

AD _____

MIPR NUMBER 92MM2581

TITLE: HIV Protein Sequence/Structure Analysis in Support of
Vaccine Development

PRINCIPAL INVESTIGATOR: Goutam Gupta, Ph.D.

CONTRACTING ORGANIZATION: Department of Energy, Albuquerque
Albuquerque, New Mexico 87185-5400

REPORT DATE: January 1994

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

19970303 156

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE January 1994	3. REPORT TYPE AND DATES COVERED Final (1 Aug 92 - 31 Dec 93)	
4. TITLE AND SUBTITLE HIV Protein Sequence/Structure Analysis in Support of Vaccine Development			5. FUNDING NUMBERS 92MM2581	
6. AUTHOR(S) Goutam Gupta, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Enery, Albuquerque Albuquerque, New Mexico 87185-5400			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200) See page 1 - Summary				
14. SUBJECT TERMS gp120 subdomains, NMR and modeling, a working of model of gp120, interacting functional epitopes			15. NUMBER OF PAGES 177	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

Where copyrighted material is quoted, permission has been obtained to use such material.

Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

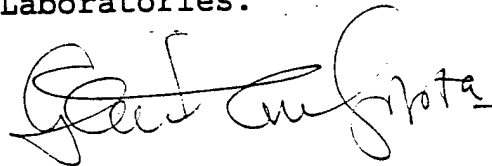
In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



GOUTAM GUPTA
PI - Signature

15 Jan 97
Date

TABLE OF CONTENTS

Summary	1
A. Research Objectives	2-4
B. Background and Rationale	4-10
C. Progress Report	10-31
D. Publications	32
Appendix I	33-39
Appendix II	40-42
References	43-50

SUMMARY

The CD4 molecules on the target macrophage and T cell are the primary receptors for the HIV-1 surface glycoprotein, gp120. In addition, chemokine receptors on the macrophage and T cell serve as co-receptors in the virus-cell interactions. An understanding of the mechanism of virus-cell interactions requires quantitative analyses of the structure-function correlations of the surface epitopes on gp120 which contains several constant (C) and variable (V) subdomains linked as C1-V1-V2-C2-V3-C3-V4-C4-V5-C5. The surface epitope inside the C4 loop is critical for CD4 binding. The epitopes inside the V1-V2 and V3 loops elicit HIV-1 neutralizing response as well as determine tropism, fusion, and infectivity of the virus. In absence of a high resolution structure of the entire gp120, we have adopted an alternative approach to analyzing the structural properties of these surface epitopes. For this purpose, we have combined theoretical and experimental techniques including sequence analysis, molecular modeling, polypeptide engineering, NMR spectroscopy, antibody binding, and neutralization assays. *First*, we have analyzed the sequence-structure-antigenicity correlations of the third variable (V3) loop of gp120 both as a cyclic 35 amino acid long peptide and in the context of the native gp120. *Second*, we have obtained average low-energy structures of various other subdomains of gp120 including the V1-V2 and V4-C4 loops. *Finally*, we have constructed a working model of gp120 based upon the knowledge of (i) the structures of the gp120 subdomains obtained by molecular modeling in conjunction with NMR and other spectroscopic data, (ii) the surface exposure data of various contiguous regions in gp120, and (iii) the data on subdomain-subdomain interactions obtained from replication competency, monoclonal antibody binding, fusion, and infectivity assays. The working model of gp120 adequately describes the sequence-structure-binding properties of the functional epitopes on gp120 such as those inside V1-V2, V3, and C4.

(Key Words: gp120 subdomains/ NMR and modeling/ a working model of gp120)/ interacting functional epitopes)

A. RESEARCH OBJECTIVES

Interaction between the surface glycoprotein, gp120, and the CD4 receptor on the target T cell and macrophage define the first step in HIV-1 pathogenesis [1-12]. gp120 is made up of several well-defined disulfide-bridged constant (C) and variable (V) subdomains or loops [4] linked as C1-V1-V2-C2-V3-C3-V4-C4-V5-C5 (see Figure 1). Receptor and antibody binding experiments [13-19] reveal that a discontinuous epitope formed by residues in C2, C3, and C4 define the contact interface for CD4 binding; the region inside C4 is most critical for binding. Although, it is not directly involved in CD4 binding, the cyclic 35 amino acid (aa) long V3 loop of gp120 contains neutralizing epitopes, i.e., monoclonal antibodies (mAb) directed against sequences inside the V3 loop can neutralize HIV-1 [21-32]. Therefore, it appears that the V3 loop is involved in the post-CD4 binding phase of viral pathogenesis. In fact, it has been reported that the V3 loop is involved in HIV-1 tropism, cell-virus fusion, and replication competency of the virus [52-56, 60-65]. However, the effective use of the V3 loop as a neutralizing target has been complicated by its extensive sequence variation across HIV-1 isolates.

(Research Objective 1) During submission of this project, our primary research objective was to theoretically predict and to test by two-dimensional nuclear magnetic resonance (2D NMR) spectroscopy how the sequence variation correlates with the global structure of the V3 loop and the local structure of the neutralizing determinant (ND) inside the V3 loop.

Since the submission of this proposal in 1991, the following major advances have been made in defining the surface epitopes of gp120 involved in HIV-1 pathogenesis.

- (i) In addition to the V3 loop, the V1-V2 loop has also been shown to contain neutralizing epitopes [33-39] and it is also implicated in cell tropism, viral fusion, and replication competency [57-59].
- (ii) In addition to the CD4 molecule (the primary receptor), the CC or CXC chemokine receptor (CCR or CXCR) on the target T cell and macrophage acts as a co-receptor during viral fusion [5-12]. There is uncertainty about the exact site on gp120 that is responsible for CCR (or CXCR) binding. Studies with different gp120 chimeras indicate that the V3 loop (and not the V1-V2 loop) is important for binding of gp120 to CCR on macrophages [10]. However, since the 35 aa long V3 loop peptide in linear or cyclic form shows no binding to CCR or CXCR [20], it implies that the V3 loop should be presented in the context of the native gp120 such that its local structure and interactions with other loops are preserved.
- (iii) Indeed, for quite some time it has been recognized that different loops of gp120, that are distant in sequence, can functionally (or spatially) interact with each other in the native protein [40-50]. Of particular importance are the interactions involving V3 and C4. This interactions may, in fact, be critical in determining HIV tropism, mAb binding, and viral fusion.

These research advances promoted us to expand our research objectives.

domain	well exposed	poorly exposed	buried or shielded	no information
C1	31-54	85-95	110-120	55-61
	62-71	100-110		71-85
V1	140-150			120-140
				150-158
V2	169-182	158-171		182-195
C2			210-231	195-210
				231-250
		275-280	250-280	280-294
V3	304-322	294-301	320-330	
C3			360-393	330-360
V4		391-405		405-423
C4	423-440		438-451	450-461
V5		461-470		
C5	499-511		465-498	

A schematic representation of HIV-1 gp120

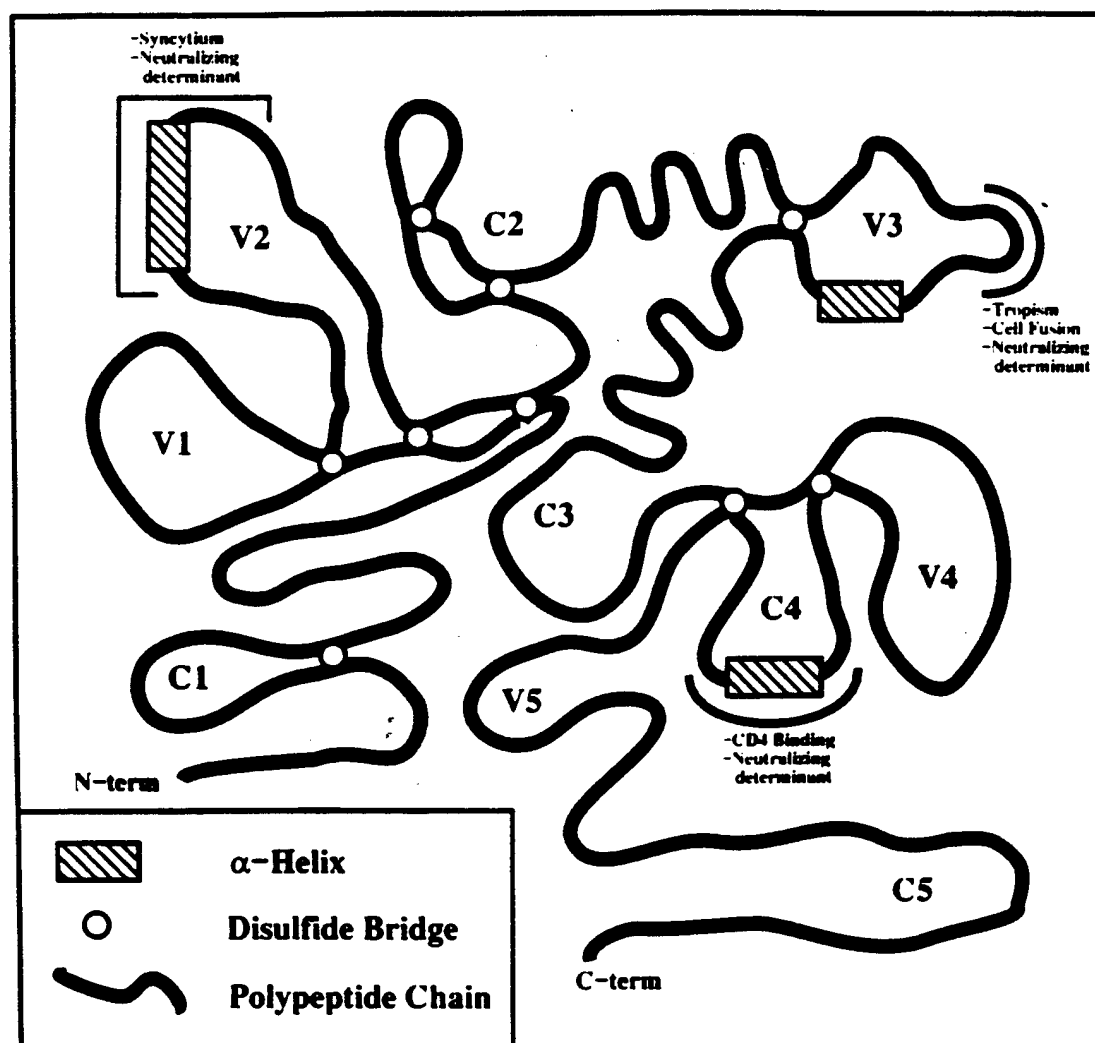


Figure 1. Constant (C) and variable (V) loops in gp120. Surface exposures of various contiguous aa segments in gp120 are also included [data taken from ref. 69].

In absence of a high resolution structure of gp120 by X-ray crystallography, the currently available data in the literature fall short of explaining the structural basis for epitope recognition either by receptors or by mAbs. Therefore, we have attempted to develop and apply an alternative method for obtaining a working model of gp120 that accurately defines the structural properties of various functional epitopes. Characterizations of these epitopes in terms of their sequence-structure-binding correlations will help us better understand the pathogenesis of HIV-1.

(Research Objective 2). To determine the structure-antigenicity correlations of the V3 loop both as a cyclic peptide and in the context of the native gp120. For this purpose, we have combined **(molecular modeling and 2D NMR spectroscopy)** with **(mAb binding and neutralization studies)**.

(Research Objective 3). To obtain a working model of gp120 that accurately defines the local structures of the individual loops and inter-loop interactions. For this purpose, we have first determined the models of V3, V1-V2, and V4-C4 subdomains of gp120. We have then used a simulated annealing method to assemble these subdomains by mainly utilizing the flexibility of the linker regions. The final set of working models has been obtained by screening the sampled structures against (i) the surface exposure data of different contiguous regions on gp120 from the immunochemical maps [21-39, 69] and (ii) the inter-domain interaction data from replication competency, fusion, and infectivity assays [40-50].

Successful completion of these research objectives enables us to obtain a comprehensive knowledge of the sequence-structure-binding correlations of various surface epitopes of gp120 that are involved either in viral pathogenesis or in eliciting neutralizing immune response. This knowledge can be utilized in designing antigens for directing immunity against HIV-1. For example, we have been able to design a multivalent HIV-1 antigen in which the conserved structural element of the V3 loop is multiply presented.

B. BACKGROUND AND RATIONALE

In the literature, three types of information are available about the disulfide-bridged subdomains (or loops) of gp120: (i) functional roles of these loops in viral replication, fusion, infectivity, and cell tropism [52-65], (ii) surface exposures of these loops [21-39, 69], and (iii) inter-loop interactions [40-50]. Very recently it has been shown that gp120 also contains epitopes for co-receptors present on CD4+ T cells and macrophages [5-12].

(i) Functional Roles

The V3 loop of gp120 has been extensively studied. Sequence determinants for HIV-1 tropism have been located inside the V3 loop [52-56]. It has also been reported that the net positive

Table 1

V1-V2 domain

HIVMN: | V1 loop (31) | | V2 loop (38) |
 CVTLNCTDLR---NTTNTN---NSTAN---NNSNS---EGTIKG-GEIKNCSEFNITTSIRDQKQYALLYKLDIVSID-----NDS---TSYRLISCNTSVITQAC

HIVRF: | V1 loop (27) | | V2 loop (51) |
 CVTLNCTDANL-----NCT-----NVTSSSG---GTMEN-GEIKNCSEFQVITTSRRDKTQKKYALFYKLDVVPKRG---NISPK-NNTS-NNT---SYGNYTLHCNNSSVITQAC

HIVSF2: | V1 loop (24) | | V2 loop (43) |
 CVTLNCTDLG-----KATNT---NSSNWK-----EEIK-GEIKNCSEFNITTSIRDQKQYALLYKLDIVSID-----NASTTT---NYT---NYRLHCNRSVITQAC

HIVJRFL: | V1 loop (23) | | V2 loop (38) |
 CVTLNCTDQV-----NAT-----NTTND---EGTMR-GEIKNCSEFNITTSIRDQKQYALLYKLDVVPIDN-----NNT---SYRLISCNTSVITQAC

HIVHAN: | V1 loop (31) | | V2 loop (45) |
 CVTLNCTDLR---NTTNTN---NSTAN---NNSNS---EGTIKG-GEIKNCSEFNITTSIRDQKQYALLYKLDIVSID-----NDS---TSYRLISCNTSVITQAC

HIVWMJ22: | V1 loop (21) | | V2 loop (40) |
 CVTLNCTDK-----NITDNK---NTTI-----IGG-GEIKNCSEFNITTSERDKQYALLYKLDVVPKRG---N---NNS---RYRLHCNNTSVITQAC

HIVZ321: | V1 loop (22) | | V2 loop (44) |
 CVTLSC-H-----NITIKD---NNTVD-----TENK-GEIKNCSEFNITTSERDKQYALLYKLDVVPKRG---N---NNSGDSKYALINNTSAITQAC

HIVMAL: | V1 loop (31) | | V2 loop (38) |
 CVTLNCTDV-----NCTAV---NCTHAGS---NRTNA-----ELKMEI-GEIKNCSEFNITTVGSDKR-QEYATFYKLDVQIDSD-----NNS---RYRLHCNNTSVITQAC

V4-C4 domain

HIVMN: | V4 loop (29) | | CD4 bind (26) |
 CGGEFFYCNTSPLFNSTWNG-----NNT---WNNTT-----GSNNNITLQCKIKQIINMWQEVGK-AMYAPPIEGQIRC

HIVRF: | V4 loop (28) | | CD4 bind (26) |
 CGGEFFYCNTTQLFNSTW-----NST---EGSNNT-----GGNDTITLPCRIKQIINMWQEVGK-AMYAPPISGQIKC

HIVSF2: | V4 loop (27) | | CD4 bind (26) |
 CRGEFFYCNTTQLFNNTWRL-----NHT---EG-TKG-----NDTIILPCRIKQIINMWQEVGK-AMYAPPIGGQISC

HIVJRFL: | V4 loop (28) | | CD4 bind (26) |
 CGGEFFYCNTSPLFNSTWN-----NNT---EGSNNT-----EGN-TITLPCRIKQIINMWQEVGK-AMYAPPIRGQIRC

HIVHAN: | V4 loop (27) | | CD4 bind (26) |
 CGGEFFYCNTSKLFNSTW-----NNT---ST---WWDNG-----NDTIILPCRIKQIINMWQEVGK-AMYAPPIGGLIRC

HIVWMJ22: | V4 loop (28) | | CD4 bind (26) |
 CGGEFFYCNTSPLFNSTW-----NGTDIKGDKNST-----LITLPCRIKQIINMWQEVGK-AMYAPPIQGQIRC

HIVZ321: | V4 loop (31) | | CD4 bind (26) |
 CGGEFFYCNTSGLFNSTWL-----NGT---SNNT---WKIDTVNDTIILPCRIKQIINMWQEVGQ-AMYAPPKGVK

HIVMAL: | V4 loop (30) | | CD4 bind (26) |
 CRGEFFYCNTSKLFNSTWQ-----NNGA---RLSNST-----ESTGSITLPCRIKQIINMWQKTKG-AMYAPPAGVINC

Table 2

Summary of HIV-1 GP120 C4 epitopes

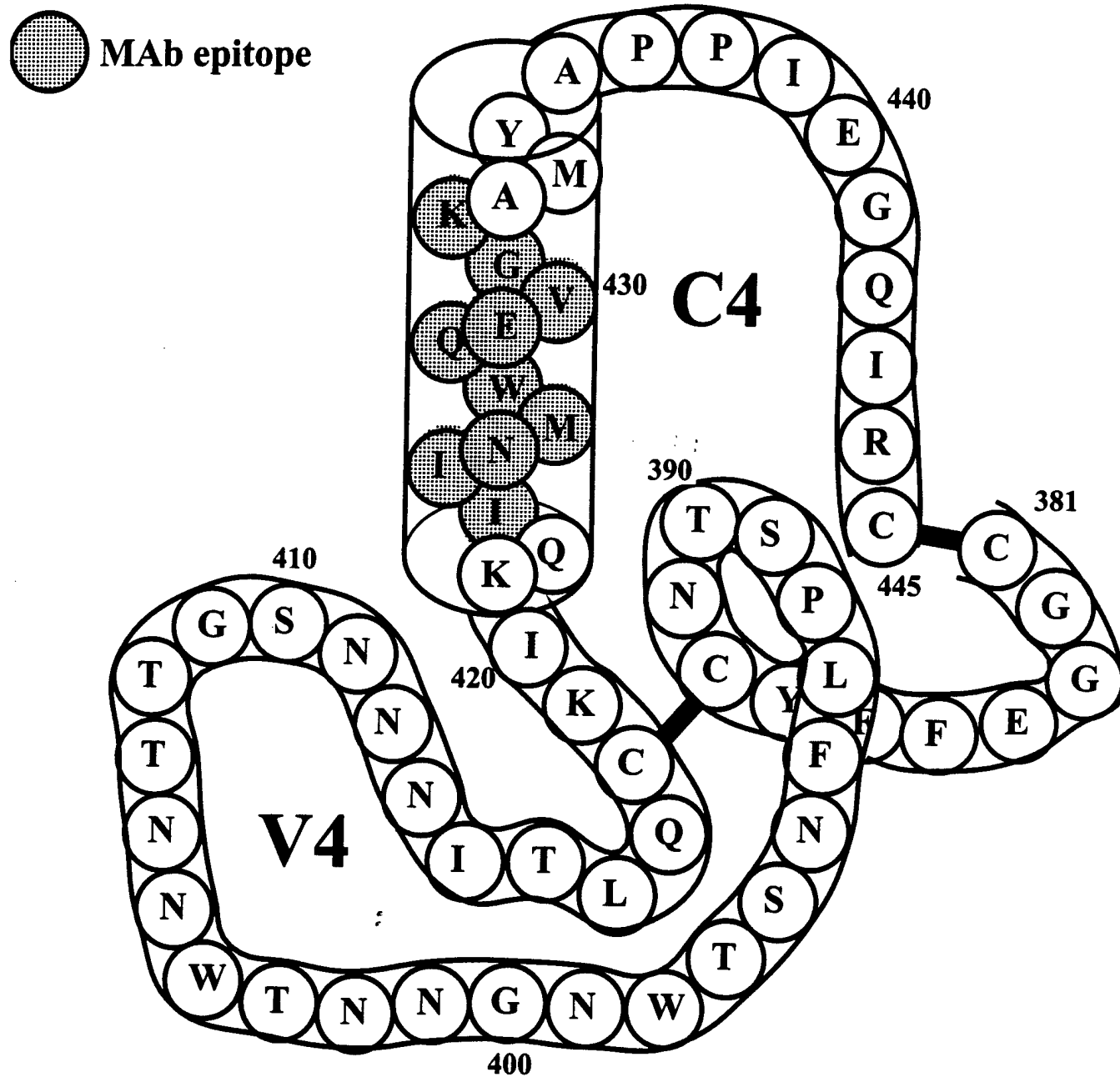


Figure 3. CD4-blocking mAbs raised against MN-gp120 [15] are directed against the C4 helix (shaded) of V4-C4. Secondary structural elements are as predicted by our modeling method (see later in Figure 11B): helix=cylinder, beta strand=arrow.

charge on the V3 loop is a possible indicator for syncytium inducing (SI) ability of a HIV-1 strain, i.e., SI strains have high net positive charge on their V3 loops whereas the non-SI strains have low net positive charges on their V3 loops [60-65]. However, it has also been argued that HIV-1 tropism and SI ability may not only be due to the V3 loop (and its charge) but also be due to its interactions with other loops (C4 in particular) of gp120. In addition to the V3 loop, the V1-V2 loop has also been implicated in HIV-1 tropism. Interestingly, the level of glycosylation, the net charge, the length of the V2 loop seem to contribute to the observed tropism [57-59]. Table 1 shows sequences of various V1-V2 loops to document variations in glycosylation, the net charge, and the length of the V2 loop. Table 2 shows sequences of various V4-C4 loops; note that although V4 shows extensive sequence variation, C4 is fairly conserved especially in the putative helical segment.

Also, both V3 and V1-V2 loops have been shown to play important roles in viral replication, virus-cell fusion, viral infectivity [57-65]. Single site mutations at the highly conserved GPGR-crest drastically reduce the virus-cell fusion [62]. Also elimination of the disulfide bridge between the conserved 1st and 35th C's in the 35 aa long V3 loop completely disables the cleavage of gp160 into gp120 and gp41, which is a pre-requisite for virus-cell fusion. Similarly, single and double site mutations inside the V1-V2 loop alter the SI ability of the mutated HIV-1 strains [59]. It has also been proposed that a putative proteolytic site inside the V3 loop may determine viral fusion to T cell which bears a membrane-bound protease [66-68].

(ii) Surface Exposures

Antibody binding data are useful for determining the surface exposure of various epitopes on gp120. Specificity of a given epitope on gp120 for a given mAb implies that the epitope in question is either permanently or transiently exposed. Mainly two classes of murine and human anti-gp120-mAbs are reported in literature: (a) mAbs that recognize epitopes on the variable V3 and V1-V2 loops and (b) mAbs that recognize epitopes on the constant C2, C3, and C4 loops. The mAbs specific for the variable regions of gp120 neutralize HIV-1 in a type-specific manner, whereas the mAbs specific for the constant regions of gp120 show broadly cross-reactive neutralizing activity [16]. However, the latter are quite difficult to raise in animals and mice probably due to epitope masking.

The majority of the V3-specific mAbs bind to the crest of the loop that contains the GPGR β turn [21-32]. For some of these mAbs, the epitopes include the N-terminal sequence flanking the GPGR, whereas for some other mAbs the epitopes include the C-terminal sequence flanking the GPGR; for a few mAbs (e.g., mAb 9284), the epitopes include both the N and C-terminal sequences flanking the GPGR. A new type of V3-specific mAb has also been identified; this mAb recognizes the sequence N-terminal to the 1st C involved in the disulfide bridge only when gp120 is denatured or in presence of a mAb that is specific for the CD4-binding region of gp120 [50]. This means that the N-terminal sequence in the V3 loop is masked in the native gp120 and it is unmasked in presence of denaturing agents or mAbs specific for the CD4-binding region.

Both murine and human V1-V2 specific mAbs have been isolated [33-39]. These mAbs are specific either for linear or conformational epitopes inside the V1-V2 loop (see Figure 2). Single and double site mutations on gp120 that diminish or enhance the mAb binding have also been reported. These mutations generally lie inside the V1-V2 loop for mAbs specific for linear epitopes whereas they lie both inside and outside the V1-V2 loop for mAbs specific for conformational epitopes [36].

Binding studies reveal that a discontinuous epitope formed by C2, C3, and C4 regions on gp120 determine the specificity for CD4 binding. CD4-blocking mAbs compete for the sites on C2, C3, and C4 (in particular) for binding (see Figure 3). CD4-blocking mAbs are generally cross-reactive across HIV-1 isolates [15-19].

Moore and co-workers have identified surface exposures of various sites on gp120 based upon a detailed analyses of mAb and CD4 binding data [69]. These analyses allow classification of different regions of gp120 as well-exposed, partially exposed, and completely buried, although the surface exposures of several other regions of gp120 remain undetermined (see Figure 1). The surface exposure of an aa segment in a gp120 model can be determined by computing the accessible surface areas of the residues (X) in the segment relative to those in the extended GXG tripeptide. In a well-exposed segment, accessible surface areas of the residues (X) should be larger than those in GXG. Similarly in a buried segment, accessible surface areas of the residues (X) should be lower than those in GXG.

(ii) Inter-loop Interactions

The binding data not only provide information about local structures and surface exposures of various loops but also information about long-range loop-loop interactions. In a continuous conformational epitope, the structure of the epitope is stabilized not only by its amino acid sequence and local disulfide bridge but also by its interactions with other regions distant in sequence; the conformational epitope in V2 is one such example [36]. Long-range loop-loop interactions are also relevant for the discontinuous epitope in which regions distant in sequence come close in space to create the contact interface for mAb binding; the discontinuous epitope for the CD4-blocking mAb, 1125H, is one such example [13]. In addition to the mAb binding data, replication competency, fusion, and infectivity assays reveal functional (and perhaps spatial) interactions between pairs of loops in gp120 [40-51], e.g., (V3 and V1-V2), (V3 and C4), (V1 and C4), etc. It is important to examine whether the inter-loop interactions from the functional assay indeed correspond to spatial interactions in our gp120 model. For example, the N-terminal V3 sequence appears to be functionally correlated with the putative helical segment in C4. We have, therefore, examined the spatial proximity of these two fragments in our gp120 model. Similarly since W427 in C4 functionally interacts with residues in V1, we have examined the spatial proximity of W427 with other residues and in particular the residues inside V1. In fact, we have tested spatial interactions for residue-pairs involved in long-range functional interactions.

