1.803	1.995	1.606
0.50	0.27	0.029	0.014	0.81	3.015	2.320	2.027	3.386	2.461	1.858	1.997	
0.30	0.27	1.177	0.446	5.4	2.255	1.680	1.300	2.539	1.525	1.568	1.995	1.300

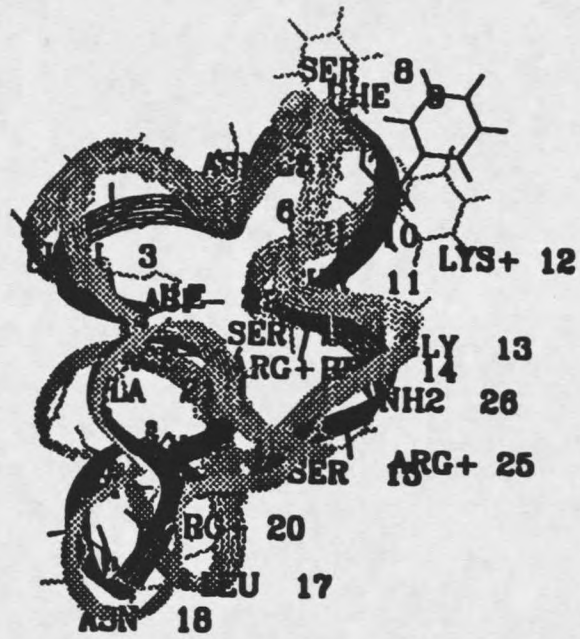
structure into an energy minimum from which it could not escape, it did not have a negative effect on our structures since they were very similar to those obtained via the use of randomized coordinates.

One of the structures from the family of 15 was used again in MARDIGRAS and its distance restraints were analyzed (Table 5). After 9 cycles of MARDIGRAS, it was determined from an evaluation of the R, Q, and $R^{1/6}$ factors that a correlation time of 0.40 ns yielded structures with the best overall agreements between the observed structures and those back-calculated from the models. A round of SA was done starting with the MARDIGRAS input structure and 15 more structures were generated. Of these, 4 showed relatively higher total energy and did not fit geometrically with the other 11. The 11 structures were further refined by MD and energy minimization until each structure converged.

Two families of structures were generated (Figure 24). Family A consisted of structures 2, 7, 8, 11, 12, 13, and 15 and are shown superimposed on structure 12, while Family B consisted of 4, 6, and 14 and are shown superimposed on 14. Structure 5 did not fit into either family. Comparisons of the families with the crystal structure of the corresponding portion of CD4 are shown in Figures 25 and 26. The families both have regions of shared similarity with the crystal structure. According to the root-mean-squared-deviation (RMSD) calculations, both families have the potential to correctly represent the structure of the bound peptide. The RMSD of Family A was 0.0088, while the RMSD of Family B was 0.00939. A plot of $[I_{cal} - I_{obs}]/I_{obs}$ (Figure 27) which is the RMSD of intensities between back-calculated intensities from structure 15 from Family A compared to observed intensities shows an overall good correlation. These back-calculations were done using CORMA. A plot of the discepancies between observed and calculated intensities as a function of residue number is shown in Figure 28 and suggests which

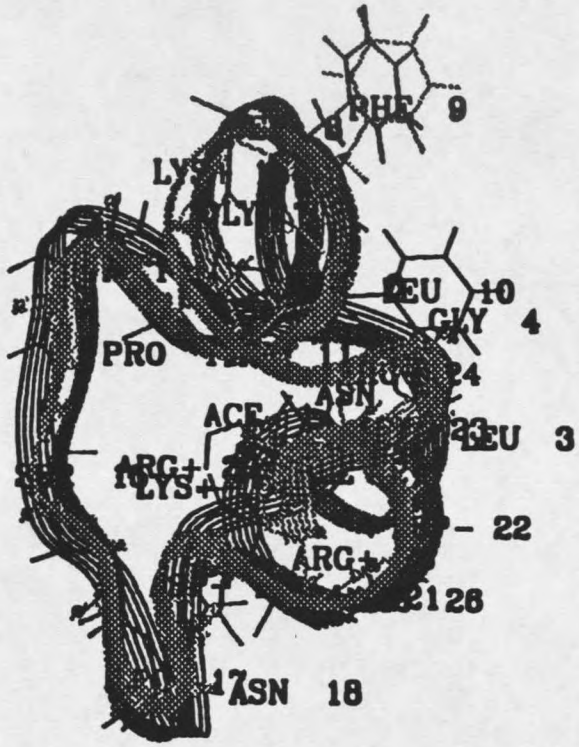


Family A

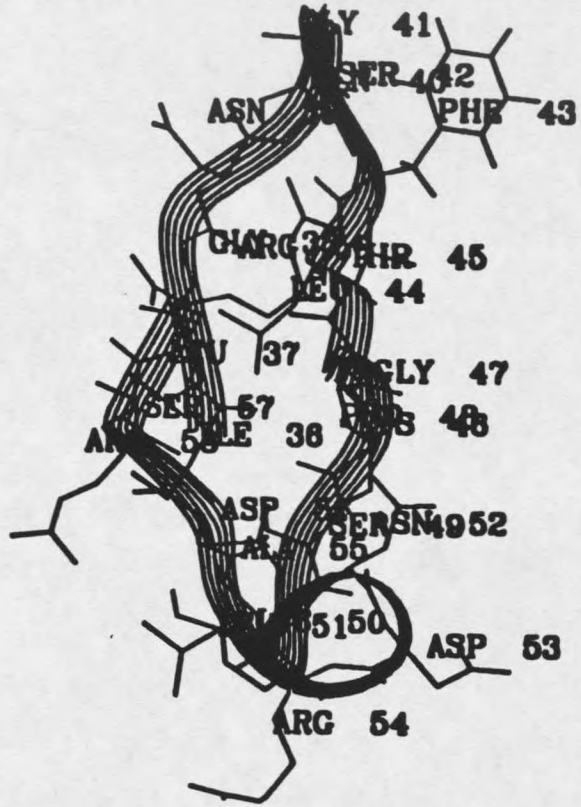


Family B

Figure 24. A comparison of gp120-bound Peptide 3 structure Families A and B generated by simulated annealing using distance constraints generated by MARDIGRAS.



Family A



CD4_36-59_II

Figure 25. A comparison of Family A with the corresponding CD4 crystal structure of Peptide 3.