In this project, special attention is paid to deciphering the role of the variable loops of gp120 in surface recognition. We believe that the interplay between the constant and variable regions of gp120 is critical in determining tropism, fusion, and infectivity although a truncated gp120 with only the

constant regions (i.e., without V1-V2 and V3) can produce high affinity binding to CD4 [73]. Previously, linear and cyclic peptides with V3 epitopes were used as immunogens [74]. Recently, a bispecific linear peptide containing the V3 and C4 epitopes [75] and peptamer containing several putative C4 helical segments [76] have also been tried as immunogens. While the potential use of these constructs as HIV vaccines remains to be seen, the subdomains chosen in our study will have equal (if not better) promise as immunogens since they closely mimic the native gp120 (the natural immunogen) due to the presence of disulfide bridges and inter-loop interactions.

(iv) Surface Epitopes on gp120 for CCR OR CXCR (or co-receptor) Binding

Recent studies have shown that the binding of gp120 to CD4 on T cells or macrophages is not enough for virus-cell fusion. Additional factors (or co-receptors) are needed to complete the gp120-mediated membrane fusion that is a prerequisite for HIV-1 infection. On macrophages these receptors are CCR5 or CCR3 [7-8, 10] and on T cells they are fusins or CXCR [9]. Both CCR and fusin are G-protein coupled receptors with seven transmembrane spanning helix segments [7-12]. Use of various gp120 chimera suggests that the V3 loop (and not the V1-V2 loop) is probably involved in the binding of gp120 to CCR. Similarly, the V3 loop may also be involved in the binding of gp120 to fusin since the V3-specific mAb, D47, blocks HIV-1 fusion to T cells bearing CD4 and fusin. However, attempts in various laboratories have failed to demonstrate that the 35 aa long V3 loop peptide itself binds to CCR or fusin (or CXCR). Since it shows characteristic difference in sequence between macrophage and T cell, the V3 loop appears to be the logical target for CCR and fusin, i.e., V3-specific recognition of gp120 by CCR and fusin is expected to guide the HIV-1 fusion to macrophage and T cell, respectively. Various studies [56] indicate that the structure of the V3 loop and its interactions with other loops (rather than the V3 loop alone) determine HIV-1 tropism. In a similar manner, the structure of the V3 loop and its interactions with other loops (rather than the V3 loop alone) may also determine the binding of gp120 to CCR or fusin.

C. PROGRESS REPORT

We have studied (in collaboration with the NCI/NIH and CDC) the structure-antigenicity correlations of the V3 loops as isolated cyclic peptides and also in the context of the native gp120. For this work, we have combined various theoretical and experimental tools including sequence analyses, molecular modeling, NMR spectroscopy, mAb binding by ELISA and BIAcore, syncytium, and neutralization assays [77-87]. The work on the V3 loop has resulted in the identification of a conserved structural element at the crest of the V3 loop and subsequently we have designed multivalent V3-specific HIV-1 antigens in which this conserved structural element has been multiply expressed. However, we have not only restricted this project to the V3 loop but also included studies on the V1-V2 and V4-C4 loops of gp120. We have also utilized the circular dichroism (CD) data that show the presence of a

putative α -helix in the V1-V2 and V4-C4 loops and explained how these data are utilized in obtaining molecular models of V1-V2 and V4-C4. Finally, we have developed and applied a molecular modeling method for assembling different subdomains of gp120 in a native fold.

C.1 Cyclic V3 Loops: Sequence-Structure-Antigenicity Correlations

C.1a Structural Requirement for Antigen-Antibody Interactions: A Case Study of the HIV-MN V3 Antigen

Theoretical studies [77-78] revealed that the variability in sequence and structure of the V3 loop is confined to the N- and C-terminal sides of the conserved GPG-crest. This leaves three regions of the V3 loop conserved both in sequence and secondary structure. Figure 4 shows the V3 loop sequences of various HIV-1 isolates; the three conserved secondary structural elements are underlined.

Figure 4. Cyclic V3 and Mini-V3 Loops

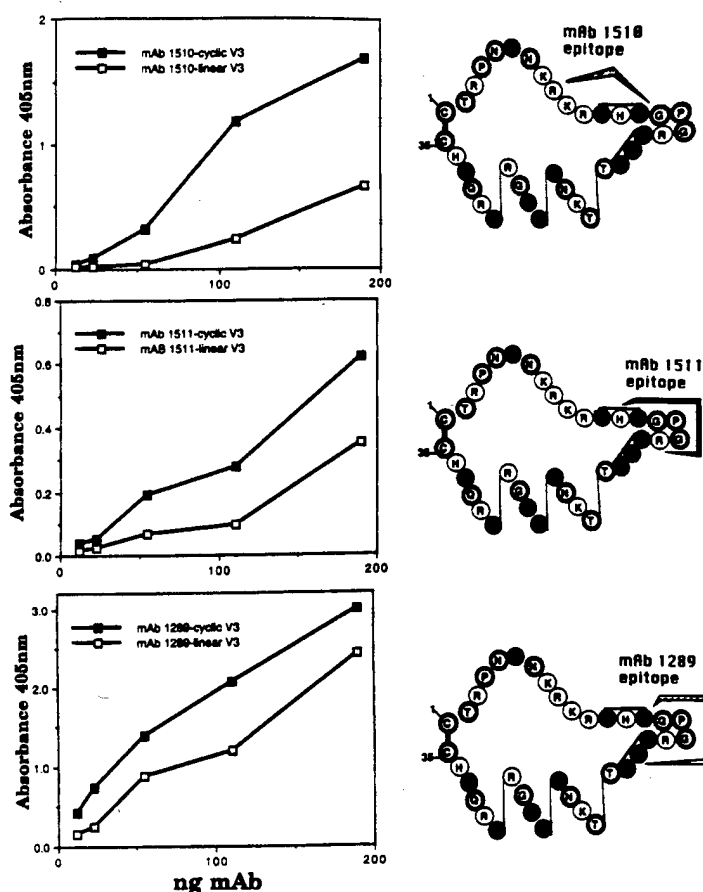
	turns	turn	helix
V3-MN	:CTRP	<u>NYNKRKRI</u>	<u>HIGPGRAFYTTKNIIGTIROAHC</u>
Mini V3	:	<u>CR</u>	<u>IHIGPGRAFYTTKC</u>
V3-RF	:-----N-T--S-TK----	VI-A-GQ---	D--K---
V3-Florida	:-----YT--G-R-----	V-AAEK---	D--R---
Mini V3	:	CG-R-----	V-AAEC
V3-Haiti	:-----D-T--S-PM---	K---	A-GD---N-----
V3-Thailand	:----SN-T-TS-T----	QV--R-GD---	D--K-Y-
Mini V3	:	CS-T----	QV--R-GC

We have carried out NMR studies [79-80, 83] to test the validity of our theoretical predictions. Structural studies were performed for the HIV-MN V3 loop in the linear and cyclic (S-S bridged) forms. While the linear V3 loop in water is devoid of ordered structure except for a loose turn at the GPG-crest, the cyclic form shows a well defined structure in water. Moreover, in (7:3) water:TFE mixed solvent (less polar than water), the cyclic V3 loop shows higher order both in terms of secondary structure content and rigidity. The three conserved regions of the HIV-MN V3 loop in the mixed solvent adopt the predicted secondary structural elements. TFE-induced helix formation in the C-terminal segment is also documented by us in the Haitian V3 loop. TFE induced, less polar environment for helix

stabilization is biologically relevant in the context of the native structure of gp120. As discussed in the following section, our modeling studies show that the proximity of the C3 region provides a hydrophobic environment for the C-terminal segment of the V3 loop (see Figure 10).

The observation of the three conserved secondary structures in the V3 loop leads to a simple rule that the sequence variability of the V3 loop can be tracked by finding the associated variability brought about by different structural elements on either side of the GPG-crest. Finally, the conformational requirement of the neutralizing determinant (ND) in the V3 loop-antibody interaction is tested by monitoring the mAb binding to the HIV-MN V3 loop in the linear and cyclic forms by ELISA [79]. The cyclization through the (S-S)-bridge between C1 and C35 and changing the solvent environment provide interesting insights into the structure-binding correlation of the HIV V3 loop. Binding of linear and cyclic V3-MN loops to three different monoclonal antibodies are compared in Figure 5. Antibodies 1510, 1511, and 1289 bind to the V3 epitopes KRIHI, HIGPGR, and GPGRF, respectively. Note that the cyclic V3-MN loop is a better ligand than the linear analog in all three cases. This is consistent with the experimental evidence that the cyclic V3-MN loop is more structured than the linear analog. As expected, the most pronounced difference in binding occurs for the mAb 1510 which recognizes the sequence KRIHI on the N-terminal side of the GPG-crest; this sequence also shows more ordered structure upon cyclization. For the other two antibodies, the difference in binding is smaller because both of them include the GPGR which even in the linear analog shows a residual turn.

Figure 5. An ELISA showing the preference of monoclonal antibodies for the cyclic over the linear form of the HIV-MN V3 loop. Human mAbs 1510 (Top) and 1511 (Center) and murine antibody 1289 (Bottom) all bind to a greater extent to the cyclic V3 loop peptide. The recognized epitopes 1510 (Top), 1511 (Center), and 1289 (bottom) are shown on the right. In the schematic representations of the HIV-MN V3 loop shown on the right, solid circles depict hydrophobic residues, open circles charged residues, and outlined circles polar uncharged residues. BIAcore measurements [87] reconfirm the ELISA data.



Therefore, the NMR and antibody binding studies imply that vaccine attempts using the cyclic V3 loop would be more effective than the linear analog in inducing protective humoral immunity to the conserved structural features. The binding profiles of mAb 1510 and 1511 (both derived from AIDS infected patients) reinforce the notion that the cyclic V3 loop presents the epitope structures similar to that found in native gp120. Interestingly, the (K10-R11-I12-H13-14-G15-P16-G17) fragment which is a part of the neutralizing epitope of the cyclic MN V3 loop shows the same structure in water and in the mixed solvent as in the co-crystal of the neutralizing antibody (mAb 50.1) and the MN V3 loop peptide antigen complex [88].

C.1b Conserved Structure at the Immunogenic Tip of the V3 Loop: Design of a Chimeric Multivalent HIV Antigen that Contains Multiple Copies of this Conserved Structural Element

We carried out molecular modeling and two-dimensional (2D) NMR studies on the V3 loops sequences shown in Figure 4 to identify the structural features of the V3 loop, especially at the ND, that remain conserved irrespective of the sequence variation. The conserved structure of the ND is a solvent accessible protruding motif or a knob (Figure 6). Interestingly, we also showed (Figure 6) by 2D NMR spectroscopy [81-82] that the HIV ND knobs are structurally isomorphous with the immunodominant knobs in the tandem repeat protein, human mucin Muc-1 (a tumor antigen for breast, pancreatic, and ovarian cancer). Each 20 amino acid repeat of Muc-1 consists of (TSAPDTRAPGSTAPPAHGV). The antigenic knob of Muc-1 is located at APDTR. Therefore, we replaced the mucin antigenic knobs by the HIV ND knobs in a set of chimeric Muc-1/V3 antigens. The repeat sequences of the chimera are: (TSGPGRAFAPGSTAPPAHGV)_n, (IHIGPGRAFAPGSTAPPAHGV)_n, and (HIGPGRAPAPGSTAPPAHGV)_n. The V3 inserts are underlined. This produced multivalent HIV antigens in which NDs are located at regular intervals and separated by extended mucin spacers. We have shown by 2D NMR spectroscopy that the multivalent antigens preserve the NDs in their native structure. We have also demonstrated by enzyme-linked-ELISA that the antigens correctly present the NDs to produce binding with monoclonal antibodies (mAbs) and polyclonal antisera from AIDS infected patients. The antibody binding of these chimera is equivalent to the cyclic form MN V3 loop.

Muc-1/V3 antigens are unique in the following ways. (i) NMR and antibody binding data [81] verify that they reproduce the native structure of the NDs even when they are presented in the context of a totally unrelated protein like mucin Muc-1. (ii) Immunogens containing identical NDs within the Muc-1 chimeras effectively allows enhanced presentation of a conserved structural feature of the virus in a fashion not possible with non-chimeric HIV antigens. The true advantage of this approach will be to induce either T-dependent or T-independent antibody responses to the ND depending on the precise construction of the antigen. (iii) Multiple NDs, present in these chimeric proteins, may be advantageous in enhancing the immune response by significantly increasing the affinity of antibody binding. The importance of multiple NDs being present in the same antigen becomes clear by analyzing the relative

binding of Muc-1/V3 (HIGPGRAPAGSTAPPAHGV)3,6 peptides to different antisera. The data show that the 120 residue peptide is a better ligand than the 60 residue peptide for the majority of the antisera we tested. This is probably due to the fact that the higher number of ND knobs in the 120 residue peptides are correctly disposed along the long axis of the molecule to facilitate the binding of bivalent antibodies. (iv) Alternatively, the nature of the Muc-1/V3 structure (Figure 6) suggests that if two or more different NDs are grafted alternately along the chain, there is enough flexibility in the spacers such that two or more antibodies specific for two different NDs can both bind bivalently, interdigitating along the molecule. Finally, there is no reason why more than two NDs cannot be introduced in the molecule. This may be critical in designing vaccines for a highly mutating pathogen like HIV.

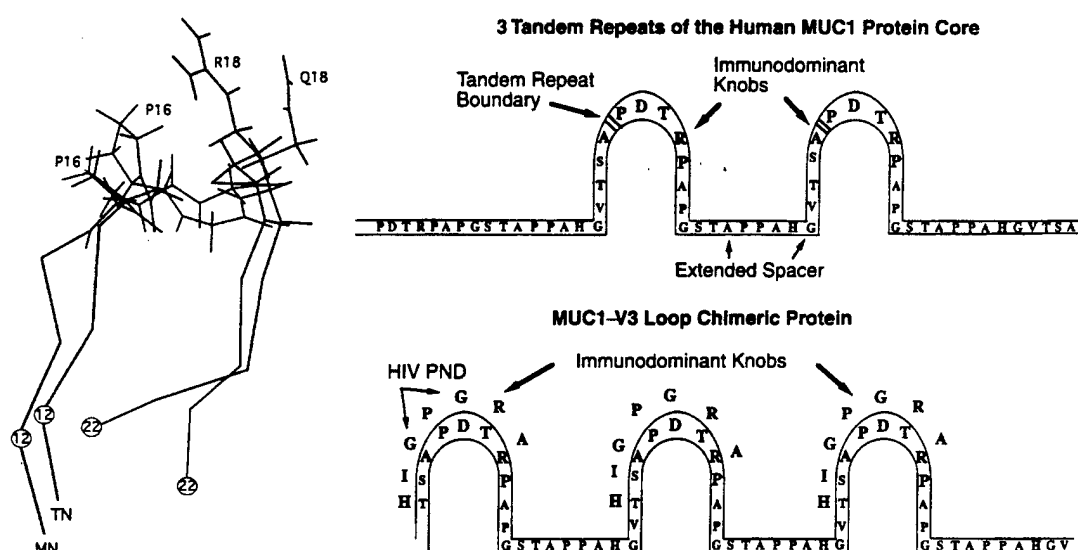
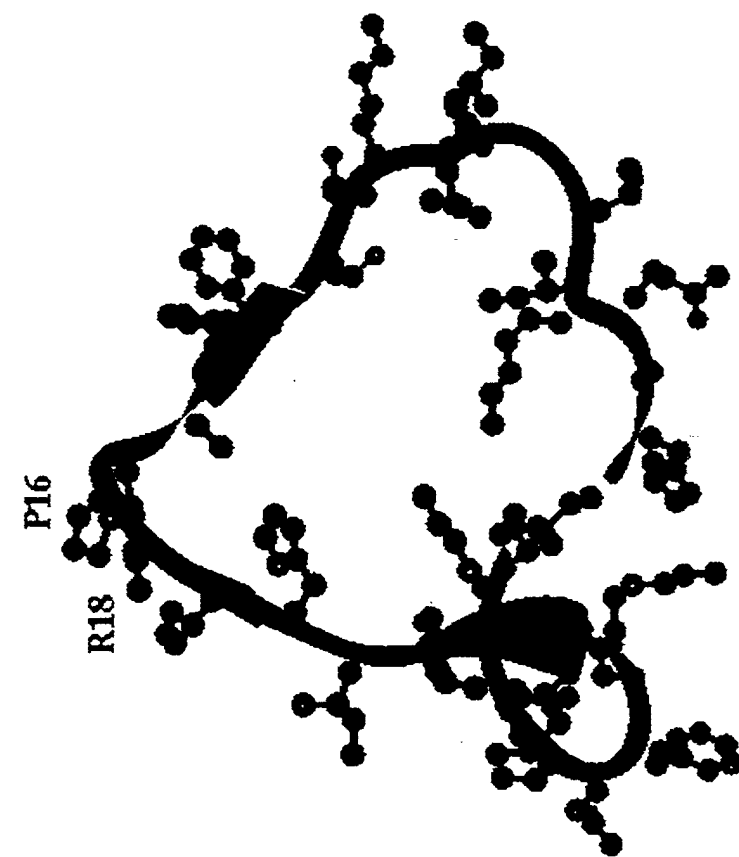


Figure 6. (A) Superimposition of the protruding motifs of two NMR structures: the V3 loop from the HIV-MN isolate (designated as MN) and that from the Thailand TN243 isolate (named TN). The sequences of two motifs are: MN, RIHIGPGRAFYT and TN, SITIGPGQVFYR. Note that the GPGR or GPGQ crests are oriented in the same way. (B) The principle of design. The V3 sequences above the Muc-1 sequences actually replace the Muc-1 residues in the chimeras.

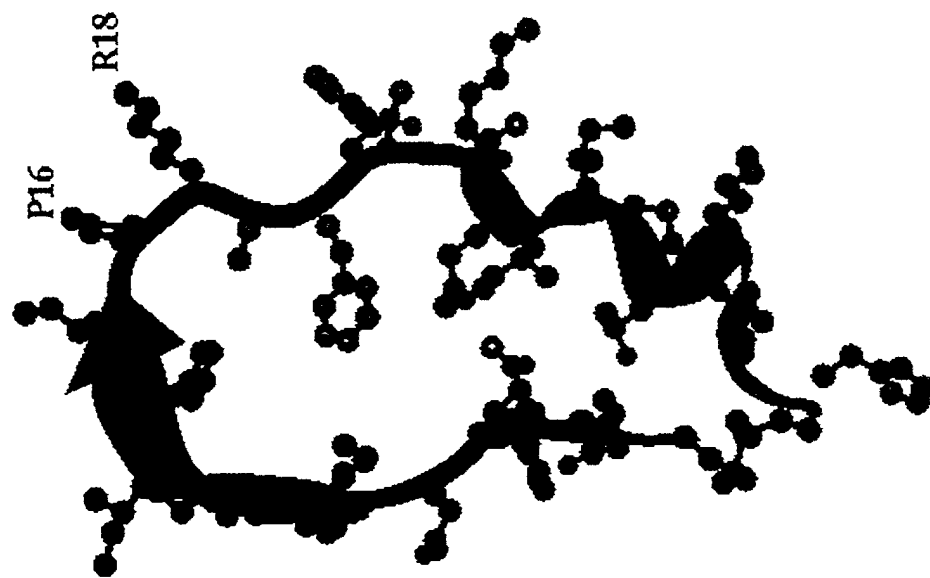
C.1c Sequence Variability at the Two Ends of the ND: Camouflaging of the Conserved Secondary Structural Element

NMR studies on the V3 loop sequences listed in Figure 4 are summarized as follows (Figure 7A-C). (i) A GPG type II turn is present at the crest of the V3 loop in all the sequences. (ii) Stretches of β -strand adjacent to the GPG-turn on the N- and C-terminal side are common to all the sequences. (iii) The residues in the C-terminal segment form a few turns in water and a helix in the less polar mixed solvent. (iv) In spite of the constraints of secondary structures [(i)-(iii)] and the disulfide

Solution Structures of V3-MN



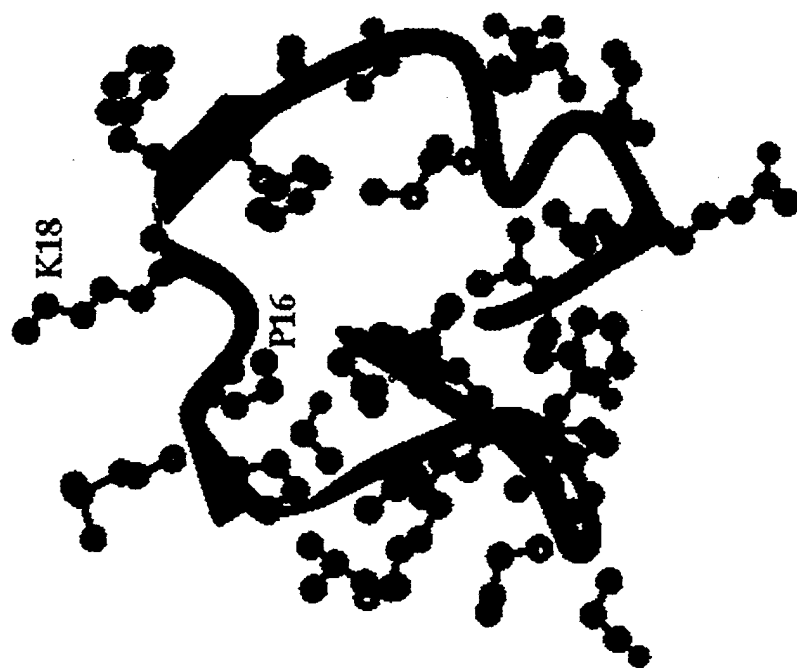
Water



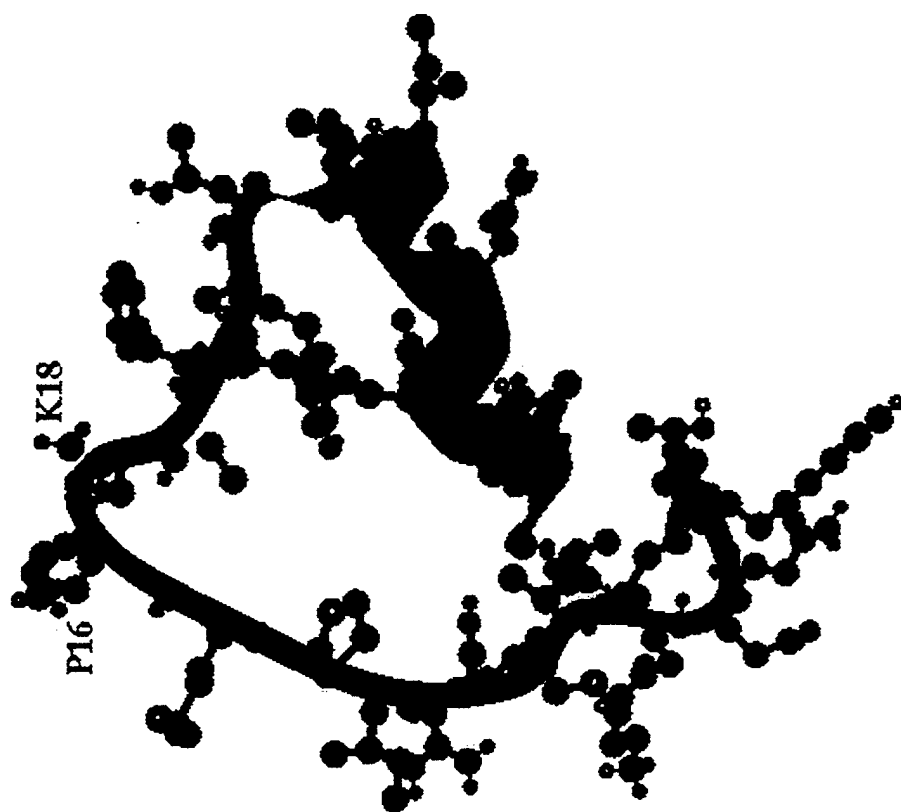
Water/TFE mixture

Figure 7A

Solution Structures of V3–Haiti



Water



Water/TFE mixture

Figure 7B

Solution Structure of the V3-RF in Water

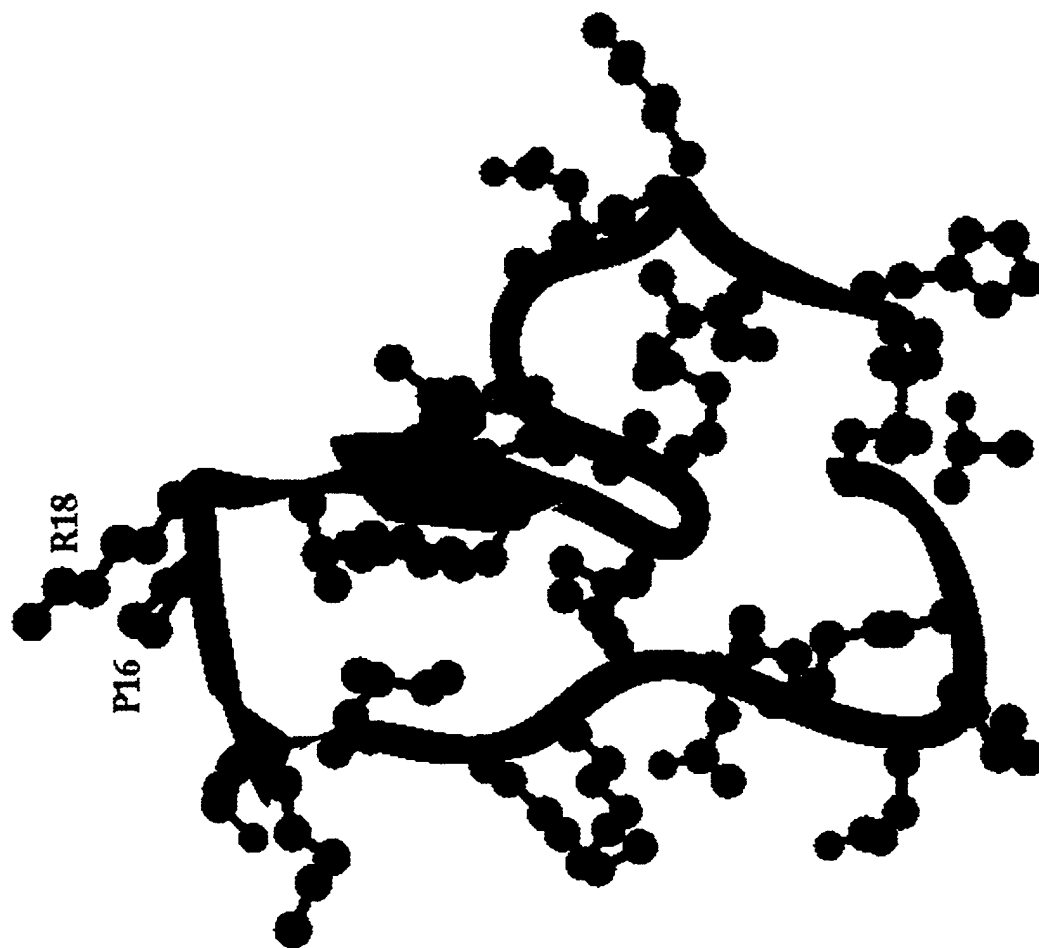


Figure 7C

Structural Camouflaging of the Neutralizing Domain inside the V3 Loop of the gp120 Haiti isolate

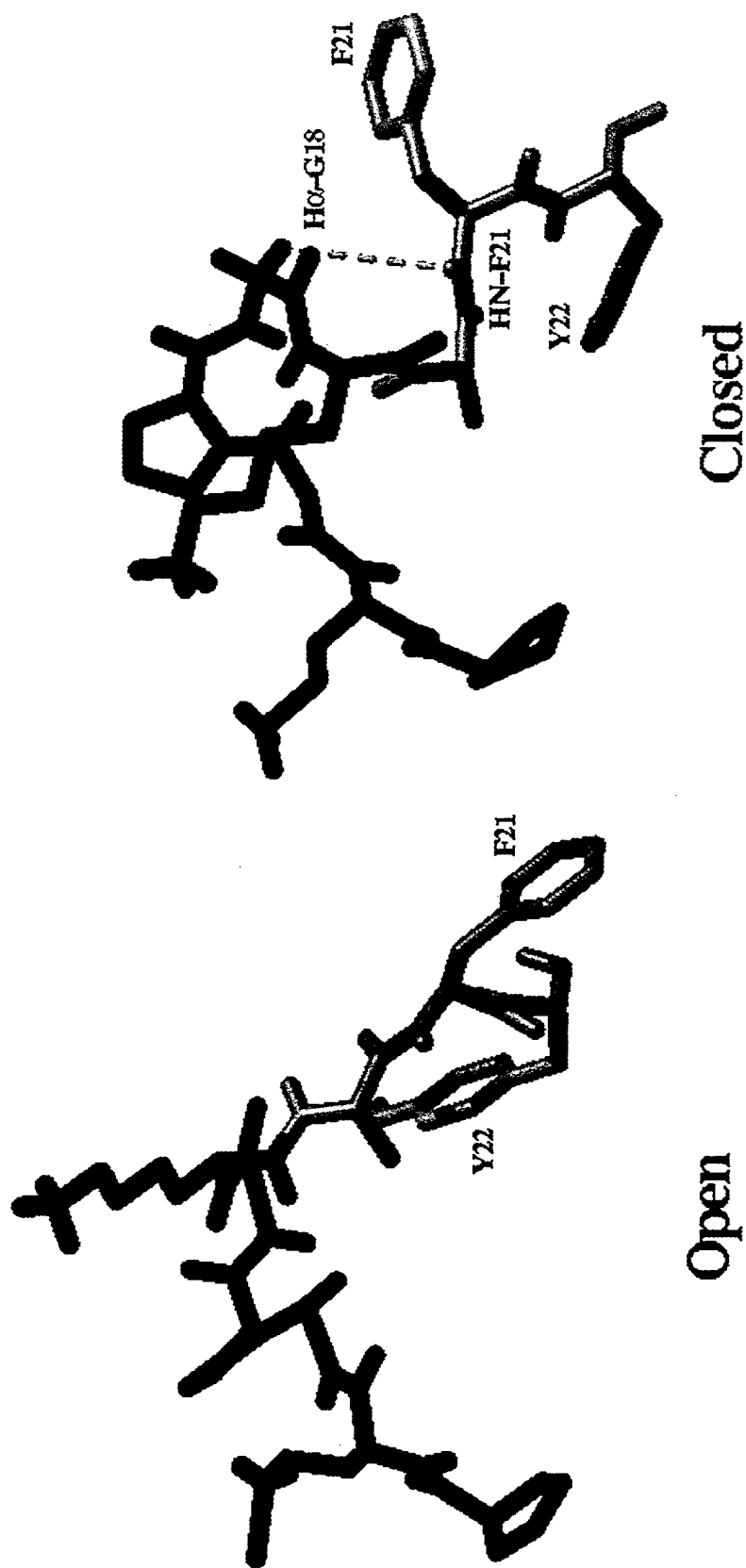


Figure 7D

bridge, the V3 loop exhibits conformational flexibility as evidenced by the absence of long range NOESY interactions commonly observed in well folded globular proteins. However, a "protruding knob" formed by the central GPG-turn and the β -strands on either side emerges as the secondary structural feature conserved among diverse V3 loop sequences. The single crystal structure of the HIV-1 neutralizing antibody (mAb 50.1) complexed to 16-residue long linear MN V3 fragment shows the hint of such a "protruding knob" although the segment on the C-terminal side of the GPGR type turn remains disordered [88]. The crystallographic observation suggests that the protruding knob of the V3 loop that includes the neutralizing epitope might well be specifically recognized by the antibody. However, we cannot count on the fact that the conserved "protruding knob" of the V3 loop will always be presented in its conformationally pure form because HIV will always find a way to mask this conserved secondary structural element. In this work we report one such mechanism of masking as revealed by the "close" state in Figure 7D. In this form of the Haitian V3 loop, the NMR data indicate an arching of the residues on the C-terminal side of the GPGK-turn. This is a departure from the "protruding knob" motif that contains the central GPG-turn and two β -strands on either side. Such an arched conformation of the neutralizing epitope has also been observed in an antibody (mAb 59.1) complexed with a linear V3 fragment [36]. When combined with the single crystal data on mAb-V3 complex, our NMR data indicate that the "closed" or "arched" conformation of the neutralizing epitope of the V3 loop is possible and can be recognized by the antibody. In addition, our data also indicate that an equilibrium between the "closed" and "open state" (Figure 7D) is possible. The arching around A20-F21 tends to mask A20 and F21 as shown by the solvent exposure data of the open and close forms of the Haitian V3 loop. The close form of the V3 loop may camouflage some essential elements of the neutralizing epitope from the immune system. For instance, this masking will interfere with the binding of antibodies that recognize the PGRAF epitope. Most importantly such a local masking of A20 and F21 should affect the proteolysis of the R/Q/K19-A20 peptide bond by thrombin and tryptase [66-67]; the second enzyme lies on the T-cell surface. When gp120 is used as a substrate unlike other proteases these two enzymes show exceptional specificity for cleavage of the R/Q/K19-A20 peptide bond inside the V3 loop. The most striking is the observation that the V3 loops of T-cell tropic virus strains are 1,000 times more susceptible to cleavage by these two enzymes than the V3 loops of macrophage tropic strains [67]. The T-cell tropic V3 loops are more positively charged than the macrophage tropic V3 loops [53]. Our studies reveal that the open state of the neutralizing epitope of the V3 loop is exclusively preferred for MN and RF V3 loops with net charges $\geq +5$ whereas the close state of the neutralizing epitope begins to appear for the Haitian V3 loop with net charge of +3. Therefore, we believe that the proteolysis data [66-67] are consistent with our structural conclusions.

We have carried out 2D NMR and molecular modeling studies on three mini V3 loops which are 17-amino-acids-long derived from the MN, Florida, and Thailand V3 sequences (Figure 5). These mini V3 loops contain the central GPG and the flanking sequences that are required for antibody binding. We show that the presence of the (S-S) bridge between the 1st and the 17th C leads to a β -

hairpin conformation for all three mini V3 loops with quite different sequences. Therefore, by this design the conformational camouflaging can be avoided.

C.2 The V3 Loop in the Context of the Native gp120

C.2a An Insight into the Neutralization Escape Mutants of HIV-1: Molecular Modeling and Neutralization Assays on gp120 Mutants with Single Site Mutations inside the V3 Loop

A dramatic effect of a single amino acid substitution in the V3 loop was encountered by monitoring a patient over a period of time. The patient, in 1985, had the virus with the GPGRA sequence at the crest of the V3 loop. mAb (M77) that binds to this V3 loop could neutralize the virus. In 1987, the virus isolated from the same patient showed only one change in the V3 loop, i.e., A21T substitution. The single A21T substitution had a dramatic effect: (i) M77 could no longer bind the singly mutated V3 loop and, therefore, (ii) M77 could not neutralize the new virus of 1987 thus leading to an escape mutant due to a single A21T substitution.

Monte Carlo (MC) Simulated Annealing studies [85-86] suggested that the residue 21 is involved in M77 recognition via "indirect reading" (Figure 8-1) as opposed to "direct reading" (Figure 8-2) in which residue 21 should make direct contacts in the antibody binding cavity. In the "indirect reading" mechanism, the contact domain of the V3 loop (amino acids 14-27 in Figure 8) in the native V3 loop forms a protruding moiety which shows surface complementarity in the antibody binding pocket of M77; A21 is not on the surface of the binding domain but in the specific cavity which can only accommodate A21 and not T21 (Figure 8-1). Therefore, A21T substitution would result in the expansion of the cavity and the surface area of the contact domain, thereby leading to the loss of M77 specificity.

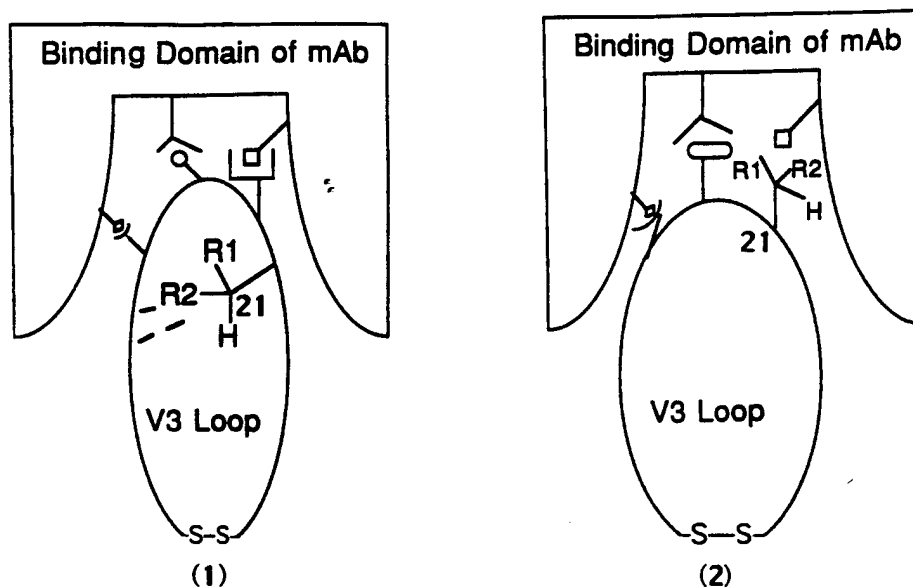


Figure 8. "Indirect Reading" (1) and "Direct Reading" (2) mechanisms of antibody (mAb M77) recognition. Four different residues, i.e., A, T, S, and I, occupy the position 21 of the V3 loop.

If the hypothesis of "indirect reading" mechanism of M77 recognition is true, the following predictions can be made: (i) A21S substitution should not alter the M77 specificity because A and S have about the same size, and therefore both should be accommodated in the same cavity, and (ii) A21I substitution should result in the loss of M77 specificity because I21 will be too large to fit in the same cavity that snugly fits A21. Both predictions were experimentally tested and were found to be true [85-86]. The work was done in collaboration with Drs. Fulvia Veronese and Marjorie Robert-Guroff of the NCI/NIH. Figure 8 shows that A21 is involved in interactions (marked by dashed lines in 8-1) with the residues that are distant from the GPG-crest. Subsequent to our work, the crystal structure of a linear V3 epitope complexed with a broadly neutralizing antibody (mAb 59.1) has been reported [89]. In this complex, the residue 21 also stays far from the GPG-crest.

C.2b Effect of Single Site Mutations at the GPGR-Crest of the V3 Loop of gp120

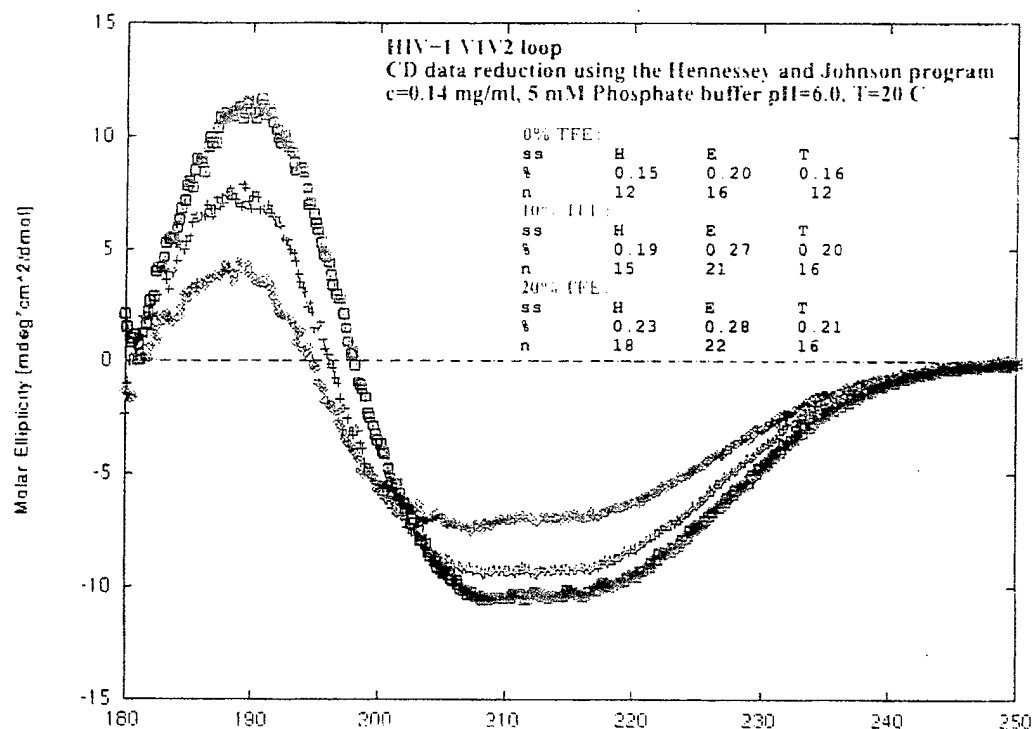
It has been shown that any mutation in the GPG-sequence of the V3 loop that destabilizes the type II turn also affects the fusion activity of the virus. This suggests that the type II turn in the V3 loop is critical in the life cycle of the virus. The four residues in GPGR/K/Q are numbered as G1, P2, G3, and R/K/Q4, respectively. The positions 1 and 3 in the (ϕ, ψ)-plot show stereochemical preference for G. This is especially true for the position 1 which should strictly prefer G and probably can accommodate A with a distortion in the turn [90]. Interestingly HIV-1 mutant with G1->T1 mutation in the type II turn leads to a non-infectious virus for Sup-T1 T cells [61]. However, after 40 days of coculture a T1->A1 revertant is identified. This revertant becomes infectious to Sup-T1 T cells. Note that T->G mutation requires 3 base changes in the codon while T->A reversion requires a single change in the second position of the codon.

Experiments discussed above and our NMR data prove the importance of a type II turn at the GPG-crest of the V3 loop for viral pathogenesis. However, it remains unclear whether residues like R/K/Q are also critically needed following the GPG sequence. We (in collaboration with the NCI/NIH) have set up the following experiment (unpublished data) in order to examine the importance of a basic (R/K) or a neutral (Q) residue after the conserved GPG sequence. We have performed molecular modeling of various V3 loop sequences with a GPGE-crest (instead of GPGR/K/Q) and our results showed that the global structure of the V3 loop remains the same as well as the type II turn at GPGE. We have then constructed HIV-1 mutants with GPGR->E mutants inside the V3 loop of gp120. The mutant virus replicated as well as the wild one. The mutant virus also appears to express the same number of gp120 molecules on the surface as the wild one. However, the mutant virus is the NSI type while the wild one is the SI type. But after a passage of three weeks in the co-culture, there has been a revertant population (70%) of the virus. This revertant virus appears to change from GPGE to GPGK sequence and the virus becomes SI active. Note that E->K reversion requires only a single base change in the first position of the triplet codon. This leads us to conclude that R, K, or Q after GPG sequence is important in the life cycle of the virus.

C.3 Molecular Modeling of the V1-V2 and V4-C4 Subdomains

We have completed small-scale chemical synthesis and purification of the cyclic V1-V2 domain of the MN isolate. This domain is 79 amino-acid-long (Table 1) and contains two internal disulfide bridges of the structure shown in Figure 2. The presence of the two disulfide bridges is confirmed by mass spectroscopy. Figure 9 shows the CD spectra of the MN V1-V2 domain with different TFE solvent fractions. TFE is used to unmask the helical propensity of the residues inside the V2 loop (Figure 2). A helical content of 15% is observed even in the polar aqueous environment. With increasing fraction of TFE in the solvent mixture the helical content levels off to 23% at 20% TFE. This probably implies that the core helical segment is still present in water. In TFE:water mixture the end-fraying is arrested and a longer helix is stabilized. Interestingly, our secondary structure prediction [77-78] identified a helical stretch of 16 residues as shown in Figure 2. This agrees with the CD data at 20% TFE. Our MC simulated annealing method has produced an ensemble of energy-minimized structures in which a stable helix is located inside the V2 loop. In this model, the percentages of β -strand and turns are also close to what have been observed for the V1-V2 domain in 20% TFE. The fact that the helix region (Figure 2; Table 1) shows TFE-induced transition we have attempted to visualize how a helix- $\rightarrow$ β -strand transition can be accommodated in the V1-V2 under the constraint of disulfide bridges and minimum changes in the rest of the molecule. Similarly, CD data [92] and secondary structure prediction algorithm seem to indicate the presence of a putative helix inside C4 of the V4-C4 subdomain (see Figure 3).

Figure 9. (A) CD spectra of the two (S-S) bridged V1-V2 sub-domain at three different water/TFE ratios. Helical content is increased upon increasing the TFE % in the mixture. Note that a 12% helix is present in the structure even in aqueous environment. pH 6-8 did not alter the CD pattern. Induction of 20% helix is consistent with a helical stretch spanning 167-182 which can undergo a helix to β -strand transition inside the V1-V2 sub-domain.



We have utilized the secondary structure prediction algorithm and/or CD data to obtain low-energy folded models of the doubly disulfide-bridged V1-V2 and V4-C4 loops. The methodology consisted of the following steps.

Step1 (Prediction of Secondary Structures). The secondary structural elements are predicted for a cyclic loop sequence (with two or more disulfide bridges) by computing the probability S of a given residue i in the loop to adopt a k -type of conformation ($k = 'h'$ for helix; $k = 'b'$ for beta sheet; $k = 'c'$ for coil; $k = 't'$ for turn), where

$$S(k,i) = \frac{\sum_{l=-\gamma}^{\gamma} P(k, i+l)}{|l|+1}.$$

(The summation is over $l = -\gamma$ to γ , where γ = size of the window chosen to account for the effect of the neighboring amino acid residues: $\gamma = 5$ for h; $\gamma = 3$ for b; and $\gamma = 4$ for c or t.) $P(k, i)$ = potential for the k type of conformation of individual residue i derived from the analysis of the single crystal structures of about 65 proteins. The highest $S(k,i)$ determines the conformation k for the i residue. Use of any existing algorithm for secondary structure prediction is only 60% accurate. In order to improve accuracy, we test our predictions by requiring an S-S bridge formation that achieves local energy minima for the loop; this leads to the next in our method.

Step 2 (Generation of the Energy-Minimized S-S-Bridged Loop). This step involves obtaining an energetically stable S-S-bridged structure for a V3 loop sequence given the secondary structural states of the constituent amino acids residues as obtained after step 1. Appropriate ranges of (ϕ , ψ) values are assigned to all amino acids. For example,

$$\phi = -55^\circ \pm 25^\circ, \quad \psi = -55^\circ \pm 25^\circ \text{ for residues in a helix,}$$

$$\phi = -140^\circ \pm 30^\circ, \quad \psi = 140^\circ \pm 30^\circ \text{ for residues in a beta strand,}$$

$$\phi_{i+1} = -65^\circ \pm 20^\circ, \psi_{i+1} = -50^\circ \pm 20^\circ, \phi_{i+2} = -90^\circ \pm 20^\circ, \psi_{i+2} = 0^\circ \pm 20^\circ$$

for residues in a type-I turn,

$$\phi_{i+1} = -65^\circ \pm 20^\circ, \psi_{i+1} = 120^\circ \pm 20^\circ, \phi_{i+2} = 90^\circ \pm 20^\circ, \psi_{i+2} = 0^\circ \pm 20^\circ$$

for residues in a type-II turn.

Residues in the coil state are set free to choose any point in the allowed space. We simplify the sequence by assuming A for residues with side chains extending beyond C^β atom, except for the Ps and the terminal Cs. Our rationale for doing this is that the allowed (ϕ , ψ) space of residues with a side chain longer than A is only a subspace of that allowed for A.

We have obtained an S-S-bridged structure of a loop by using a linked-atom-least-square refinement equation that minimizes function F in the space (ϕ, ψ) :

$$F = \sum_{\lambda} \lambda |G_{\lambda}| + \sum_{ij} (d_{ij}^{mn} - D^{mn})^2,$$

where $G_{\lambda} (= |r_1 - r^0_1| = 0)$ indicates distance constraints for an S-S bridge. Distances in the S-S-bridged V3 loop configuration are defined as $r_1 = S(C1) - S(C35)$, $r_2 = C^{\beta}(C1) - S(C35)$, $r_3 = C^{\beta}(C35) - S(C1)$, and $r_4 = C^{\beta}(C1) - C^{\beta}(C35)$; corresponding equilibrium distances are $r^0_1 = 2.04 \text{ \AA}$, $r^0_2 = r^0_3 = 3.05 \text{ \AA}$, $r^0_4 = 3.85 \text{ \AA}$. λ_l indicates Lagrangian multipliers; d_{ij}^{mn} indicates distance between atom i (type m) and atom j (type n); and D^{mn} indicates the contact limit between atom (type m) and atom (type n). In this refinement (ϕ, ψ) of various residues are treated as elastic variables (i.e., variables with weights) such that by appropriate choice of weights the predicted secondary structural states of residues (after step 1) are minimally altered. This method guarantees a stereochemically orthodox structure for the S-S-bridged sequence. Finally, appropriate sidechains are attached (in place of As) to generate an actual sequence and the potential energy of the system is minimized in the $(\phi, \psi, \omega, \chi)$ -space using the force-field of Scheraga and co-workers. The total conformational energy, ETOT (KCal/Mole), has the following components:

$$\begin{aligned} \text{ETOT} = & \text{EES (Coulomb interactions between pairs of partial charges, dielectric constant} = 80) \\ & + \text{ETOR (Torsional energy due barriers around single and partially double C-N bonds)} \\ & + \text{ENB (van der Waal attraction and repulsion terms between non-bonded atom-pairs)} \\ & + \text{ESS (constraint energy due to S-S bonds)} \\ & + \text{EDIS (energy due to distance constraints as present in different secondary} \end{aligned}$$

Several starting configurations are chosen within the allowed domains of the (ϕ, ψ) -space.

Step 3 (Exploration of the Conformational Flexibility By Monte Carlo Simulated Annealing). The simulated annealing is performed in the following manner. *First*, a starting energy-minimized structure is chosen and Monte Carlo (MC) simulations are performed for 50,000 steps at 600K in the $(\phi, \psi, \omega, \chi)$ -space and the lowest energy configuration is stored. *Second*, 50,000 MC steps are repeated in several cycles of gradually decreasing temperature until a temperature of 100K is reached. *Third*, the lowest energy configuration at 100K is further energy minimized to a low energy-gradient.

Steps 1-3 are repeated for several different starting configurations. This results in several low-energy structures of V1-V2 and V4-C4. If these structures belong to the same energy basin, an average model can be obtained that represents a given global fold.

C4. A Model of gp120

We have used molecular modeling to construct a model of MN-gp120. Details are given in reference 86. Briefly, our methodology is as follows. The mature form of the gp120 protein of the MN variant of HIV-1 contains 513 residues. At present, we have modeled a fragment of gp120 spanning the V1-V2-C2-V3-C3-V4-C4 region (see Figure 1); the C1 and V5/C5 fragments are ignored. This fragment contain 6 disulfide bridges. The first six C's define the V1 and V2 loops (V1-V2), the 9th to

12th C's lock a central domain of the C2 region (C2b), the 13th and 14th C's define the V3 loop and the last four C's form the V4 and the C4 loops (V4-C4). We will use the following steps to model the fragment:

(1) We have chosen the average energy-minimized models of the following sub-domains except for V3 for which we have used the NMR model:

(a) the V1-V2 region (residues: 118-210)

(b) the C2b region (residues: 223-252)

(c) V3 loop (residues: 301-335)

(d) the V4-C4 region (residues: 381-445)

(2) We have joined these sub-domains with the appropriate linkers (i.e., C2, C3 etc., as shown in Figure 1).

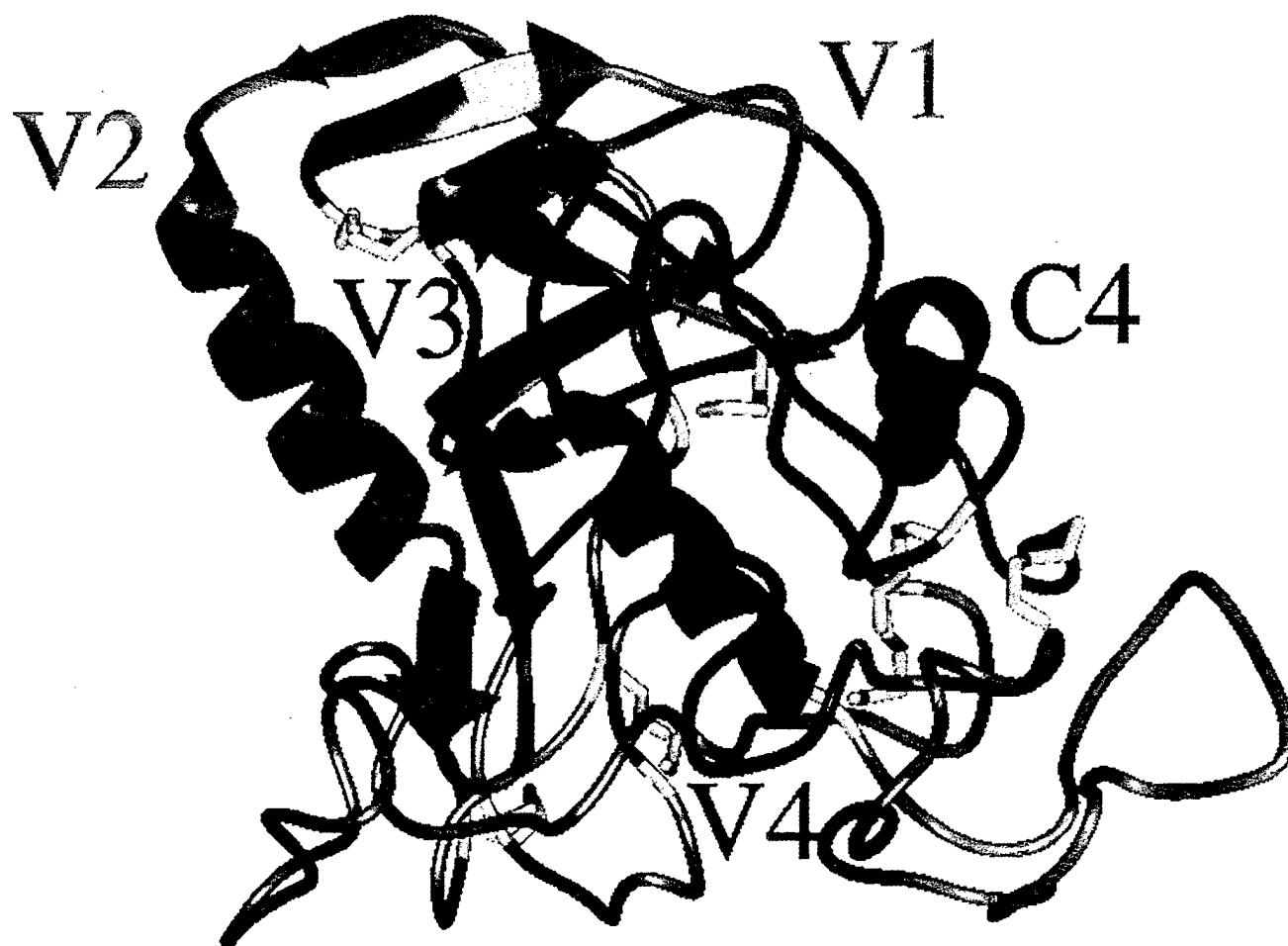
(3) We have performed MD simulated annealing to generate the (e) V1-V2-C2-V3 (residues:118-335) and the (f) V3-C3-V4-C4 (residues:301-445) fragments. V1-V2, C2b, V4-C4, and V3 have been varied around their average structures in accordance with their observed conformational flexibility as revealed by NMR or modeling. We have started with linkers in fully extended conformations. These regions include the N- and C-terminal peptides of the C2 region (residues: 211-222, C2a region, and residues: 253-300, C2c region). The two fragments have then been subjected to 200 ps of molecular dynamics (MD) at 500 K using united atom Amber 4.0 force field. We have used a bulky terminal group of van der Waals radius 5 Å to model minimal N-glycosylation. 100 structures (i.e., one after every 2 ps) have been selected from a 200-ps MD trajectory after equilibration. Representative structures for each fragment have been annealed and then minimized using 2000 steps of conjugate gradients.

(4) Once folded, the two fragments have been fused together by superimposing the helical region of the V3 loop present in both fragments. The structure have then been minimized to relieve initial strains from the docking of the two structures and equilibrated at 300 K for 50 ps of molecular dynamics. The configuration at the end of the 50-ps trajectory have been annealed and minimized for 2000 conjugate gradient steps. This has resulted in an energy-minimized model of the (V1-V2-C2-V3-C3-V4-C4) fragment of gp120. We have repeated this step with several different starting models of the (V1-V2-C2-V3) and (V3-C3-V4-C4) fragments.

(5) An ensemble of energy-minimized structures of the (C1-V1-V2-C2-V3-C3-V4-C4) have been tested against the surface accessibility data from the immunochemical maps [21-39, 69] and long-range interaction data from various functional assays [40-50].

(6) The screened gp120 models have been analyzed to study (i) the overall tertiary folding, (ii) the structures of various linear epitopes, (iii) specific interactions stabilizing intradomain interactions, (iv) masking of the constant regions by the variable regions, and (v) the role of inter-domain interactions in creating discontinuous epitopes. Finally, a few of the critical inter-domain interactions have been identified for the experimental testing of the model.

We have used the Molecular Surface Package, version 2.6.2., for computing the solvent accessible surfaces. For the solvent accessibility calculations a probe sphere of 1.5 Å of radius has been



The HIV-1 GP120

Figure 10. A ribbon diagram of the (V1-V2-C2-V3-C3-V4-C4) fragment of MN gp120. This is a representative model of gp120 that is consistent with the data from immunochemical maps and other functional assays. Color coding: green for V1, blue for V2, red for V3, magenta for C4, gray for C2 & C3, and yellow for the (S-S) bridges.

Interdomain hydrogen-bonds:

V2/V1:

N	165	ASD	O	128	LEU
N	124	LEU	O	200	SER
NZ	173	LYS	O	127	THR
N	126	VAL	OD1	172	ASP
N	127	THR	OD1	172	ASP
N	127	THR	OD2	172	ASP
N	148	SER	OD2	185	ASP
OG	169	SER	OG1	127	THR
OG	200	SER	OG	148	SER
OG	200	SER	OG	150	SER
ND2	145	ASN	OD2	185	ASP

C2/V1:

N	204	SER	O	129	ASD
N	208	GLN	O	119	VAL
N	209	ALA	O	119	VAL
N	119	VAL	O	209	ALA
N	121	LEU	O	206	ILE
N	131	THR	O	204	SER
OG	204	SER	O	122	THR
OG1	207	THR	O	153	THR
NZ	212	LYS	O	148	SER
NZ	241	LYS	O	142	SER
N	147	ASN	OE1	216	GLU
NZ	237	LYS	OG1	143	THR
NZ	241	LYS	OE1	151	GLU
ND2	146	ASD	OE1	216	GLU
ND2	146	ASD	OE2	216	GLU

C2/V2:

N	203	THR	O	199	ILE
N	215	PHE	O	185	ASP
N	190	ASP	O	271	ALA
N	192	ASP	O	285	ASN
N	202	ASD	OE2	177	GLU
NZ	236	LYS	O	187	VAL
NZ	236	LYS	O	192	ASP
NE	278	ARG	O	184	LEU
N	287	LYS	OD1	192	ASP
N	288	THR	OD1	192	ASP
N	289	ILE	OD2	192	ASP
NZ	183	LYS	O	273	GLU
N	188	SER	OE2	273	GLU
NZ	212	LYS	OD1	185	ASP
NZ	236	LYS	OG1	194	THR
OG1	288	THR	OD1	192	ASP
NZ	183	LYS	OE1	273	GLU
NZ	183	LYS	OE2	274	GLU
OG1	194	THR	OG1	288	THR

V3/V1:

ND2	307	ASN	OD1	138	ASN
NZ	308	LYS	OG1	136	THR
NZ	308	LYS	OD1	138	ASN

V3/V2:

OG1	323	THR	O	195	SER
N	197	ARG	OG1	323	THR

V3/C2:

N	310	LYS	OE1	208	GLN
N	312	ILE	OG1	203	THR
NH1	331	ARG	O	288	THR
NH1	331	ARG	O	291	VAL
NE2	208	GLN	O	310	LYS
NH2	303	ARG	OD1	300	ASD
ND1	313	HIP	OD1	202	ASD
NH1	331	ARG	OD1	235	ASP
NH2	331	ARG	OD1	235	ASP
NH2	331	ARG	OD2	235	ASP

C3/C2:

N	338	SER	O	296	SER
OG	338	SER	O	296	SER
OH	387	TYR	O	289	ILE

C3/V3:

N	336	ASD	O	334	HIP
---	-----	-----	---	-----	-----

V4/V3:

N	396	SER	O	329	THR
N	408	THR	O	309	ARG
N	309	ARG	O	406	ASD

V4/V3 (continue):

N	319	ALA	O	404	TRP
OG	396	SER	O	329	THR
N	397	THR	OE1	332	GLN
N	402	ASN	OG1	322	THR
N	403	THR	OG1	322	THR
N	307	ASN	OG1	408	THR
NH1	309	ARG	O	404	TRP
NH2	309	ARG	O	404	TRP
NZ	310	LYS	O	406	ASD
N	311	ARG	OD1	405	ASD
ND2	325	ASN	O	400	GLY
N	325	ASN	OG1	403	THR
OG1	403	THR	OG1	322	THR
ND2	305	ASN	OG1	397	THR
NZ	310	LYS	OD1	406	ASD
NE	311	ARG	OD1	405	ASD
NZ	324	LYS	OG1	408	THR
NE2	332	GLN	OG1	397	THR

V4/C3:

N	417	GLN	O	377	HIP
N	371	ASP	O	390	THR
OG	391	SER	O	373	GLU
ND1	377	HIP	O	415	THR
OG1	390	THR	OD2	371	ASP
NE2	377	HIP	OG	391	SER

C4/V1:

NE2	428	GLN	O	137	THR
ND2	138	ASN	O	427	TRP
OH	435	TYR	OD1	140	ASD

C4/C2:

N	299	ILE	O	438	PRO
NE2	442	GLN	OD1	300	ASD

C4/V3:

ND2	300	ASD	O	436	ALA
ND2	300	ASD	O	437	PRO
N	302	THR	OE1	442	GLN
NH1	303	ARG	O	442	GLN
NH2	303	ARG	O	441	GLY

C4/C3:

N	383	GLY	O	443	ILE
N	442	GLN	OD1	344	ASP
N	443	ILE	OD1	344	ASP
NE	444	ARG	OD2	344	ASP
NH1	444	ARG	OD2	344	ASP
NH2	444	ARG	OE1	355	GLU
NH1	444	ARG	OE2	355	GLU
NH2	444	ARG	OE2	355	GLU
NZ	341	LYS	OE1	440	GLU
NZ	341	LYS	OE2	440	GLU
OG1	345	THR	OE1	440	GLU

C4/V4:

N	419	LYS	OD1	413	ASD
NE2	422	GLN	O	416	LEU
N	423	ILE	OD1	412	ASN
NE1	427	TRP	O	409	GLY
ND2	413	ASD	O	419	LYS
ND2	413	ASD	O	420	ILE
NE2	422	GLN	OD1	413	ASD

Interdomain hydrophobic interactions:

V1-V2/V3:

186	ILE	327	ILE
196	TYR	327	ILE
199	ILE	313	HIP
199	ILE	321	TYR

V4-C4/V3:

394	PHE	330	ILE
398	TRP	326	ILE
404	TRP	319	ALA
404	TRP	320	PHE
416	LEU	306	TYR
426	MET	306	TYR
427	TRP	306	TYR
434	MET	304	PRO

Table 3

Figure 11A. Inter-domain interactions of the residues in the (V1-V2) sub-domain. Color coding: purple for V1, yellow for the (S-S) bridges, cyan for the linear epitope in V2, purple for the conformational epitope in V2, gray for the rest of the V2, green for the interacting polar residues, and red for the interacting non-polar residues. Note that the linear V2 epitope adopts a helix whereas the conformational V2 epitope adopts a b-hairpin in the folded gp120. Also note that the b-hairpin is involved in several inter-domain long-range interactions with the residues in C2 and V4. The turn at residues 135-140 of V1 makes contact with W427 in C4; W427 is critical for the binding of the CD4-blocking antibodies.

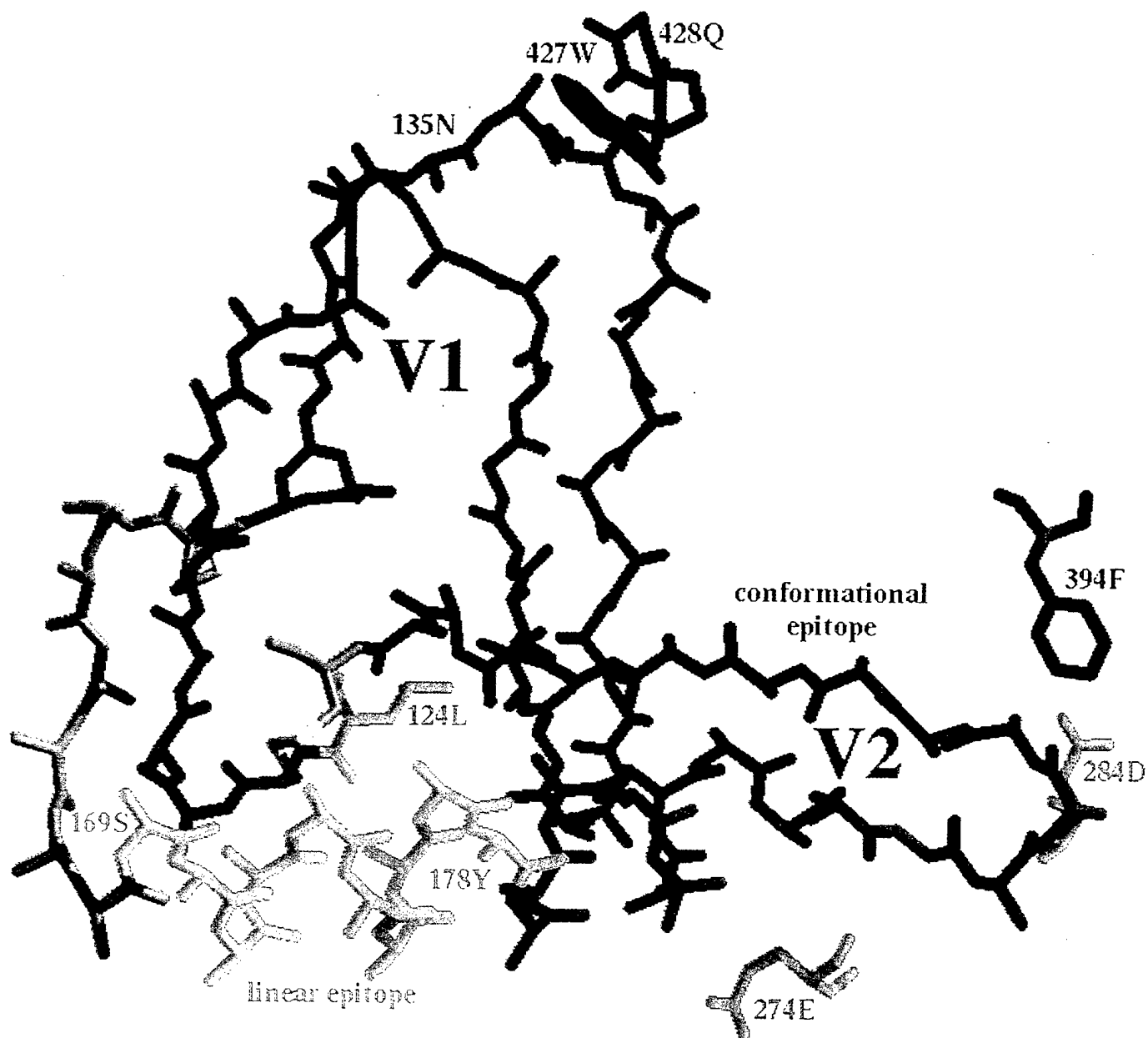
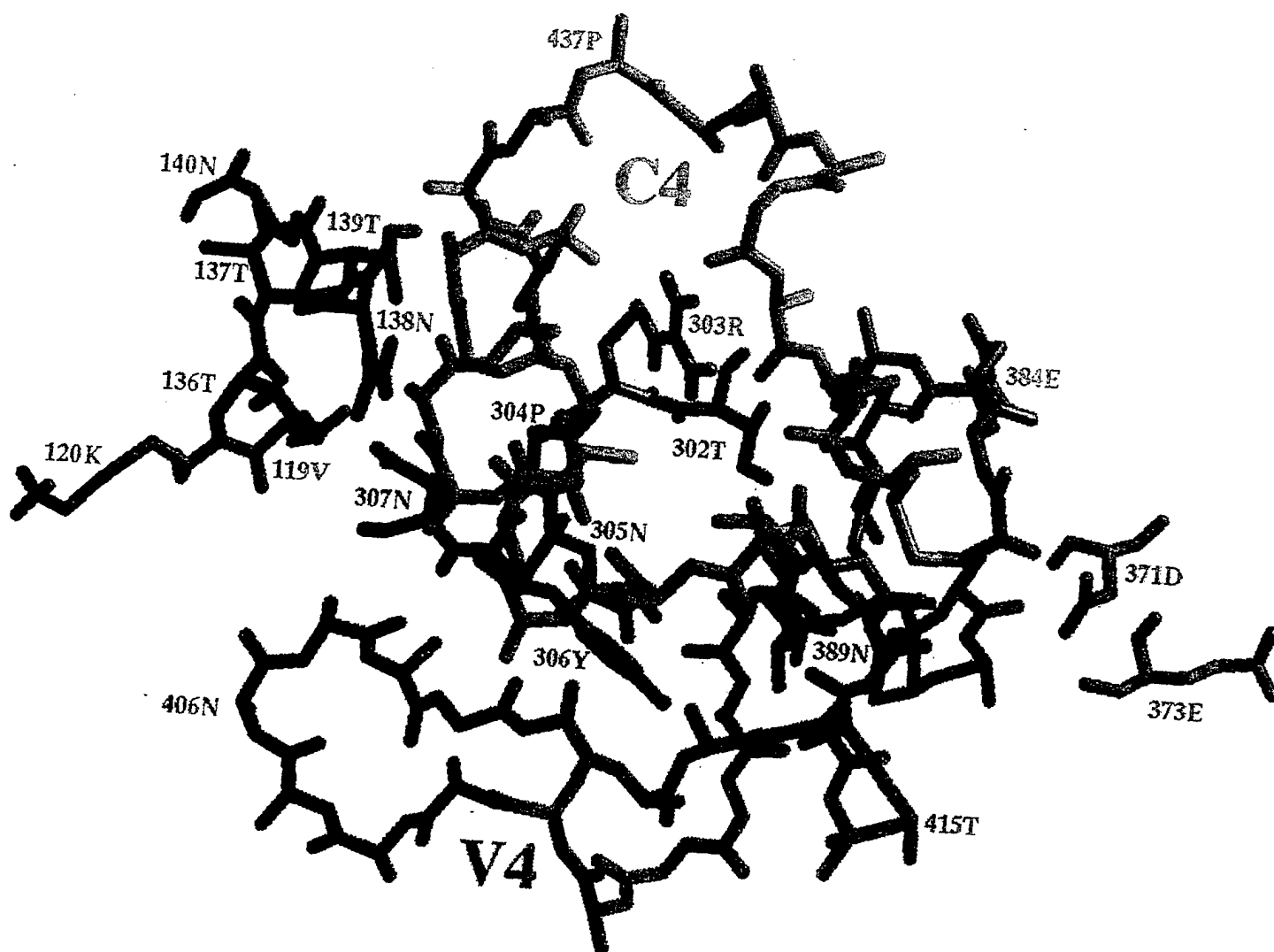


Figure 11B. Long-range inter-domain interactions involving residues from the V1, V3, and C4 loops. Color coding: purple for V4, blue for C4, yellow for the (S-S) bridges, green for interacting charged residues from the V1, V3, and C3 regions, red for the interacting non-polar residues from the V1, V3, and C3 regions. A turn at V1, the N-terminal V3 segment, and a helical stretch in C4 form an intricate contact interface. P437 is critical in inducing a sharp turn in C4.



used. To compute the residue-specific accessible surfaces, the accessible surface for each residue (X) is normalized by the surface of the same residue (X) in a fully extended conformation of G-X-G.

Figure 10 shows a representative low energy structure of the V1-V2-C2-V3-C3-V4-C4 fragment of gp120. It is clear from Figure 10 that the tip of the V3 loop, the helix inside the V2, and a part of the C4 loop are all solvent exposed. Theoretical solvent accessibility data are computed for all the residues in our model of the V1-V2-C2-V3-C3-V4-C4 fragment; note that (see Appendix I) the computed values agreed well with the observed data [69]. Table 3 lists a set of H-bonding interactions stabilizing intra- and inter-domain structures. Three types of H-bonds in Table 3 originate from three different types of interactions, namely, backbone-backbone, backbone-sidechain, and sidechain-sidechain (the last two interactions are sequence specific). For example, in our model (Figure 11) OH-Y435(of C4) is H-bonded to OD1-D22(of V1) and NZ-K341(of C3) is H-bonded to OE1/OE2-E440 (of C4). Table 3 also lists key hydrophobic interactions involving residue-pairs from different domains. Most of these interactions involve residue-pairs inside a tight cavity showing either van der Waal contacts [e.g., I186(of V2)---I337 (of V3)] or stacking overlaps [e.g., Y306 (of V3) and W427 (of C4)]. Figure 11 shows the inter-domain interactions involving (A) (V3 and V1/V2) and (B) (V3 and C4) , i.e., **N*135-T136-T137-N138-N139-N140** (of V1), **N*300-C301-T302-R303-P304-N305-Y306-N307** (of V3), and **Q422-I423-I424-N425-M426-W427-Q428-E429-V430-G431-K432-A433-M434-Y435-A436** (of C4)- key residues in these segments are marked in bold (N*=glycosylated N). As shown in Figure 11A, the linear V2 epitope adopts a helix whereas the conformational V2 epitope adopts a β -hairpin in the folded gp120. Also note that the β -hairpin is involved in several inter-domain long-range interactions with the residues in C2 and V4. The turn at residues 135-140 of V1 makes contact with W427 in C4; W427 is critical for the binding of the CD4-blocking antibodies. Experimental data also support the presence of (V1 and C4) interactions in the native gp120. Table 3 shows that K183/L184/D185 of the V2 loop are all involved in inter-domain H-bonds. The C=O of L184 is H-bonded to the basic sidechain of R278 whereas the acidic sidechain of D185 is H-bonded to the basic sidechain of K212; also the sidechains of E274 and L184 are within 5 Å. Therefore, if such a β -hairpin is critical for antibody binding a double site mutation, L184/D185-D184/L185, would be catastrophic because this would bring D184 close to E274 and L185 in steric clash with K212 (both of which would destabilize the β -hairpin). Indeed, a L184/D185-D184/L185 double mutation reduces the binding affinity of antibodies specific for conformational epitopes inside the V2 loop. Table 3 shows that R303 and N*300 form sidechain-sidechain H-bonding which stabilizes a turn at residues 300-303. This turn (as shown in Figure 11) brings the Y306 ring in close proximity with the W437 ring. The turn at residues 300-303 also locks R303 in H-bonding interactions with (G441 and Q442) and N*300 in H-bonding interactions with (A436 and P437)-- see Table 3. The key residues, W427 and Y435, also show sidechain-backbone H-bonds with N138 and N*140 (Table 3). Therefore, we predict that the residues N*300 and R303 of V3 and the residue W427 of C4 are critical in bringing V1, V3, and C4 in spatial proximity. Interestingly from antibody binding studies, it has been demonstrated that R303G substitution exposes the N-terminal V3 fragment; from our model we predict that such a substitution would abolish several key inter-interactions

involving the (V1 and V3) and (V3 and C4) loops. Similarly, antibody binding studies reveal that W427S substitution exposes the residues 420-435 in the C4 fragment; from our model we also predict that such a substitution would disable several key inter-interactions involving the (V1 and V3) and (V3 and C4) loops.

Additional data on gp120 modeling in Appendix I include the following.

- (i) **(Figure 12)** The MN-gp120 sequence is aligned with IIIB-gp120 sequence (note that most of the experimental data have been obtained for IIIB-gp120).
- (ii) **(Figure 13)** A (ϕ, ψ) plot of the residues in V1-V2-C2-V3-C3-V4-C4 as shown in Figure 10 (this plot shows that most of the residues fall within the allowed region of the plot).
- (iii) **(Figure 14)** Solvent accessibility of various contiguous amino acid stretches computed for V1-V2-C2-V3-C3-V4-C4 as shown in Figure 10 (the computed values agree well those obtained from the immunochemical maps of gp120).
- (iv) **(Figure 15)** A close-up view of the (A) V1-V2 and (B) (V4-C4) loops taken from Figure 10 (note that the average low-energy models of these two loops are also quite similar).
- (v) **(Figure 16)** A close-up view of the β -hairpin (residues 183-188) that is an important part of the conformational epitope inside V1-V2 (the effect of L184/D185-D184/L185 mutations are described).
- (vi) **Table 4** Listing of inter-domain H-bonds involving backbone-backbone, backbone-sidechain, and sidechain-sidechain interactions.

Appendix II includes our structural biology work on the human mucin, Muc-1, a breast cancer tumor antigen. This tandem repeat protein is critically important in breast cancer research, a major research initiative in the US Army.

D. PUBLICATIONS

1. Catasti, P. and Gupta, G. Structural Studies Involving Different HIV-1 V3 Loops (1996) HIV Data Base (Eds. Myers et al.) Los Alamos National Laboratory III-125-129.
2. Gupta, G. and Myers, G. Analyses of the Various Folding Patterns of the HIV V3 Loops. In Peptides: Design, Synthesis, and Biological Activity (C. Basava and G. M. Anantharamiah, eds.), New York: Birkhauser (1994), Ch 17:259-277
3. Catasti, P., Fontenot, J.D., Bradbury, E.M., and Gupta, G. Local and Global Structural Properties of the HIV-MN V3 Loop. (1995) J. of Biol. Chem. 270:2224-2232.
4. Catasti, P., Bradbury, E.M., and Gupta, G. Structure and Polymorphism of the HIV-1 V3 Loops. (1995) J. of Biol. Chem. 271:8236-8242.
5. Catasti, P., Fontenot, J.D., Mariappan, S.V.S., Bradbury, E.M., and Gupta, G. Pau, C-P., Parekh, B.S., George, J.R. and Gupta, G. Human Immunodeficiency Virus (HIV) Antigens: Structure and Serology of Multivalent Human Mucin MUC1-HIV V3 Chimerical Proteins. (1995) Proc. Natl. Acad. Sci. USA 92:315-319.
6. Fontenot, J.D., Gatewood, J.M., Mariappan, S.V.S., Catasti, P., Domenech, P., Finn, O.J., and Gupta, G. Structure of a Tumor Associated Antigen Containing Tandemly Repeated Immunodominant Epitope. (1995) J. Biomol. Str. & Dyn. 13-245-260.
7. Gupta, G., Anantharamaiah, G.M., Scott, D.R., Eldrige, J.H. and Myers, G. Solution Structure of the V3 Loop of a Thailand HIV Isolate. (1993) J. Biomol. Struc. & Dyn. 11:345-366.
8. Vernose, F.D.M., Reitz, M., Guioff, M.R., Thompson, C.B., Gallo, R.C., Gupta, G. and Russo, P. and P. Russo. Generation of a Neutralizing Epitope by a Spontaneous Mutation in the V3 Loop of HIV-1(IIIB) Isolated From an Infected Laboratory Worker. (1993). J. Biol. Chem. 268:25894-25901.
9. Vernose, F.D.M., Reitz, M., Gupta, G., Guioff, M.R., Thompson, C.B., Gallo, R.C., and Russo, P. and P. Russo. Loss of a Neutralizing Escape Mutant by a Local Change in the HIV-1 V3 Loop Conformation. (1994). Vaccines 94.
10. Catasti, P., Bradbury, E.M., and Gupta, G. A Stereochemical Modeling of the HIV-1 Surface Envelope Protein, gp120. (1995). J. Virol. to be submitted.
11. Fontenot, J.D., Van Cott, T.C., Parekh, B.S., Pau, C-P., George, J.R., Birx, D.R., Zolla-Pazner, S., Gorny, M.K., and Gatewood, J.M. Presentation of HIV V3 Loop Epitopes for Enhanced Antigenicity, Immunogenicity, and Diagnostic Potential. (1995). AIDS 9: 1121-1129.

APPENDIX I

Figure 12. In-frame alignment of MN and IIIB gp120. Note that most of the immunochemical and other functional assays are performed on single and double site mutants of IIIB gp120 [21-50].

HIVMN

10	20	30	40	50	60	70	80								
MRVKGIRRN	YQHWWG	WTMLLGLLM	ICSATEKL	WVTVYYG	VPVWKEA	TTTLFCAS	DAKAYDTE	VHNVWATQ	ACVPTD	PNP					
90	100	110	120	130	140	150	160								
QEVELVN	VNTENF	NMWKNNM	VEQMHE	DIISLWD	QSLKPCV	KLTPLCV	TNLNCTD	LRNTTNT	NNNSTAN	NNNSNSE	GTIKGG	EMK			
170	180	190	200	210	220	230	240								
NCSFNIT	TSIRD	KMQKEY	ALLYKL	DIVSID	NNDSTS	YRLISC	NSTSVIT	QACPKIS	FEPIPI	HYCAPA	GFAILK	CNDKKF	SFG		
250	260	270	280	290	300	310	320								
KGSCKN	VSTVQ	CTHGIR	PVSTQ	LLNGSL	AEEVVIR	SENFTD	NAKTI	IIVHLN	ESVQIN	CTRPN	YNKRK	RIHIGP	GRAF		
330	340	350	360	370	380	390	400								
YTTKNI	IGTIR	QAHCN	ISRAK	WNDTL	RQIVSK	LKEQF	KNKTIV	FNQSSG	GDPEIV	MHSFNC	GGEFFY	CNTSPL	FNSTW	NG	
410	420	430	440	450	460	470	480								
NNTWN	NNTGS	NNNITL	QCKIK	QIINM	WQEVG	KAMYAP	PIEGQ	IRCSS	NITGL	LLLTR	DGGKDT	DTNDTE	IFRPGG	GDMRD	N
490	500	510													

WRSELYKYKVVTIEPLGVAPTAKRRVVQREKR

HIVHXB2R

10	20	30	40	50	60	70	80								
MRVKEKYQ	HLWRWG	WRWGTM	LLGMLM	ICSATEKL	WVTVYYG	VPVWKEA	TTTLFCAS	DAKAYDTE	VHNVWAT	HACVPT	DPN				
90	100	110	120	130	140	150	160								
PQEVVLN	VNTENF	NMWKND	MVEQM	HEDIISL	WDQSLK	PCVKLT	PLCVSL	KCTDLK	NDTNTN	SSSGRM	IMEKGE	IKNCSF	N		
170	180	190	200	210	220	230	240								
ISTSI	RKGVQ	KEYAFF	YKLDI	IPIDND	TTSYKL	TSCNTS	VITQAC	PKVSFE	PIPIHY	CAPAGF	AILKCN	NKTFNG	TGPCT		
250	260	270	280	290	300	310	320								
NVSTV	QCTHG	IRPVV	STQLL	NGSLA	EEVVIR	SVNFTD	NAKTI	IIVQLN	TSVEIN	CTRPN	NNNTRK	RIRIQ	RGPGR	AFVTI	
330	340	350	360	370	380	390	400								
GKIGN	MRQAH	CNISRA	KWNNTL	KQIAS	KLREQ	FGNKT	IIIFKQ	SSGDPE	IVTHS	FNCGGE	FFYCN	STQLFN	STWFN	STW	
410	420	430	440	450	460	470	480								
STEGS	NNTEG	SDTITL	PCRIK	QIINM	WQKVG	KAMYAP	PIISGQ	IRCSS	NITGL	LLLTR	DGGNS	NNNESE	IFRPGG	GDMRD	NWR
490	500	510													

SELYKYKVVKIEPLGVAPTAKRRVVQREKR

Figure 13. (ϕ, ψ) values of the residues in the gp120 model of Figure 10. Note that most of the residues lie within the allowed regions of the Ramachandran plot. A few non-Glycine residues in the disallowed region (i.e., the center of the right bottom quadrant) are marked. This plot proves the stereochemical feasibility of our model and the validity of our conformational search.

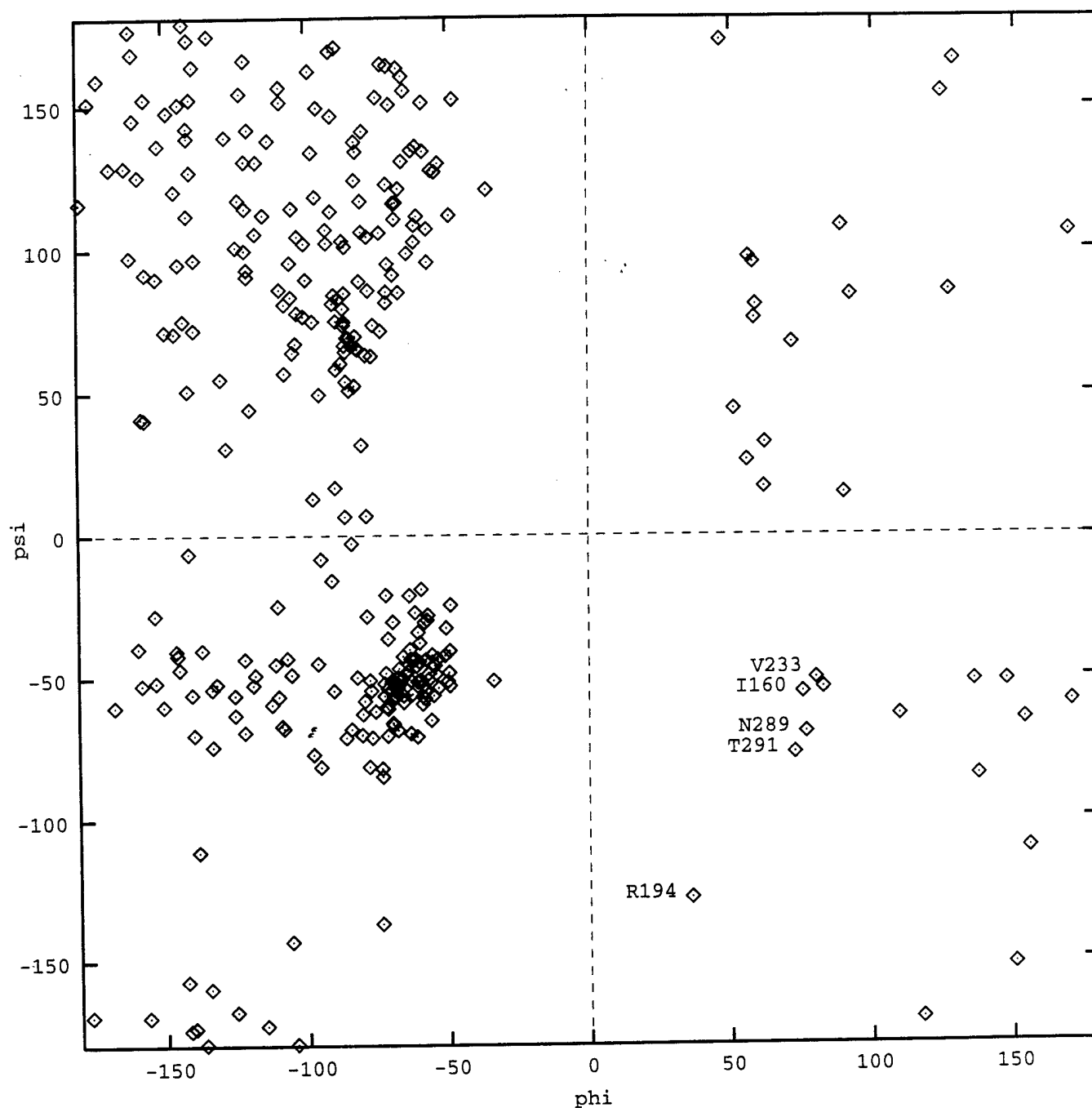


Figure 14. Site-specific solvent accessible surface area (in $\text{\AA}^2$). The MAS (maximum accessible surface) values are given for the most exposed non-hydrogen atom in a residue after scaling it by the surface area of the same atom in the same residue (X) of the extended G-X-G. Therefore MAS values well above 1 implies well exposed residues. Color coding: green for V1, cyan for V2, black for C2 and C3, red for V3, magenta for V4, and blue for C4. Observed accessibility data for different segments of gp120 are shown as bars, i.e., the bars with MAS > 4 represent well exposed segments, the bars with MAS ~1 represent poorly exposed segments, and the bars with MAS ~0 represent buried segments. The location and length of the bars define the epitope on gp120. Note that the calculated surface accessibility areas of the epitopes inside the V1, V2, V3, V4, and C4 loops agree well with the observed data [69].

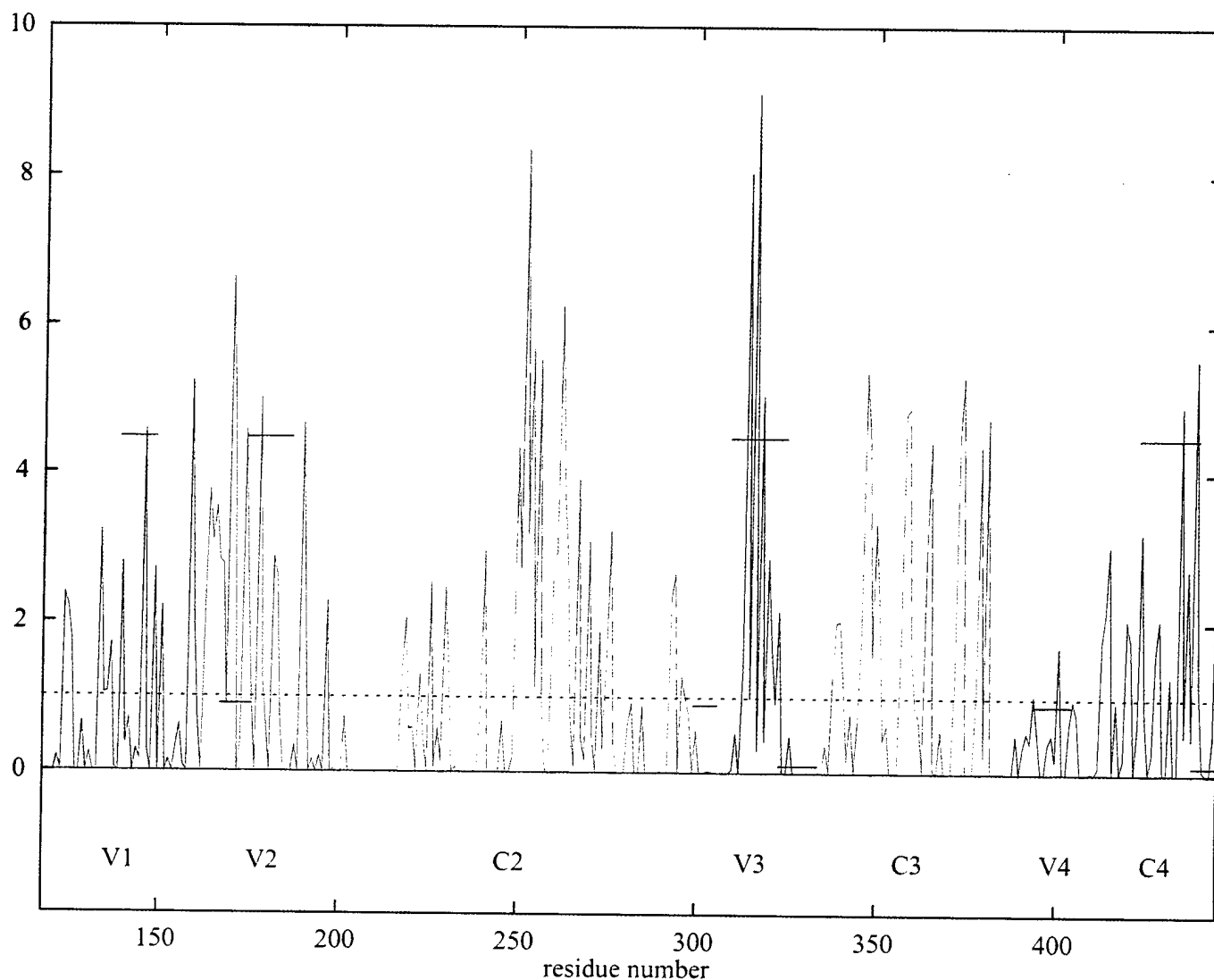


Table 4. Summary of intra- and inter-domain backbone-backbone (bb), backbone-sidechain (bs), and sidechain-sidechain H-bonds. The cut-off for the donor-acceptor distance is 3 Å and the cut-off for the donor-proton-acceptor angle is 120°. The presence of a large number of intradomain H-bonds inside the C2 and C3 sub-domains imply that they form well folded structures. Also note that there are a significant number of inter-domain H-bonds between the C2 and C3 sub-domains (which constitute a part of the discontinuous epitope for CD4 binding). The number of intra- and inter-domain H-bonds are averages over sampled gp120 models that are consistent with the data from the immunochemical maps [69] and other functional assays [40-50].

	V1	V2	C2	V3	C3	V4	C4	
V1	16	4	6	0	0	0	0	bb
	14	8	6	0	0	0	2	bs
	10	4	4	3	0	0	1	ss
V2	4	26	4	0	0	0	0	bb
	8	9	9	2	0	0	0	bs
	4	10	6	0	0	0	0	ss
C2	6	4	28	0	1	0	1	bb
	6	9	41	5	2	0	2	bs
	4	6	16	5	0	0	1	ss
V3	0	0	0	16	1	4	0	bb
	0	2	5	14	0	11	4	bs
	3	0	5	5	0	6	0	ss
C3	0	0	1	1	16	4	1	bb
	0	0	2	0	29	2	3	bs
	0	0	0	0	9	2	8	ss
V4	0	0	0	4	4	7	0	bb
	0	0	0	11	2	12	7	bs
	0	0	0	6	2	2	1	ss
C4	0	0	1	0	1	0	17	bb
	2	0	2	4	3	7	5	bs
	1	0	1	0	8	1	3	ss

Figure 15A. A close-up view of the V1-V2 loop taken from Figure 10. As in Figure 2, three disulfide bridges (colored yellow) are present in this subdomain. The V1 loop is shown in green whereas the V2 loop is shown in magenta and the rest in gray.

V1-V2 Domain

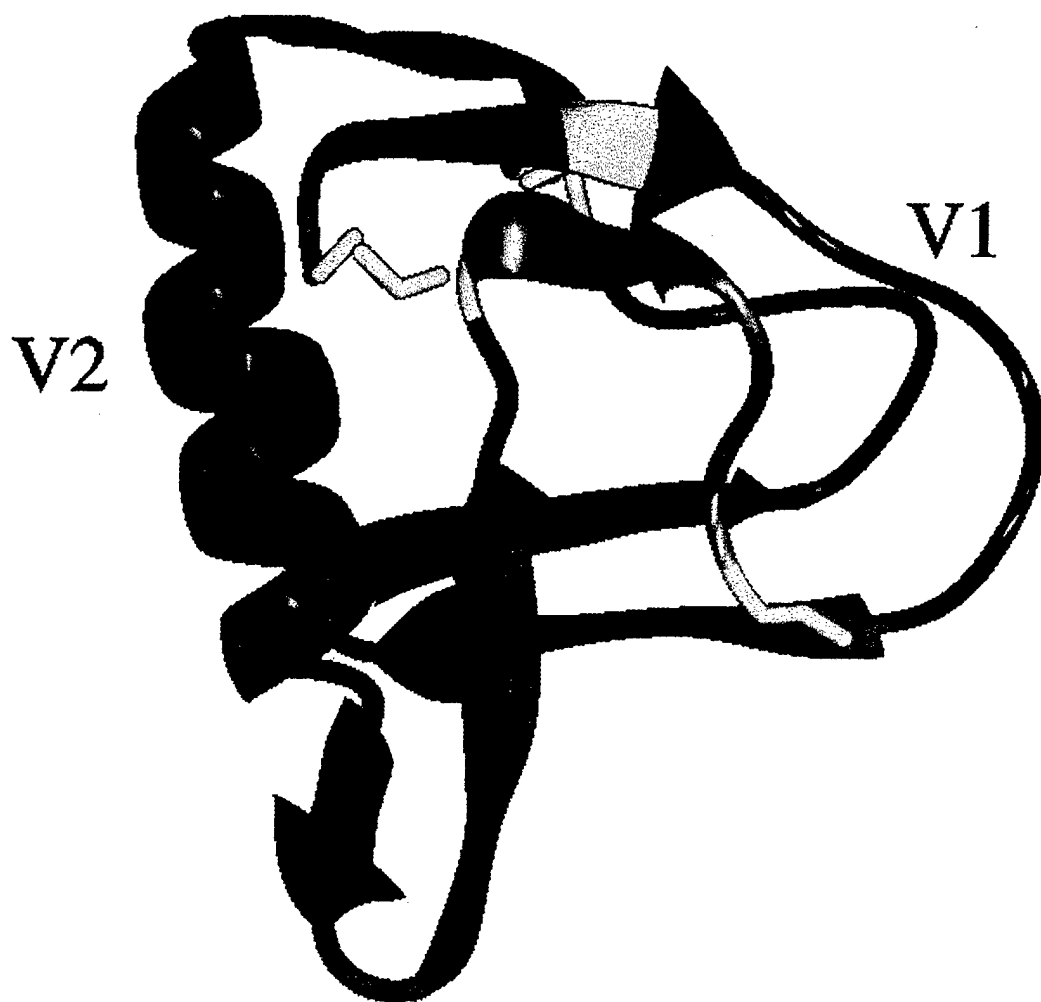


Figure 15B. A close-up view of the V4-C4 loop taken from Figure 10; color coding: V4 in green, C4 in magenta, and the two disulfide bridges in yellow.

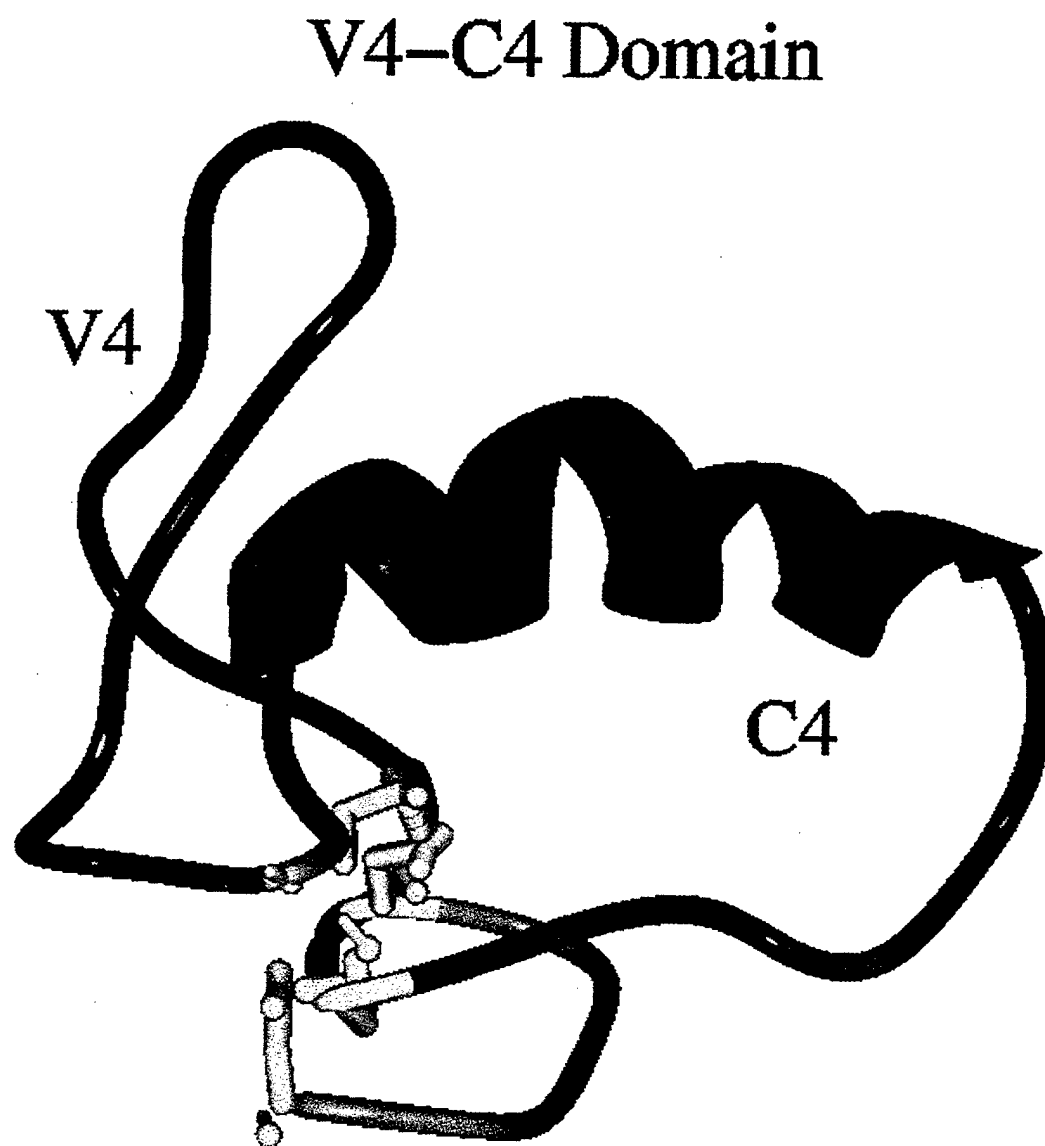
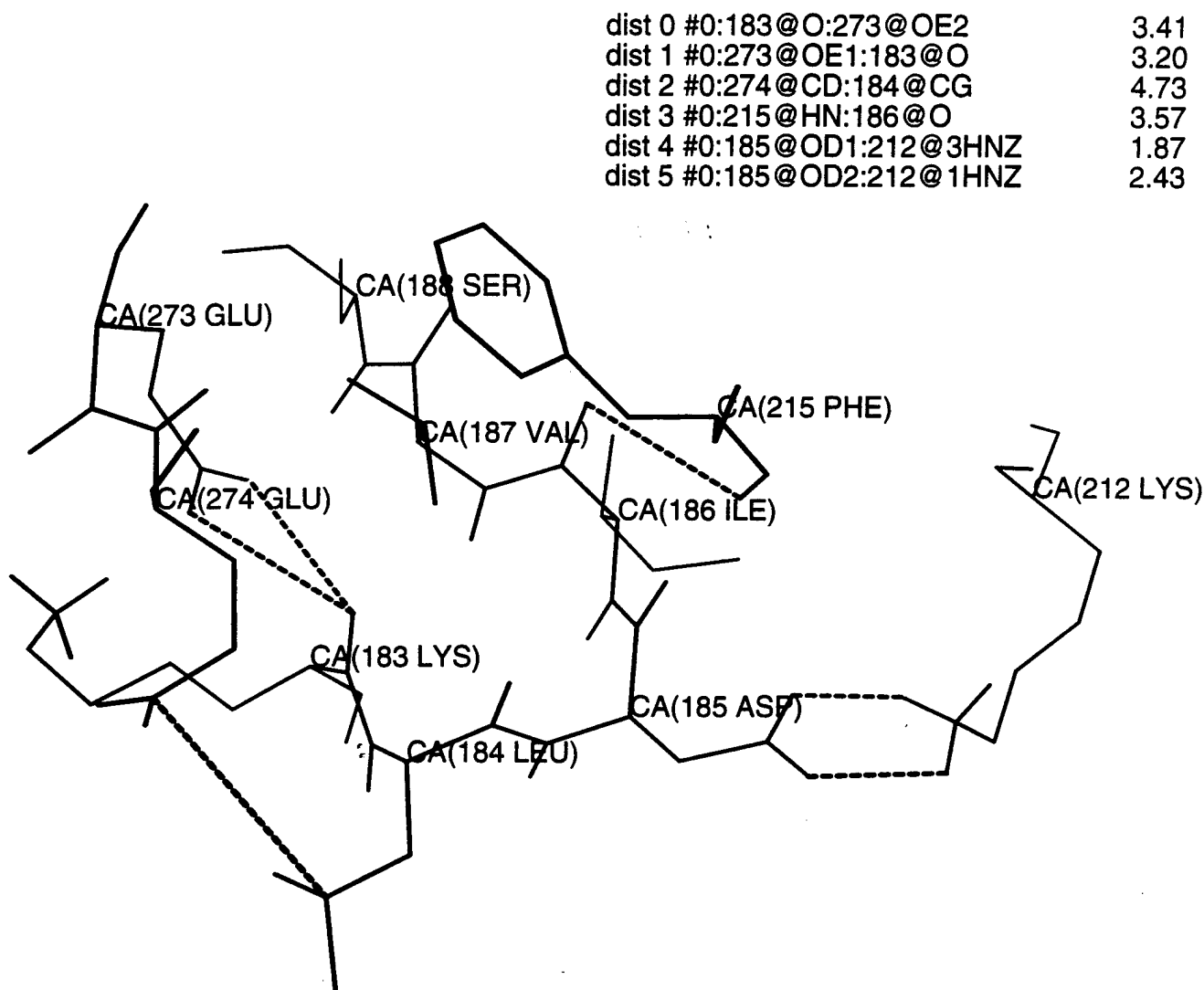


Figure 16. A part of the β -hairpin (residues 183-188). Note that K183/L184/D185 of the V2 loop are all involved in inter-domain H-bonds (see also Table 3). The C=O of L184 is H-bonded to the basic sidechain of R278 whereas the acidic sidechain of D185 is H-bonded to the basic sidechain of K212; also the sidechains of E274 and L184 are within 5 Å. Therefore, if such a β -hairpin is critical for antibody binding a double site mutation, L184/D185-D184/L185, would be catastrophic because this would bring D184 close to E274 and L185 in steric clash with K212 (both of which would destabilize the β -hairpin). Indeed, a double L184/D185-D184/L185 mutation reduces the binding affinity of antibodies specific for conformational epitopes inside the V2 loop [33-39].



UCSF MidasPlus

Conformational epitope inside the V2 loop

APPENDIX II

Structural Studies on the Human Mucin, Muc-1: effect of O-Glycosylation at Threonine Residues

Human mucins are a family of high molecular weight, heavily O-glycosylated proteins which are dominated by large tandem repeat domains. Mucin tandem repeat domains vary in size, proline content, and potential extent of O-glycosylation. Underglycosylation of the human mucin Muc-1 tandem repeat domain in certain breast, pancreatic and ovarian tumors results in the unmasking of protein core epitopes. Tumor reactive, mucin-specific monoclonal antibodies reveal differences between the surface of Muc-1 derived from tumors and normal tissues. Synthetic peptide studies show that most tumor specific antibodies recognize an epitope within the tandem repeat protein core of Muc-1. Two-dimensional nuclear magnetic resonance experiments are performed on chemically synthesized mucin tandem repeat polypeptides, (P1D2T3R4P5A6P7G8S9T10A11P12P13A14H15G16V17T18S19A20)_n for n=1,3. These studies demonstrate how the tandem repeats assemble in space giving rise to the overall tertiary structure, and the local structure and presentation of the (underlined) antigenic site (APDTR) at the junction of two neighboring repeats. The NMR data (Figure 17) reveal repeating knob-like structures connected by extended spacers. The knobs protrude away from the long-axis of Muc-1 and the predominant antigenic site (APDTR) forms the accessible tip of the knob. Multiple tandem repeats enhance the rigidity and presentation of the knob-like structures. We have enzymatically O-glycosylated the Muc-1 tandem repeat (PDTRPAPGSTAPPAHGVTSA)₃. The glycosylation occurs preferentially at the T-residues except at the T belonging to the immunodominant (APDTR) knob. Detailed NMR analyses show that the pentapeptide, (APDTR) maintains the same turn as in the unglycosylated form while additions of GalNAc at T10 and T18 in the glycosylated mucin alter the local structure (Figure 18). Also the glycosylated structure is less flexible.

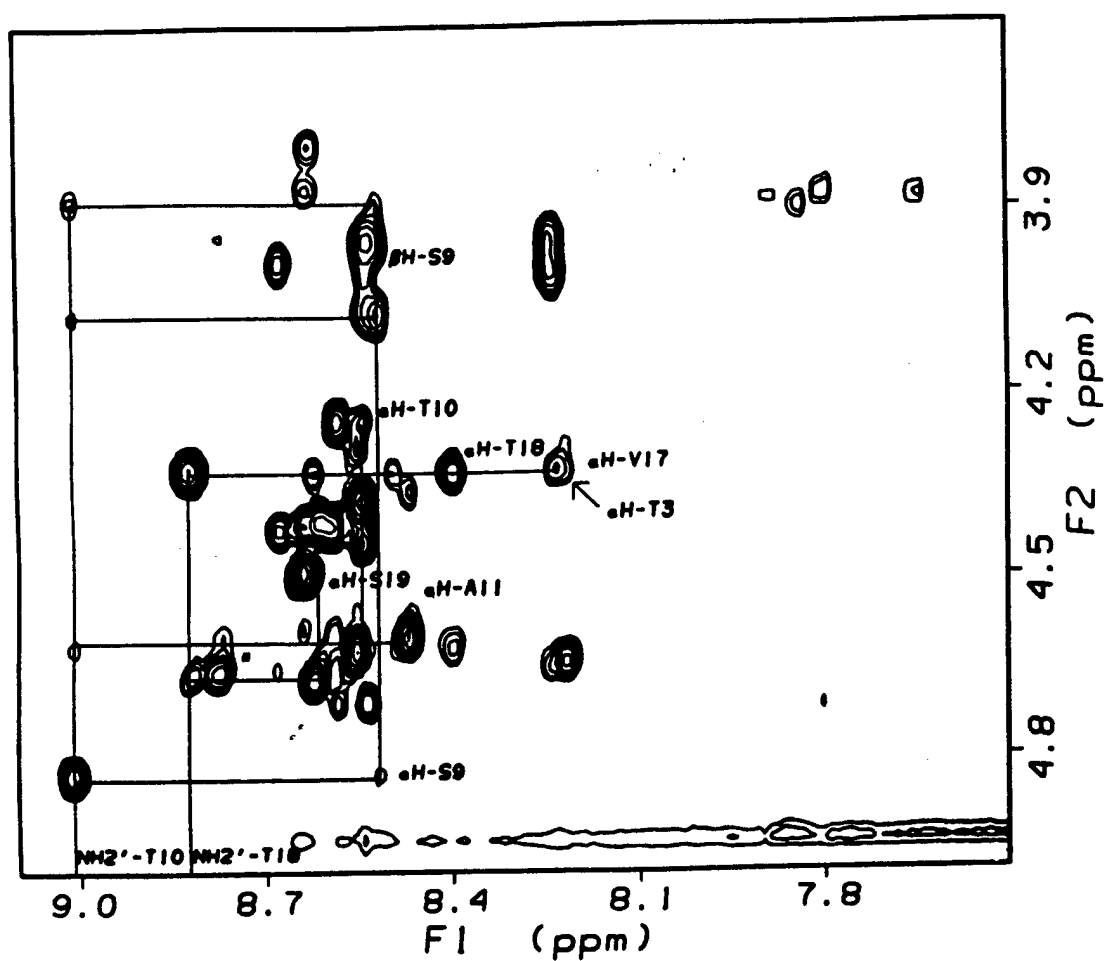


Figure 17. 2D NOESY crosssection of the O-linked GalNaC-conjugated T10 and T18 of Muc-1, (P1D2T3R4P5A6P7G8S9T10A11P12P13A14H15G16V17T18S19A20)3. Mixing time = 400 ms, temperature = 10 °C.

REFERENCES

1. Pantaleo, G., Graziosi, C., and Fauci, A.S.. The Immunopathogenesis of Human Immunodeficiency Virus Infection. (1993) *Mechanisms of Disease*, Review Article 328:327-335.
2. Sattentau, Q.J. and Weiss, R.A., The CD4 Antigen: Physiological Ligand and HIV Receptor (1988) *Cell* 52:631-633.
3. Robey, E. and Axel, R. CD4: Collaborator in immune recognition and HIV infection (1990) *Cell* 60:697-700.
4. Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N. and Gregory, T.J Assignment of Intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells (1990) *J. of Biol. chem.* 265:10-21.
5. Baggiolini, M., Dewqald, B. and Moser, Beatrice, D. and Moser, B. Interleukin-8 and related chemotatic cytokines-CXC and CC chemokines. (1994) *Adv. Immun.* 55:97-165.
6. Murphy, P.M. The molecular biology of leukocyte chemoattractant receptors. (1994) *Annu. Rev. Immun.* 12:593-633.
7. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C. M., Davis, C.B., Peiper, S.C., Schall, T.J., Littman, D.R., and Landau, N.R. Identification of major co-receptor for primary isolates of HIV-1. (1996) *Nature* 381:661-66.
8. Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P. and Paxton, W.A. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5.(1996) *Nature* 381:667-672.
9. Feng, Y., Broder, C.C., Kennedy, P.E. and Berger, E.A. HIV-1 Entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Sci.* 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. The Beta-Chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV isolates. (1996) *Cell* 85: 1135-1148.
11. Doranz, B.J., Rucker, J., Yi, Y., Smyth, R.J., Samson, M., Peiper, S.C., Parmentier, M., Collman, R.G., Doms, R.W. A dual-tropic primary HIV-1 isolate that uses fusion and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. (1996) *Cell* 85:1149-1158.
12. Zhang, L., Huang, Y., He, T., Cao, Y., and Ho, D.D. (1996) HIV-1 subtype and second receptor. *Nature* 383:768-772.
13. Thali, M., Furman, C., Ho, D.D., Robinson, J., Tilly, S., Pinter, A. and Sodroski, J. Discontinuous, Conserved Neutralization Epitopes Overlapping the CD4-Binding Region of Human Immunodeficiency Virus Type 1 gp120 Envelope Glycoprotein. (1992) *J. Virol.* , 66:5635-5641.
14. Ho, D.D., McKeating, J.A., Li, X.L., Moudgil, T. Daar, E.S., Sun, N-C. and Robinson, J.E. Conformational Epitope on gp120 Important in CD4 Binding and Human Immunodeficiency Virus Type 1 Neutralization Identified by a Human Monoclonal Antibody. (1991) 489-493.

15. Nakamura, G.R., Byrn, R., Wilkes, D.M., Fox, J.A., Hobbs, M.R., Hastings, R., Wessling, H.C., Norcross, M.A., Fendly, B.M. and Berman, P.W. Strain Specificity and binding affinity requirements of neutralizing monoclonal antibodies to the C4 domain of gp120 from human immunodeficiency virus type 1 (1993) *J. Virol.* 67:6179-6191.
16. Chamat, S., Nara, P., Berquist, L., Whalley, A., Morrow, W.J.W., Kohler H. and Kang, C-Y. Two major groups of neutralizing anti-gp120 antibodies exist in HIV-infected individuals (1992) *The J. of Immunology* 149:649-654.
17. Lasky, L.A., Nakamura, G., Smith, D.H., Fennie, C., Shimasaki, C., Eric Patzer, E., Berman, P., Timothy Gregory, T. and Capon, D.J. Delineation of a Region of the Human Immunodeficiency Virus Type 1 gp120 Glycoprotein Critical for Interaction with the CD4 Receptor. (1987) *Cell*, 50:975-985.
18. Moore, J.P., Cao, Y., Ho, D.D. and Koup R.A. Development of the anti-gp120 response during seroconversion to human immunodeficiency virus type 1 (1994) *J. Virol.* 68:5142-5155.
19. Cao, Y., Quin, L., Zhang, L., Safrit, J. and Ho, D.D. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection (1995) 332:201-208.
20. Appella, E., Gupta, G., Murphy, P.M. (personal communication).
21. LaRosa, G.J., Davide, J.P., Weinhold, K., Waterbury, J.A., Profy, A.T., Lewis, J.A., Langlois, A.J., Dreesman, G.R., Bowell, R.N., Shadduck, P., Holley, L., Karplus, M., Bolognesi, D.P., Matthews, T.J., Emini, E.A., and Putney, S.D. Conserved sequence and structural elements in the HIV-1 principal neutralizingdeterminant. (1990) *Sci.* 249:932-936.
22. Ohno, T., Terada, M., Yoneda, Y., Shea, K.W., Chambers, R.F., Storka, D, Nakamura, M., and Kufe. D.W. A Broadly Neutralizing Monoclonal Antibody that Recognizes the V3 Region of Human Immunodeficiency Virus type 1 Glycoprotein gp120. (1991) *PNAS USA* 88:10726-10729.
23. Goudsmit, J., Debouk, C., Meloen, R.H., Smith, L., Bakker, M., Asher,, Wolff, A.V., Gibbs, C.G.Jr. and Gajdusek, D.C. Human Immunodeficiency Virus Type 1 Neutralization epitope with conserved Architecture Elicits Early Type-specific Antibodies in Experimentally Infected Chimpanzees. (1988) *PNAS USA* 85:4478-4482.
24. Bou-Habib, D.C., Roderiquez, G., Oravec, T., Berman, P.W., Lusso, P.K., and Norcross, M.A. Cryptic Nature of Envelope V3 Region Epitopes Projects Primary Monocytotropic Human Immunodeficiency Virus Type 1 from Antibody Neutralization. (1994) *J. of Virol.* 68:6006-6013.
25. Gorny, M.K., Xu, J-Y., Giankakos, V., Karwowska, S., Williams, C., Sheppard, H.W., Hanson, C.V., and Zolla-Pazner, S. Production of site-selected neutralizing Human Monoclonal Antibodies Against the Third Variable Domain of the Human Immunodeficiency Virus Type 1 Envelope Glycoprotein. (1991) *PNAS* 88:3238-3242.
26. Zwart, G. Langedijk, H., van der Hoek, L., de Jong, J-J., Wolfs, T.F.M., Ramautarsing, C., Bakker, M.B., De Ronde. A., and Goudsmit. J. Immunodominance and Antigenic Variation of the Principle Neutralization Domain of HIV-1. (1991) *J. Virol.* : 481-489.

27. Javaherian, K., J. Langlois, A.J., McDanal, C. Ross, K.L., Eckler, L.I., Jellis, C.L., Profy, A.T., Rusche, J.R., Bolognesi, D.P., Putney, S.D., and Matthews, T.J. Principal Neutralizing Domain of the Human Immunodeficiency Virus Type 1 Envelope Protein. (1989) PNAS 86:6768-6772.
28. Steimer, K.S., Scandella, C.J., Skiles, P.V. Haigwood, N.L. Neutralization of Divergent HIV-1 Isolates by Conformation-Dependent Human Antibodies to Gp120. (1991) Science: 254: 105-108.
29. Gorny, M.K., Xu, J-Y., Karwowska, S., Buchbinder, A. and Zolla-Pazner. Repertoire of Neutralizing Human Monoclonal Antibodies Specific for the V3 HIV-1 gp120. (1993) J. of Immun. 150:635-643.
30. Goudsmit, J., Debouck, C., Melen, R.H., Smith, L., Bakker, M., Asher, D. M., Wolff, A.V., Gibbs, C.J. and Gajdusek, D. C. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees (1988) PNAS USA 85: 4478-4482.
31. Rusche, J.R., Javeherian, K., McDanal, C., Petro, J., Lynn, D.L., Grimaila, R., Langlois, A., Gallo, R.C., Arthur, L.O., Fischinger, P.J., Bolognesi, D.P., Putney, S.D., and Matthews, T.J. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120 (1988) 85:3198-3202.
32. Gorney, M. K., Xu, J-Y., Gianakakos, V., Karwowska, S., William C., Sheppard, H.W., Hanson, C.V. and Zolla-Pazner, S. Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein (1991) Immunology 36:3238-3242.
33. Gorny, M.K., Moore, J.P., Conley, A.J., Karwowska, S., Sodroski, J., Williams, C., Burda, S. Boots, L.J. and Zolla-Pazner, S. Human anti-V2 monoclonal antibody that neutralizes primary but not laboratory isolates of human immunodeficiency virus type 1 (1994) J. Virol. 68: 8312-8320.
34. Shotton, C., Arnold, C., Sattentau, Q., Sodroski, J. and McKeating, J.A., Identification and characterization of monoclonal antibodies specific for polymorphic antigenic determinants within the V2 region of the human immunodeficiency virus type 1 envelope glycoprotein (1995) J. Virol. 69:222-230.
35. Fung, M.S.C., Cecily, R.Y. Sun, C.R.V., Gordon, W.L., Liou, R-S., Chang, T.W., Sun, W.N.C., Daar, E.S.D. and Ho, D.D. Identification and Characterization of a Neutralization Site within the Second Variable Region of Human Immunodeficiency Virus Type 1 gp120. (1992) J. Virol. , 66:848-856.
36. Moore, J.P., Sattentau, Q.J., Yoshiyama, H., Thali, M., MacArthur, C., Sullivan, N., Poon, S-W., Fung, M.F., Traincard, F., Pinkus, M., Robey, G., Robinson, J.E., Ho, D.D. and Sodroski, J. Probing the Structure of the V2 domain of Human Immunodeficiency Virus Type 1 Surface Glycoprotein gp120 with a Panel of Eight Monoclonal Antibodies: Human Immune Response to the V1 and V2 Domains. (1993) J. of Virol. 67:6136-6151.
37. Shotton, C. Arnold, Q. Sattentau, J. Sodroski, and McKeating, J.A. Identification and Characterization of Monoclonal Antibodies Specific for Polymorphic Antigenic Determinants within the

- V2 Region of the Human Immunodeficiency Virus Type 1 Envelope Glycoprotein. (1995) *J. of Virol.* 69:222-230.
39. McKeating, J.A., Shotton, C., Cordell, J., Graham, S., Balfe, P., Sullivan, N., Charles, M., Page, M. Bolmstedt, A., Olofsson, S., Kayman, S.C., Wu, Z., Pinter A., Dean, C., Sodroski, J. and Weiss, R.A. Characterization of neutralizing monoclonal antibodies to linear and conformation-dependent epitopes within the first and second variable domains of human immunodeficiency virus type 1 gp120 (1993) *J. Virol.* 67:4932-4944.
 40. Willey, R.L., Ross, E.K., Buckler-White, A.J., Theodore, T.S. and Martin, M. Functional Interaction of Constant and Variable Domains of Human Immunodeficiency Virus Type 1 gp120. (1989) *J. of Virol.* 63:3595-3600.
 41. Andeweg, A. C., Leeftang, P., Osterhaus, A.D.M.E., and Bosc, M.L. Both the V2 and V3 Regions of the Human Immunodeficiency Virus Type 1 Surface Glycoprotein Functionally Interact with Other Envelope Regions in Syncytium Formation. *J. Virology* 67: 3232-3239.
 42. Moore, J.P, Willey, R.L., Lewis, G.K, Robinson, J. and Sdrosski, J., Immunological Evidence for Interactions between the First, Second, Fifth Conserved Domains of the gp120 Surface Glycoprotein of Human Immunodeficiency Virus Type 1. *J. Virol.* 68:6836-6847.
 43. Carrillo, A. and Ratner, L. Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells (1996) *J. Virol.* 70:1310-1316.
 44. Carrillo, A. and Ratner, L. Human immunodeficiency virus type 1 tropism for T-Lymphoid cell lines: Role of the V3 loop and C4 envelope determinants *J. Virol.* 70:1301-1309.
 45. Koito, A., Harrowe, G., Levy, J.A., and Cheng-Mayer, C., Functional Role of the V1/V2 Region of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein gp120 in Infection of Primary Macrophages and Soluble CD4 Neutralization (1994) *J. Virol.* 66:2253- 2259.
 46. Stamatatos, L. and Cheng-Mayer, C. Evidence that the Structural conformation of Envelope gp120 Affects Human Immunodeficiency Virus Type 1 Infectivity, Host Rang, and Syncytium-Forming Ability. (1993) *J. of Virol.* 5635-5639.
 47. Wang, W-K., Essex, M. and Lee, T-H. Single amino acid substitution in constant region 1 or 4 of gp120 causes the phenotype of a human immunodeficiency virus type 1 variant with mutations in hypervariable regions 1 and 2 to revert (1996) *J. Virol.* 70:607-611.
 48. Freed, E.O. and Martin, M.A. Evidence for a functional interaction between the V1/V2 and C4 domains of human immunodeficiency virus type 1 envelope glycoprotein gp120 (1994) *J. Virol.* 68:2503-2512.
 49. Stamatatos, L. and Chen-Mayer, C. Structural modulations of the envelope gp120 glycoprotein of human immunodeficiency virus type 1 upon oligomerization and differential V3 loop epitope exposure of isolates displaying distinct tropism upon virion-soluble receptor binding (1995) *J. Virol.* 69:6191-6198.

50. Wyatt, R., Thali, M., Tilley, S., Pinter, A., Posner, M., Ho, D., Robinson, J., and Sodroski, J. Relationship of the Human Immunodeficiency Virus Type gp120 Third Variable Loop to a Component of the CD4 Binding Site in the Fourth conserved Region. (1992) *J. of Virol.* 66:6997-7004.
51. Tilley, S.A., Honnen, W.J., Racho, M.E., Chou, T-C. and Pinter, A. Synergistic neutralization of HIV-1 by human monoclonal antibodies against the V3 loop and the CD4-binding site of gp120. (1992) *AIDS Res. and Human Retrovir.* 8:461-467.
52. Hwang, S.S., Boyle, T.J., Lyster, H.K., and Cullen, B.R. Identification of the envelope V3 Loop as the Primary Determinant of Cell Tropism in HIV-1 (1991) *Reports*, P 71.
53. Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. Macrophage-Tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: Definition of critical amino acids involved in cell tropism (1992) *J. Virol.* 66: 6547-6554.
54. Ebenbichler, C., Westervelt, P., Carrillo, A., Henkel, T., Johnson, D., and Ratner, L. (1993). Structure-function Relationship of the HIV-1 Envelope V3 Loop Tropism Determinant: Evidence for two distinct conformations. *AIDS* 7:639-646.
55. Hwang, S.S. Boyle, T.J., Lywely, H.K., and Cullen, B.R., Identification of the Envelope V3 Loop as the Primary Determinant of Cell Tropism in HIV-1. (1987) *Science*, 253:71-74.
56. Kim, F.M., Kolson, D.L., Valliet, J.W., Srinivasan, A., and Collman, R.G. V3-Independent determinants of macrophage tropism in primary human immunodeficiency virus type 1 isolate (1995) *J. Virology* 69:1755-1761.
57. Boyd, M.T., Simpson, G.R., Cnn, A.J., Johnson, M.A., and Weiss, R.A. A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. *J. Virology* 67:3649-3652.
58. Groenink, M., Fouchier, R.A., Broerson, Baker, C., Koot, M., vant Wout, A.B., Huisman, H.G., Miedema, F., Termette, M., and Schuitemaker, H. Envelope V2 Configuration and HIV-1 Phenotype. (1993). *Science* 260:1513-1516.
59. Cornelissen, M., Hogervorst, E., Zorgdrager, F., Hartman, S., and Goudsmit, J. Maintenance of syncytium-inducing phenotype of HIV type 1 is associated with positively charged residues in the HIV type 1 gp120 V2 domain without fixed positions, elongation, or relocated N-linked glycosylation sites. (1995) *AIDS Res. and Human Retrovir.* 11:1169-1175.
60. Page, K.A., Stearns, S.M. and Littman, D.R. Analysis of mutations in the V3 Domain of that affect fusion and ineffectivity (1992) *J. Virol.* , p 6:524-533.
61. Morris, J.F., Sternberg, E.J., Gutshall, L., Petteway, S.R., and Ivanoff, L .A. Effect of a single amino acid substitution in the V3 domain of the human immunodeficiency virus type 1: Generation of revertant viruses to overcome defects in infectivity in specific cell types (1994) *J. Virology* 8380-8385.
62. Reed, E.O., Myers, D.J., and Risser, R. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain (1991) *J. Virol.* 65:190-194.

63. Wang, W-K., Essex, M. and Lee, T-H. The highly conserved aspartic acid residue between hypervariable regions 1 and 2 of human immunodeficiency virus type 1 gp120 is important for early stages of virus replication. (1995) *J. Virol.* 69:538-542.
64. Hogervorst, E., de Jong, J., van Wijk, A., Bakker, M., Val, M., Nara, P., and Goudsmit, J. Insertion of Primary Syncytium-Inducing (SI) and NON-SI envelope V3 loops in human immunodeficiency virus type 1 (HIV-1) LAI reduces neutralization sensitivity to autologous, but not heterologous, HIV-1 antibodies (1995) *J. Virol.* 69:6342-6351.
65. Sullivan, N., Thali, M., Furman, C., HO, D.D. and Sodroski, J. Effect of amino acid changes in the V1/V2 region of the human immunodeficiency virus type 1 gp120 glycoprotein on sub-unit association, syncytium formation, and recognition by a neutralizing antibody (1993) *J. Virology* 67:3674-3679.
66. Clements, G.J., Price-Jones, M.J., Stephens, P.E., Sutton, C., Schulz, T.F., Clapham, P.R., McKeating, J.A., McClure, M.O., Thomson, S., Marsh, M., Kay, J., Weiss, R.A., and Moore, J.P. The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: A possible function in viral fusion? (1991) *AIDS Research and Human Retroviruses* 7:3.
67. Niwa, Y., Yano, M., Futaki, S., Okumura, Y., and Kido, H. (1996). T-cell Membrane-Associated Serine Protease TL2, Binds HIV-1 gp120 and cleaves the V3 Loop. *Eur. J. Biochem.* 237:64-70.
68. Ivanoff, L.A., Looney, D.J., McDanal, C., Moris, J.F., Wong-Staal, F., Langlois, A.J., Petteway, S.R. and Matthews, T.J. Alteration of HIV-1 infectivity and Neutralization by a single amino acid replacement in the V3 loop domain (1991) *AIDS Research and Human Retrovir.* 7:595.
69. Moore, J.P., Jameson, B.A., Sattentau, Q.J., Willey, R. and Sodroski, J. Towards a Structure of the HIV-1 Envelope Glycoprotein gp120: an Immunochemical Approach. (1993) *Phil. Trans. R. Soc. Lond. B.* 342:83-88.
70. Gallaher, W.R., Ball, J.M., Garry, R.F., Martin-Amedee, A.M. and Montelaro, R.C. A General Model for the surface Glycoproteins of HIV and Other Retroviruses. *Aids Research and Human Retrovir.* 11: 191-201.
71. Gabriel, J.L. and Mitchell, W.M. Proposed atomic structure of a truncated human immunodeficiency virus glycoprotein gp120 derived by molecular modeling: Target CD4 recognition and docking mechanism (1993) *PNAS USA* 90:4186-4190
72. Andreassen, H., Bohr, H., Bohr, J., Brunak, S., Bugge, T., Cotterill, .M.J., Jacobsen, C., Kusk, P., Lautrup, B., Petersen, S. B., Saermark, T. and Ulrich, K. Analysis of the secondary structure of the human immunodeficiency virus (HIV) proteins p17, gp120, and gp41 by computer modeling based on Neural Network Methods (1990) *J. of AIDS* 3:615-622 .
73. Pollard, S. R., Rosa, M.D., Rosa, J.J. and Wiley, D.C. Truncated Variants of gp120 bind CD4 with High Affinity and Suggest a CD4 Binding Region (1992). *EMBO J.* 11:585-591.

74. Vu, H.M., de Lorimier, R., Moody, M.A., Haynes, B.F. and Spicer, L.D. Conformational Preference of Chimeric Peptide from the C4-V3 Domains of gp120 Envelope Protein of HIV-1 CAN0A Based on Solution NMR (1996). *Biochemistry* 35:5158-5165.
75. de Lorimier, R., Moody, M.A., Haynes, B.F. and Spicer, L.D. NMR-Derived solution conformations of a hybrid synthetic peptide containing multiple epitopes of envelope protein gp120 from the RF strain of human immunodeficiency virus (1994) *Biochemistry* 33:2055-2062.
76. Robey, F.A., Kelson-Harris, T., Robert-Guroff, M. Batinic, D., Ivanov, B., Lewis, M.S. and Roller, P. P. A synthetic Conformational Epitope from the C4 Domain of gp120 that binds CD4 (1996). *J. Biol. Chem.* 271:17990-17995.
77. Gupta, G. and Myers, G. Computer Analyses of HIV Epitope Sequences. 1990. Proc. of the 5th Colloque des "Cent Gardes," Paris, Oct 29-31, 1990 (M. Girard and Valette, eds.), Paris: Pasteur Vaccins, 1990, pp. 99-104.
78. Gupta, G. and Myers, G. Analyses of the Various Folding Patterns of the HIV V3 Loops. In *Peptides: Design, Synthesis, and Biological Activity* (C. Basava and G. M. Anantharamiah, eds.), New York: Birkhauser (1994), Ch 17:259-277
79. Catasti, P., Fontenot, J.D., Bradbury, E.M., and Gupta, G. Local and Global Structural Properties of the HIV-MN V3 Loop. (1995) *J. of Biol. Chem.* 270:2224-2232.
80. Catasti, P., Bradbury, E.M., and Gupta, G. Structure and Polymorphism of the HIV-1 V3 Loops. (1995) *J. of Biol. Chem.* 271:8236-8242.
81. Catasti, P., Fontenot, J.D., Mariappan, S.V.S., Bradbury, E.M., and Gupta, G. Pau, C-P., Parekh, B.S., George, J.R. and Gupta, G. Human Immunodeficiency Virus (HIV) Antigens: Structure and Serology of Multivalent Human Mucin MUC1-HIV V3 Chimerical Proteins. (1995) *Proc. Natl. Acad. Sci. USA* 92:315-319.
82. Fontenot, J.D., Gatewood, J.M., Mariappan, S.V.S., Catasti, P., Domenech, P., Finn, O.J., and Gupta, G. Structure of a Tumor Associated Antigen Containing Tandemly Repeated Immunodominant Epitope. (1995) *J. Biomol. Str. & Dyn.* 13-245-260.
83. Gutpa, G., Anantharamaiah, G.M., Scott, D.R., Eldrige, J.H. and Myers, G. Solution Structure of the V3 Loop of a Thailand HIV Isolate. (1993) *J. Biomol. Struc. & Dyn.* 11:345-366.
84. Vernose, F.D.M., Reitz, M., Guiraff, M.R., Thompson, C.B., Gallo, R.C., Gupta, G. and Russo, P. and P. Russo. Generation of a Neutralizing Epitope by a Spontaneous Mutation in the V3 Loop of HIV-1(IIIB) Isolated From an Infected Laboratory Worker. (1993). *J. Biol. Chem.* 268:25894-25901.
85. Vernose, F.D.M., Reitz, M., Gupta, G., Guiraff, M.R., Thompson, C.B., Gallo, R.C., and Russo, P. and P. Russo. Loss of a Neutralizing Escape Mutant by a Local Change in the HIV-1 V3 Loop Conformation. (1994). *Vaccines* 94.
86. Catasti, P., Bradbury, E.M., and Gupta, G. A Stereochemical Modeling of the HIV-1 Surface Envelope Protein, gp120. (1995). *J. Virol.* to be submitted.

87. Fontenot, J.D., Van Cott, T.C., Parekh, B.S., Pau, C-P., George, J.R., Birx, D.R., Zolla-Pazner, S., Gorny, M.K., and Gatewood, J.M. Presentation of HIV V3 Loop Epitopes for Enhanced Antigenicity, Immunogenicity, and Diagnostic Potential. (1995). *AIDS* 9: 1121-1129.
88. Rini, J.M., Stanfield, R.L. Stura, E.A., Salinas, P. A., Profy, A.T. and Wilson. I.A.. Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop. (1993) *Proc Natl. Acad. Sci. USA* 90: 6325-6329.
89. Ghiara, J.B., Stura, E.A., Stanfield, R.L., Profy, A.T., Wilson, I.A.. Crystal Structure of the Principal Neutralization Site of HIV-1. (1994) *Science* 264: 82-85.
90. G. Ramachandran and V. Sasisekharan. Conformation of Polypeptide and Proteins. 1986. *Advances in Protein Chemistry* 23:284-438.
91. Reed, J., and Kinzek, V., A conformational switch is associated with receptor affinity in peptides derived from the CD4-Binding domain of gp120 from HIV I (1991) *Biochemistry* 30: 4521-4528.
92. Reed, J. and Kinzel, V. Primary structure elements responsible for the conformational switch in the envelope glycoprotein gp120 from human immunodeficiency virus type 1: LPCR is a motif governing folding *PNAS USA* (1993) 90:6761-6765.
93. Bush, C.A. Yan, Z-Y. and Rao, B.N. (1986), *J. Am. Chem. Soc.*, 108, 6168-6172.
94. Poolex-Kruger, A., Myer, B., Stuike-Prill, R., Sinnwell, V., Matta, K.L. and Brockhausen, I. (1993), 100, 365-380.

Solution Structure Of The V3 Loop Of A *Thailand* HIV Isolate

**Goutam Gupta¹, G.M. Anantharamaiah², David R. Scott³,
John H. Eldridge⁴ and Gerald Myers¹**

¹Theoretical Biology and Biophysics Group
Los Alamos National Laboratory, T-10, MS K710
Los Alamos, NM 87545

²Department of Medicine
University of Alabama at Birmingham Medical Center
Birmingham, AL 35294

³NMR Instrument Facility, Department of Chemistry
Iowa State University of Science and Technology
Ames, IO 50011

⁴Department of Medicine and Microbiology
University of Alabama at Birmingham Medical Center
Birmingham, AL 35294

Abstract

The principal neutralizing determinant (PND) of human immunodeficiency virus (HIV) is located inside the third variable loop (designated the V3 loop) of the envelope glycoprotein *gp120*. The V3 loop is typically 35 amino-acids long, and the 1st and the 35th residues in the loop are invariant cystines involved in a disulfide-bridge. Although PNDs from different HIV isolates contain a conserved GPG-sequence, the amino acids flanking the conserved sequence show hypervariability among HIV isolates; the GPG and the two flanking regions are collectively referred to as the GPG-crest or the PND. The amino acid sequence variability in the GPG-crest gives rise to different antigenic specificities for different PNDs from different HIV isolates. By combining two-dimensional nuclear magnetic resonance (2D NMR) and molecular modeling techniques, we have developed a method to study (1) the global tertiary fold of the V3 loops of HIV and (2) the local structure of the PND at the tip of the V3 loop. In this article, we report the results of our structural studies on the V3 loop of a *Thailand* HIV isolate. The sequential assignment is made by combining DQF-COSY, TOCSY, and NOESY/ROESY experiments. Various intra- and inter-residue inter-proton distances are estimated by full-matrix analyses of the NOESY data at 100 and 400 ms of mixing times and of the ROESY data at 60 and 200 ms of mixing times. 100 inter-residue distances are used

Abbreviations: Human Immunodeficiency Virus (HIV), Principal Neutralizing Determinant (PND), Correlation Spectroscopy (COSY), Double Quantum Filtered COSY (DQF-COSY), Total COSY (TOCSY), Nuclear Overhauser and Exchange Spectroscopy (NOESY), Rotating Frame Nuclear Overhauser and Exchange Spectroscopy (ROESY).

as structural constraints in a simulated annealing procedure to derive energetically stable structures. Two functional motifs in the V3 loop, i.e., the glycosylation site and the GPG-crest, form defined structures: a turn is located at the glycosylation site, and the GPG-crest forms a protruding domain with a type-II GPGQ turn. The other regions of the V3 loop are rather flexible—especially the C-terminal DIRKAYC-stretch. These flexible regions of the V3 loop lead to conformational flexure of the entire V3 loop without altering the local structures of the glycosylation site or the GPG-crest. However, the ROESY experiments revealed no slow exchange among different V3 loop conformations, and therefore the flexible conformations are in fast exchange within the NMR time scale. The extent of this conformational flexibility is also discussed.

Introduction

Studies on the feasibility of a subunit vaccine to protect against HIV-1 infection have mainly focused on the outer envelope glycoprotein, *gp120* [1]. The PND is located inside the V3 loop of *gp120* [2]. Antibodies elicited by the PND block virus infectivity, thus neutralizing the virus [3-5]. Neutralizing antibodies can also block viral infection by inhibiting fusion of HIV-infected cells with CD4-positive uninfected cells [6]. The role of the PND in virus neutralization and inhibition of cell fusion has made them the focus of intense research in the development of vaccine directed against HIV infection. However, progress in vaccine development has been impeded by amino acid sequence variability among different HIV isolates, particularly in the V3 loop [7]. Neutralizing antibodies elicited by the PND from one HIV isolate do not neutralize other HIV isolates [8]. Sequence and structure analyses of the V3 loops from a large number of HIV isolates are therefore required to accurately define global and local structural differences among different V3 loop sequences. Because an antibody recognizes a specific three-dimensional structure of the antigen [9], performing the above analyses is central to understanding the specific interactions that govern PND-antibody complex formation. Such studies are also relevant because the (S-S)-bridged V3 loop is the smallest part of *gp120* that is likely to present the PND to the antibody in a manner similar to the entire envelope protein. A systematic study of several different V3 loop sequences will allow us to map the amino acid sequence variability in terms of the global structure of the entire V3 loop and the local structures of various functional motifs located inside the V3 loop, i. e., the glycosylation site and the GPG-crest. In this article we outline the results of our studies on the Thailand TN243 viral V3 loop sequence (Figure 1) combining 2D NMR and molecular modeling techniques. This V3 loop sequence which is quite different from the viral V3 loop sequence found in North America [8], may account for the virulence and transmissibility of this viral strain [25,26].

Materials and Methods

Chemical Synthesis and purification

The V3 loop sequence was reported by McCutchan *et al.* [25]. First, the 36-amino-acids-long linear peptide was synthesized using Merrifield's solid phase synthesis protocol. Two cystine residues (i.e., C2 and C36) involved in the (S-S)-bridge were protected using a methyl benzyl group (MeBzl), and the terminal C1 was protected by using an acetamide-methyl group (ACM). The MeBzl groups on C2 and C36 were removed by using anhydrous hydrogen fluoride (HF) while the terminal C1 was still blocked. A pro-

longed oxidation experiment was carried out for the formation of the (S-S)-bridge between C2 and C36. After the completion of the oxidation step, the cyclic V3 loop was purified by analytical HPLC.

NMR

All NMR experiments are performed using a 500 MHz (Varian Unity) NMR spectrometer.

The following steps are performed for the analysis of the NMR data.

(i) Sequential Assignment

At first, fingerprint HN-H^α correlations are obtained from the corresponding TOCSY and DQF-COSY cross-sections of the V3 loop in 90%₂H₂O+10%₂D₂O. The sidechains are identified from HN-H^α-H^β & H^α-H^β-H^γ-H^δ correlations in the TOCSY spectra and H^α-H^β, H^β-H^γ, & H^γ-H^δ correlations in the DQF-COSY spectra of the V3 loop in 90%₂H₂O+10%₂D₂O and in 100%₂D₂O. Once isolated spin-systems belonging to constituent amino acids are identified the sequential connectivity is obtained by monitoring the NOESY/ROESY cross-peaks for H^α(i)-HN(i+1), H^β(i)-HN(i+1), & HN(i)-HN(i+1) interactions. NOESY data are collected for mixing times of 100 and 400 ms. ROESY data are collected for mixing times of 60 and 200 ms.

(ii) Extraction of Inter-proton Distances as Structural Constraints

This step involves obtaining an energetically stable (S-S)-bridged structure for the Thailand TN243 loop sequence given the secondary structural states of the constituent amino acids residues as obtained by analyzing the sequential NOESY and ROESY pattern. Appropriate ranges of (φ, ψ) values are assigned to all amino acids. For example,

$$\phi = -55^\circ \pm 25^\circ, \quad \psi = -55^\circ \pm 25^\circ \text{ for residues in a helix;}$$

$$\phi = -140^\circ \pm 50^\circ, \quad \psi = 140^\circ \pm 50^\circ \text{ for residues in a beta strand or in an extended conformation;}$$

$$\begin{aligned} \phi_{i+1} &= -65^\circ \pm 20^\circ, & \psi_{i+1} &= -50^\circ \pm 20^\circ \\ \phi_{i+2} &= -90^\circ \pm 20^\circ, & \psi_{i+2} &= 0^\circ \pm 20^\circ \end{aligned} \text{ for residues in a type-I turn;}$$

$$\begin{aligned} \phi_{i+1} &= -65^\circ \pm 20^\circ, & \psi_{i+1} &= 120^\circ \pm 20^\circ \\ \phi_{i+2} &= 90^\circ \pm 20^\circ, & \psi_{i+2} &= 0^\circ \pm 20^\circ \end{aligned} \text{ for residues in a type-II turn.}$$

(φ, ψ) of residues in the coil state are set free to choose any point in the allowed space [for definitions of different secondary structures and corresponding (φ, ψ)-values, reference 10]. First we obtain (S-S)-bridged structure for a pseudo (CA₃₃C)-sequence in the following manner. All residues with sidechains extending beyond C^β atoms are treated as A, except the Ps & Gs and the terminal Cs. Our rationale for doing this is that the allowed (φ, ψ) space of residues with a sidechain longer than A is only a subspace of that allowed for A [10]. We obtain an (S-S)-bridged structure of a V3 loop by a linked-atom-least-square refinement [11] by minimizing a function, F, only in the (φ, ψ) space.

$$F = \text{R-Factor} + \sum_l \lambda_l G_l + \sum_{ij} (d_{ij}^{mn} - D^{mn})^2 \dots \dots \dots [1]$$

$$\text{R-Factor} = \frac{\sum ||o - c|}{\sum |o|}$$

|o = observed NOESY intensity and |c =calculated NOESY intensity by full-matrix NOESY simulations. The sum extends over all pairs (i,j) of observed NOESY cross-peaks.

$G_l (=|r_l - r_l^o| = 0)$ indicates distance constraints for an (S-S) bridge. Distances in the (S-S)-bridged V3 loop configuration are defined as $r_1 = \text{S(C1)-S(35)}$, $r_2 = \text{C}^\beta(\text{C1})\text{-S(C35)}$, $r_3 = \text{C}^\beta(\text{C35})\text{-S(C1)}$ and $r_4 = \text{C}^\beta(\text{C1})\text{-C}^\beta(\text{C35})$; corresponding equilibrium distances are $r_1^o = 2.04\text{\AA}$, $r_2^o = r_3^o = 3.05\text{\AA}$, $r_4^o = 3.85\text{\AA}$ [12,13]. λ_l indicates Lagrangian multipliers; d_{ij}^{mn} indicates the distance between atom i (type m) and atom j (type n); and D^{mn} indicates the contact limit between atom (type m) and atom (type n) [10]. In this refinement, the (ϕ, ψ) values of various residues are treated as elastic variables (i.e., variables with weights) such that by appropriate choice of weights the experimentally determined secondary structural states of residues are minimally altered [14]. This method guarantees a stereochemically orthodox structure for the (S-S)-bridged (CA_{33}C)-like sequence. Finally appropriate sidechains are attached to generate an actual V3 loop sequence and the potential energy of the system is minimized in the (ϕ, ψ, ω, χ)-space using the force-field of Scheraga and co-workers [12]. The minimization of the function, F, for the virtual (S-S)-bridged (CA_{33}C) system followed by the energy minimization for the actual V3 loop is repeated for 50 different starting structures (Fletcher, 1984). At the end of this step we obtain a set of 50 models in agreement with the NOESY and ROESY data. From these models, a set of inter-proton distances are extracted as structural constraints required for agreement with the NOESY/ROESY data. Each pairwise distance representing a structural constraint provides an upper and a lower limit of the distance. Two types of constraints are identified.

Type I

This is given as

$$\begin{aligned} \text{EDIST} &= 0 \text{ if the distance } r \text{ is within a specified range } r_1 \text{ \& } r_2 \\ &= k(r-r_1)^2 \text{ if } r < r_1 \\ &= k(r-r_2)^2 \text{ if } r > r_2. \text{ k : force constant.} \end{aligned}$$

Type II

This is given as

$$\begin{aligned} \text{EDIST} &= 0 \text{ if } r \geq r_1 \\ &= k(r-r_1)^2 \text{ if } r < r_1. \end{aligned}$$

This type is particularly useful for an unobserved NOE where we can set a lowest allowable distance limit for the corresponding proton pair.

(iii) Exploration of the Conformational Flexibility Subject to the NMR Data

The energy term, EDIST, is added to the force-field as in Scheraga and co-workers [12; QCEP 454]. The simulated annealing is performed [15] in the following manner.

First, a starting energy-minimized structure is chosen and Monte Carlo (MC) simulations are performed for 50,000 steps at 600K in the $(\phi, \psi, \omega, \chi)$ -space; the last accepted configuration is stored to be subsequently used as a starting configuration in the next lower temperature-cycle. Second, 50,000 MC steps are repeated in several cycles of gradually decreasing temperature until a temperature of 100K is reached. Third, the lowest energy configuration at 100K is further energy-minimized to a low energy gradient. This is the "temperature quenching" step in which thermally excited single bond rotations around the equilibrium positions are quenched. Finally, first through third steps are repeated for 100 different starting configurations which comprise of the 50 starting structures obtained after the Step (ii) discussed in the previous paragraph and 50 other structures as obtained after carrying out first through third steps of the simulated annealing protocol on the very same 50 starting structures as discussed in the previous paragraph.

(iv) Analyses of the V3 Loop Structures

Low energy structures are analyzed in terms of their $(\phi, \psi, \omega, \chi)$, agreement with the NMR data and agreement with the ϕ -values as estimated from the HN-H^α J-coupling data obtained from the DQF-COSY experiments. For $J_{\text{HN-H}^\alpha} > 7$ Hz the stipulated constraint on ϕ is $110^\circ \pm 40^\circ$. However, no ϕ -constraint is imposed during simulated annealing; but in the final model if ϕ falls far outside the expected range (as determined from the DQF-COSY data in Figures 3 & 7), then the corresponding model is discarded.

Results

NMR Data

Figure 1 shows the amino acid sequence of the *Chiang Mai* V3 loop sequence. The N-terminal modified C1 is added for facilitating conjugation in antibody production experiments. Note that C2 and C36 are involved in (S-S) bridge formation. A total of 35 HN protons are present in the system, i.e., excluding two Ps in the sequence and including the terminal blocking group at C1. Figure 2 shows the TOCSY HN-H^α fingerprint region of the Thailand TN243 V3 loop at 10°C . All the spin systems are observed; the only exception is the HN of the blocking group. The sequential assignment of all the cross-peaks are indicated in Figure 2. The amino acid residues that occur only once in the sequence are easily assigned, e.g., Q19, V20, F21, K33, & A34. The spin systems of G and S are distinguished by examining the DQF-COSY $\text{H}^{\alpha 1}$ - $\text{H}^{\alpha 2}$ & H^α - H^β cross-peak patterns. The spin system of the branched sidechains of T and I are identified by monitoring the H^α - H^β - H^γ TOCSY and DQF-COSY coupling patterns. The spin system of R is identified by monitoring the H^α - H^β - H^γ J-coupling pattern and H^γ - H^δ -NH NOE pattern. The spin system of Y is assigned by monitoring the H^α - H^β & 2,6H-3,5H J-coupling patterns and the H^β -Hring NOE pattern. The spin systems of D and N are distinguished by the presence of the NH2 group. Figure 3 displays the fingerprint region of a DQF-COSY spectrum of the Thailand TN243 V3 loop at 10°C showing the HN-H^α cross-peaks. Positive and negative contour are plotted without distinction. Here the shapes of the peaks are distorted because both sine-bell and shifted sine-bell apodization functions are used to sharpen the peaks.

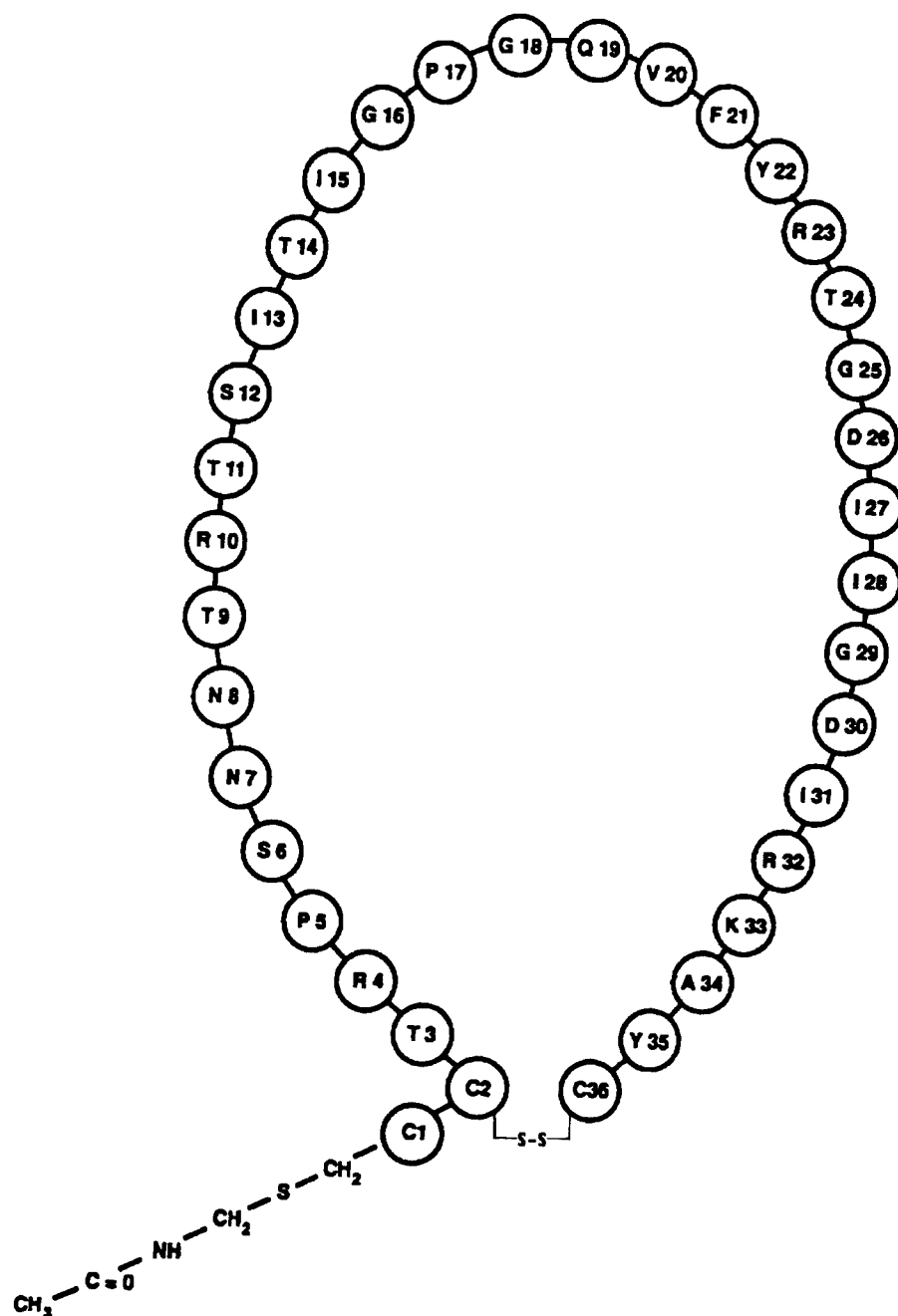
Chiang Mai U3 Loop

Figure 1: The amino acid (aa) sequence of the Thailand TN243 V3 loop. C2 and C36 form a (S-S) bridge. A modified C1 is present at the N-terminal. The protecting group on the N-terminal is removed before conjugation to BSA- a step involved in antibody production. Amino acid containing the GPG-crest and the two flanking regions (extending up to 4-5 aa on each side) define the immunogenic tip of the V3 loop. N7 (within the recognition element N7-N8-T9) is the site of N-linked glycosylation. Although epitope mapping for this particular V3 loop is not reported, amino acids within the aa13-27 stretch of the V3 loop generally make contact with the antibody [22,23].

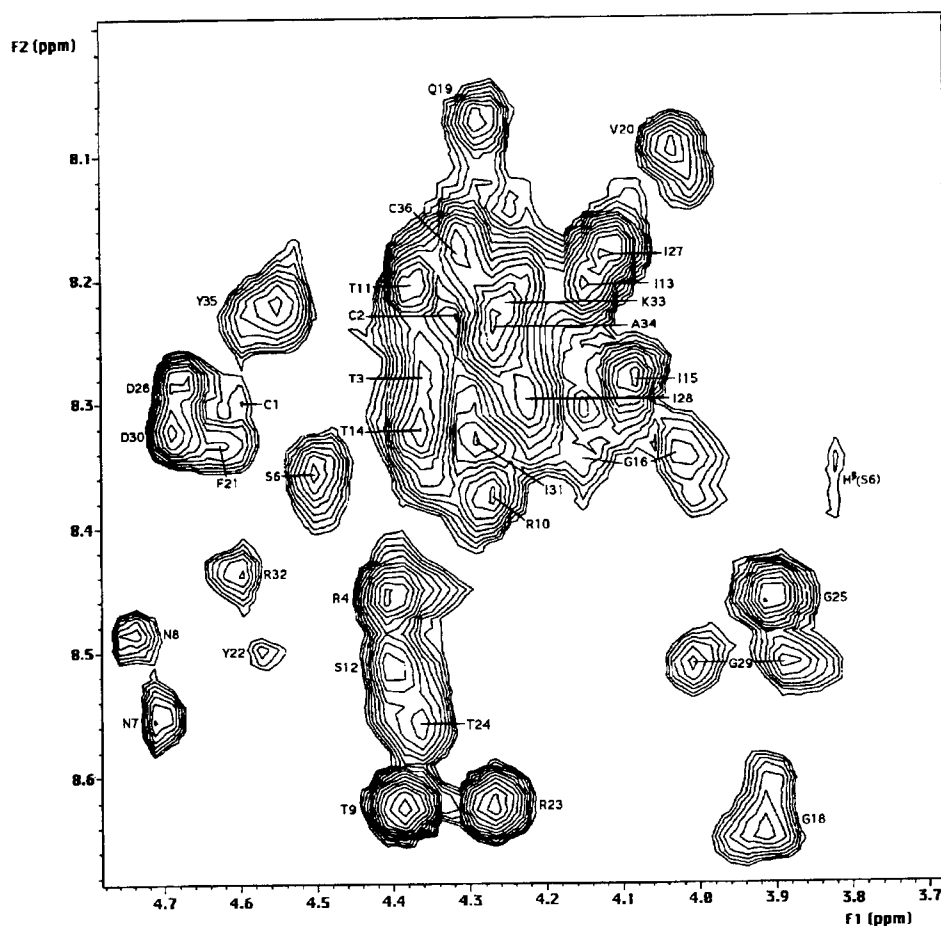


Figure 2: HN-H α fingerprint TOCSY cross-section. The TOCSY experiment is performed under the following conditions: temperature = 10 °C, pH = 4.5, polypeptide concentration = 8 mM, solvent composition = 90% H_2O +10% D_2O . The acquisition parameters: data points in t_2 = 2048; complex data points in t_1 = 256; relaxation delay; RD = 1.5 s; mixing time = 35 ms. The HDO signal is suppressed by employing the pulse sequence of Sklenar and Bax [15]. Several contours are drawn starting from the lower level of intensity in order to show all the (HN-H α) crosspeaks.

However, for accurate measurement of the $J(\text{H}^\alpha\text{-HN})$ couplings (see Figure 7 below), appropriate apodization functions and line-width corrections are used to make the two extrema as symmetric to each other as possible [17]. Out of the 34 possible (HN,H α) spin systems, $J(\text{HN-H}^\alpha)$ coupling data are obtained for 27 residues. No DQF-COSY HN-H α cross-peaks are obtained for C1, G18, & G29. Overlap of the HN-H α cross-peaks for K33 & A34 precluded measurements of the corresponding J couplings. The HN-H α cross-peak of T3 is too weak and poorly defined to allow measurement of J coupling. A weak HN-H α cross-peak is observed for only one of the G16 H α protons.

Even though isolated spin systems of all the amino acids are identified from the TOCSY and the DQF-COSY experiments, the sequential assignment is possible only after examining the inter-residue H $^\alpha(i)$ -HN($i+1$) and H $^\beta(i)$ -HN($i+1$) NOESY connectivities as shown in Figures 4 and 5, respectively.

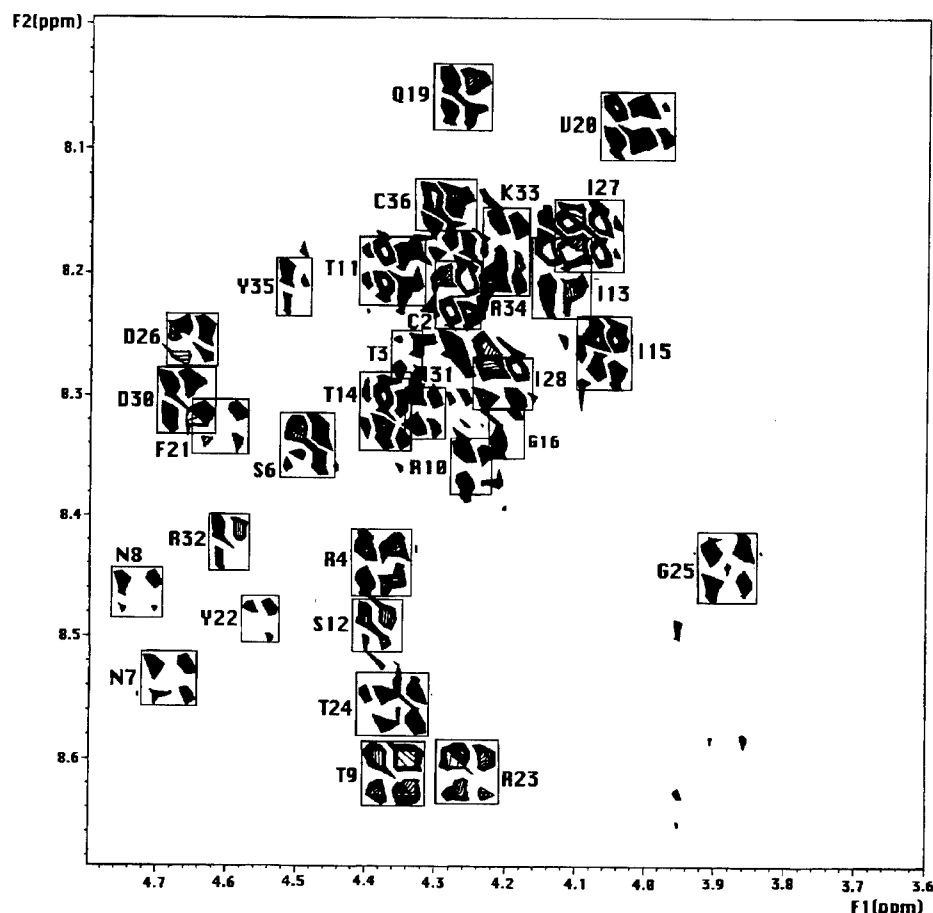


Figure 3: HN- H^{α} fingerprint DQF-COSY cross-section. The DQF-COSY experiment is performed under the following conditions: temperature = 10 °C, pH = 4.5, polypeptide concentration = 8 mM, solvent composition = 90% H_2O +10% D_2O . The acquisition parameters: data points in t_2 = 2048; t_1 increments = 256; relaxation delay; RD = 1.5 s. The HDO signal is presaturated for 1 s during the relaxation delay. Fourier transformation is performed on a (2048X1024) data matrix with a combination of gaussian, sin-bell, and shifted sine-bell apodization functions.

Figure 4 shows the H^{α} -HN NOESY cross-section (mixing time = 400 ms) of the Thailand TN243 V3 loop at 10 °C. Note that a NOESY cross-peak corresponding to H^{α} -HN interaction of the N-term is observed in this cross-section. A few intra- and inter-residue H^{β} -HN interactions are also observed in this cross-section. The NOE pattern is consistent with the presence of three turns located at R4-P5-S6-N7, G16-P17-G18-Q19, and T24-G25-D26-I27. The sequential H^{α} -HN NOEs and distinctive long range NOEs for HN(C2)- H^{α} (S6) & HN(C2)- H^{β} (S6) support the presence of a type-I turn at R4-P5-S6-N7. The sequential H^{α} (P17)-HN(G18) overlaps with the intra-residue H^{α} -HN of T9; however, other sequential H^{α} -HN in the G16-P17-G18-Q19 is consistent with a type-II turn with a single H-bond between C=O(G16) & HN(Q19). In view of the fact that a strong H^{α} (P17)-HN(G18) NOE distinguishes a type-II from a type-I turn [18], the overlap of this cross-peak in the spectrum (Figure 4) prevents us from unequivocally suggesting a G16-P17-G18-Q19 type-II turn. However, in other V3 loop sequences with GPGR-sequence, a type-II turn is indicated

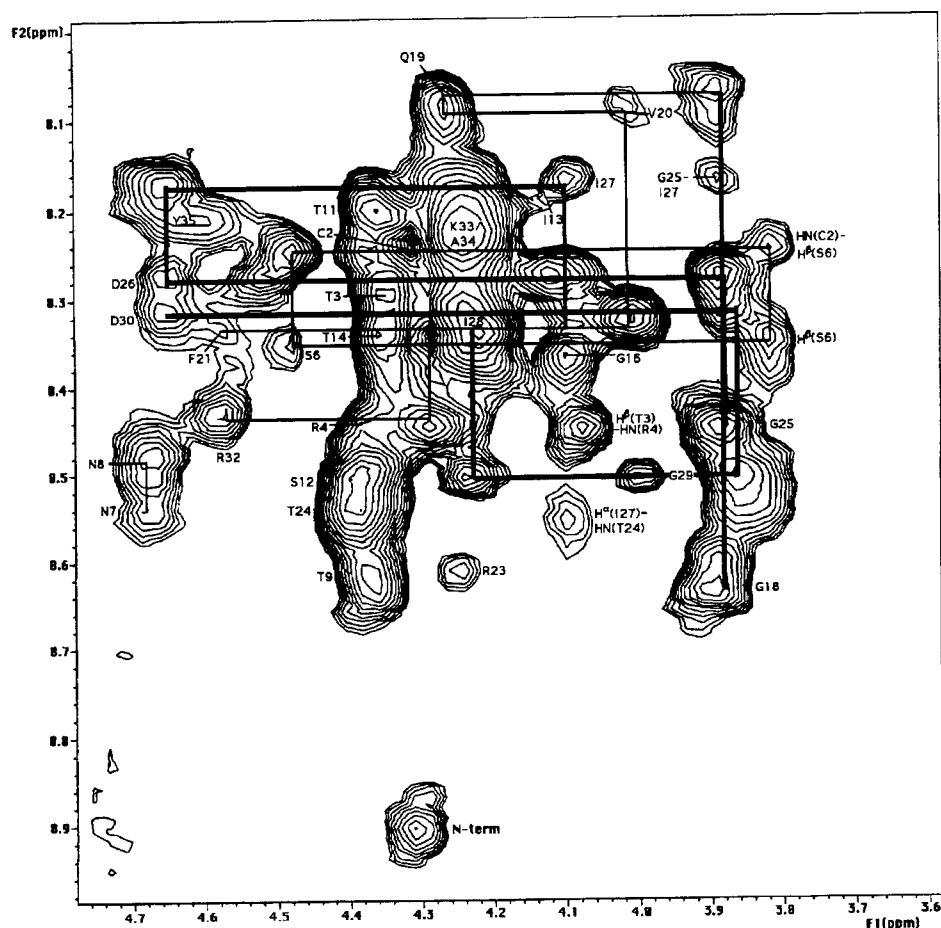


Figure 4: HN-H α NOESY cross-section. The NOESY experiments are performed under the following conditions: temperature = 10 °C; pH = 4.5, polypeptide concentration = 8 mM; solvent composition = 90% H₂O + 10% D₂O. The acquisition parameters: data points in t₂ = 2048, complex data points in t₁ = 256; relaxation delay, RD = 1.5 s; mixing time = 400 ms. The HDO signal is suppressed by presaturation (for 1 s) during the relaxation delay.

[19,20]. In addition, simulated annealing of the GPGQ-fragment shows the energetic preference of type-II over type-I turn. Therefore, we considered a type-II turn for the GPGQ-sequence of the *Chiang Mai* V3 loop. The sequential H α -HN NOE pattern and weak H α (G25)-HN(27) and HN(T24)-H α (I27) NOEs suggest a type-I turn for the T24-G25-D26-I27 sequence. An extended conformation (which includes the single β -strand) is identified for the I28-G29-D30-I31 sequence because of the presence of mostly strong sequential H α -HN NOESY cross-peaks in Figure 4. An extended conformation is also located at the T11-S12-I13-T14-I15 sequence; residues in this segment show sequential H α -HN NOESY cross-peaks of medium intensity (see Figure 7 below for the summary of NOE and ROE data). The centers of H α -HN NOESY cross-peaks for the T11-S12-I13-T14-I15 sequence are not clearly defined in Figure 4. However, the centers of H α -HN NOESY cross-peaks for T11-S12 & S12-I13 are distinguished by examining the NOESY cross-section at 100 ms mixing time

Table I

Chemical shift values in ppm. The assignment is obtained by analyzing TOCSY, DQF-COSY, & NOESY data of the Thailand TN243 V3 loop in 90% H_2O +10% D_2O and TOCSY data in 100% D_2O . TSP is used as an internal standard. Note that the complete spin system of R4, R10 and R23 beyond $\gamma CH/\beta CH$ could not be assigned. The sidechain NHs of these residues appear as broad peaks at (6.52,6.81 ppm).

Residue	HN	H ^a	H ^b	Other
C1	8.29	4.59	3.12,3.10	
C2	8.23	4.31	3.19,3.00	
T3	8.28	4.36	4.24	γCH_3 1.18
R4	8.46	4.40	1.82,1.82	γCH_2 1.59
P5	-	4.40	2.20,2.20	γCH_2 1.96,1.96; δCH_2 3.75,3.56
S6	8.35	4.50	3.83,3.83	
N7	8.55	4.70	2.78,2.78	NH 6.94,7.62
N8	8.49	4.73	2.82,2.76	NH 6.96,7.64
T9	8.63	4.39	4.24	γCH_3 1.18
R10	8.37	4.28	1.92,1.92	γCH_2 1.62
T11	8.20	4.37	4.28	γCH_3 1.18
S12	8.51	4.40	3.89,3.84	
I13	8.20	4.14	1.82	γCH_2 1.43,1.20; γ & δCH_3 0.85
T14	8.32	4.36	4.11	γCH_3 1.12
I15	8.28	4.08	1.80	γCH_2 1.43,1.20; γ & δCH_3 0.85
G16	8.34	4.16,4.03		
P17	-	4.41	2.24,2.24	γCH_2 2.00,2.00; δCH_2 3.75,3.56
G18	8.64	3.92,3.94		
Q19	8.07	4.28	2.12,1.90	
V20	8.10	4.02	1.81	γCH_3 0.81,0.70
F21	8.33	4.62	2.88,2.88	2,6H 7.14; 3,5H 7.24
Y22	8.50	4.57	2.93,2.97	2,6 H 6.78; 3,5,4H 7.09
R23	8.63	4.26	1.98,1.98	
T24	8.56	4.37	4.24	γCH_3 1.18
G25	8.46	3.92,3.92		
D26	8.28	4.68	2.82,2.79	
I27	8.18	4.12	1.85	γCH_2 1.43,1.20; γ & δCH_3 0.85
I28	8.30	4.22	1.63	γCH_2 1.43,1.20; γ & δCH_3 0.85
G29	8.51	4.01,3.89		
D30	8.32	4.69	2.84,2.76	
I31	8.33	4.28	1.82	γCH_2 1.43,1.20; γ & δCH_3 0.85
R32	8.43	4.59	1.84	γCH_2 1.62; δCH_2 3.10; NH 7.24,7.30
K33	8.22	4.24	1.84,1.84	γCH_2 1.36; δCH_2 1.65; ϵCH_2 2.95; NH 7.54
A34	8.24	4.26	1.26	
Y35	8.22	4.31	2.98,2.87	2,6H 6.74; 3,5H 7.05
C36	8.18	4.31	2.82,2.90	

and the ROESY cross-section at 60 ms mixing time (data not shown). The inter-residue H^a -HN NOESY cross-peaks for I13-T14 & T14-I15 are always partially overlapping in all NOESY and ROESY cross-sections. The assignments of (H^a ,HN) of C2, T3, & R4 do not readily follow from the connectivity route because H^a of C2, T3, & R4 are not sufficiently resolved in chemical shift so as to allow construction of a clear $H^a(i)$ -HN(i+1) connectivity. Because there are other possible $H^a(i)$ -HN(i+1) connectivities involving T-R/R-T sequences, we first constructed the $H^a(i)$ -HN(i+1) connectivity route for the relatively non-overlapping regions of the spectrum. We deferred the assignment of C2, T3, & R4 till the end when C2 is distinguished from

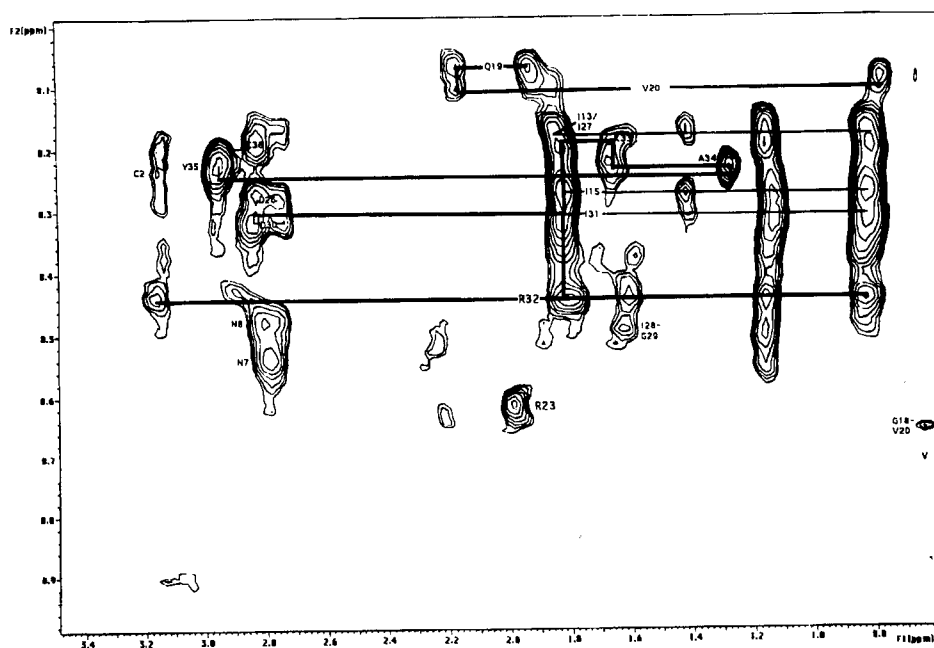


Figure 5: HN-H^β NOESY (mixing time = 400 ms) cross-section. The experimental and solutions conditions are the same as in Figure 4.

C36 by the presence of C2-S6 interactions and T3 and R4 are easily distinguished in terms of their intra-residue TOCSY/NOESY HN-H^β spin correlation. Note that the centers of the (H^α-HN) NOESY cross-peaks (Figure 4) are not as clearly defined as the corresponding TOCSY (Figure 2) or DQF-COSY (Figure 3) cross-peaks. Therefore, there are mismatches between the the centers of TOCSY cross-peak and the apparent centers of the corresponding NOESY cross-peaks. We, however, have been consistent by adhering to the peak centers as determined from the TOCSY (Figure 2) or DQF-COSY (Figure 3) cross-sections and therefore, lines in the NOESY cross-section (Figure 4) appear not to exactly pass through the apparent centers of two (H^α-HN) cross-peaks. Chemical shift values are given in Table I correspond to the centers of TOCY peaks.

Figure 5 shows the H^β-HN NOESY cross-section (mixing time = 400 ms) of the Thailand TN243 V3 loop at 10°C. Note the presence of sequential H^β(i)-HN(i+1) NOESY connectivity for the residues in the C-terminal I31-R32-K33-A34-Y35 segment. Such an NOE pattern is expected for residues in an α-helical segment. However, the absence of sequential HN-HN NOESY cross-peaks for all the residues in the segment (Figure 6) suggests that the residues in this segment are only partially folded and do not form a regular rigid helix (see Figures 8 & 9 below). The R32 shows the intra-residue HN-H^δ NOE and the inter-residue H^β(I31)-HN(R32) NOE. Inter-residue NOE is also observed for the sequential H^β(Q19)-HN(V20) interaction. The NOEs involving H^{β2}(P5)-HN(N7) are consistent with a turn at the T4-P5-S6-N7 sequence. The strong and medium NOEs in Figures 4 and 5 are also present in the corresponding NOESY cross-sections at 100 ms of mixing. However, the sequential HN-HN NOESY cross-peaks are only observed at 400 ms of mixing. Figure 6 shows

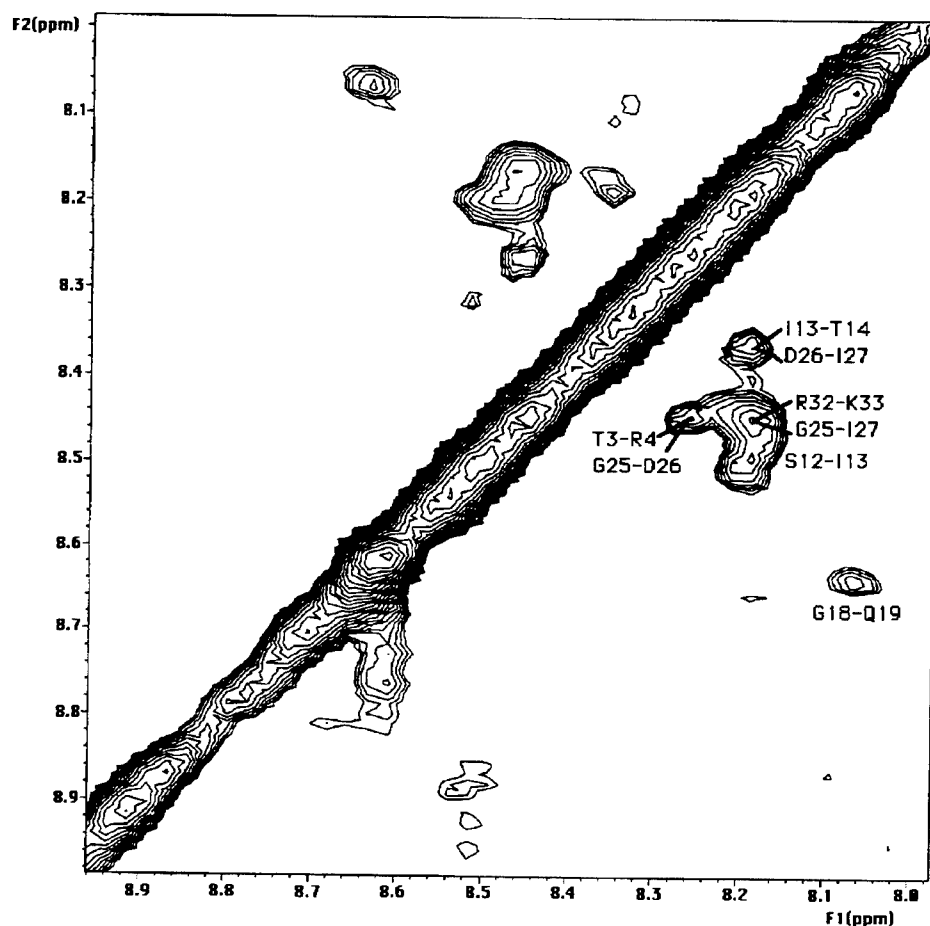


Figure 6: HN-HN NOESY (mixing time = 400 ms) cross-section. The experimental and solutions conditions are the same as in Figure 4.

this cross-section. The cross-peak for HN(G18)-HN(Q19) interaction is consistent with a turn at G16-P17-G18-Q19. In the partially folded segment, the cross-peak HN(R32)-HN(K33) is clearly visible; the cross-peaks due to HN(K33)-HN(A34) & HN(A34)-HN(Y35) are too close to the diagonal to be clearly visible. Considering the fact that strong NOESY cross-peaks are obtained for H^{β} (K33)-HN(A34) & H^{β} (A34)-HN(Y35) interactions even at 100 ms mixing time, medium NOESY intensities with soft force constants ($k = 1 \text{ KCal/Mole}$) are assumed for these two distances. The cross-peak for HN(G25)-HN(D26) is consistent with type-I turn T24-G25-D26-I27.

Solution NMR studies were previously reported for two other types of the V3 loop sequences [19,20]. As in the present case, the structure determination from the 2D NMR data in the previous two cases, was complicated by poor resolution and peak overlap in the NOESY spectrum. As also noted in the previous studies, the complications in the NMR spectrum were primarily due to the flexible nature of the V3 loops. Although some local structures are formed mostly in the form of putative

turns [19,20], the overall folding of the two previously reported V3 loops were rather loose. In addition, the simultaneous observation of the sequential H^{α} -HN (typical for an extended or a beta conformation) and the sequential HN-HN (typical for an alpha helical conformation) NOESY crosspeaks indicated that perhaps the V3 loop (especially for the HIV-MN isolate, reference 19) was sampling two widely different conformations. However, in the case of the Thailand TN243 V3 loop, only weak sequential HN-HN NOESY crosspeaks are observed for a few residues [except for NH(G18)-NH(Q19) which is of medium intensity]; whereas the sequential H^{α} -HN crosspeaks for the residues assigned in the extended or beta conformation are quite strong. This coupled with the (H^{α} -HN) J-coupling data (Figure 7) suggest that most of the residues (though flexible) are, indeed, in the extended or beta conformation.

As discussed in the following section, here we have made an attempt to visualize the nature of the conformational flexibility of a particular V3 loop from a *Thailand* HIV isolate. The flexibility is analyzed subject to the NMR data which shows the presence of three localized turns and two stretches of extended conformation.

Structure and Flexibility

Figure 7 summarizes the results of NMR experiments [18]. The majority of the observed NOEs are due to intra-residue and nearest-neighbor inter-residue interactions. Seven observed inter-residue NOEs are due to long-range interactions, i.e., pairwise interactions involving residues that are not nearest neighbors. The interaction involving P5 and N7 suggests a type-I turn at the site of glycosylation R4-P5-S6-N7* (N7* being the locus of N-linked glycosylation). Similarly, the interaction involving T24 and I27 suggests a type-I turn at T24-G25-D26-I27. In addition, a type-II turn at G16-P-17-G18-Q19 and two β -strands/extended conformations flanking this turn are indicated on the basis of intra-residue and nearest-neighbor inter-residue NOEs. The secondary structural features mentioned above and the constraints of the (S-S) bridge between C2 and C36 (Figure 1) are used in *Step (ii) of Materials and Methods* to obtain various models in agreement with the NOESY/ROESY data. From these models a set of 100 inter-residue distances are extracted as structural constraints that are crucial for agreement with the NMR data. Observation of almost all intra-residue H^{α} -HN NOESY cross-peaks suggests that the effective tumbling time of the individual peptide units in the V3 loop is long enough so that we are able to capture distances less than 3 Å in the laboratory frame NOESY experiment. However, the inherent flexibility of the link between two neighboring peptides allows a whole range of dynamics for the sequential interactions. The absence of a given sequential NOESY cross-peak does not necessarily imply that the corresponding distance is beyond a certain limit. The local dynamics of the link between two neighboring peptide units may reduce the effective correlation time and thereby place $\omega\tau_c$ in the intermediate range and elude the observation of corresponding NOESY peaks in the laboratory frame experiments. We, therefore, supplemented 2D NOESY (100 & 400 ms of mixing time) with the ROESY (60 & 200 ms of mixing time) data. This allowed us to observe inter-proton distances below a limit regardless of their effective correlation time. Full Relaxation-Matrix analyses allowed us to stipulate two types of constraints: (1) upper and lower distance limits for the

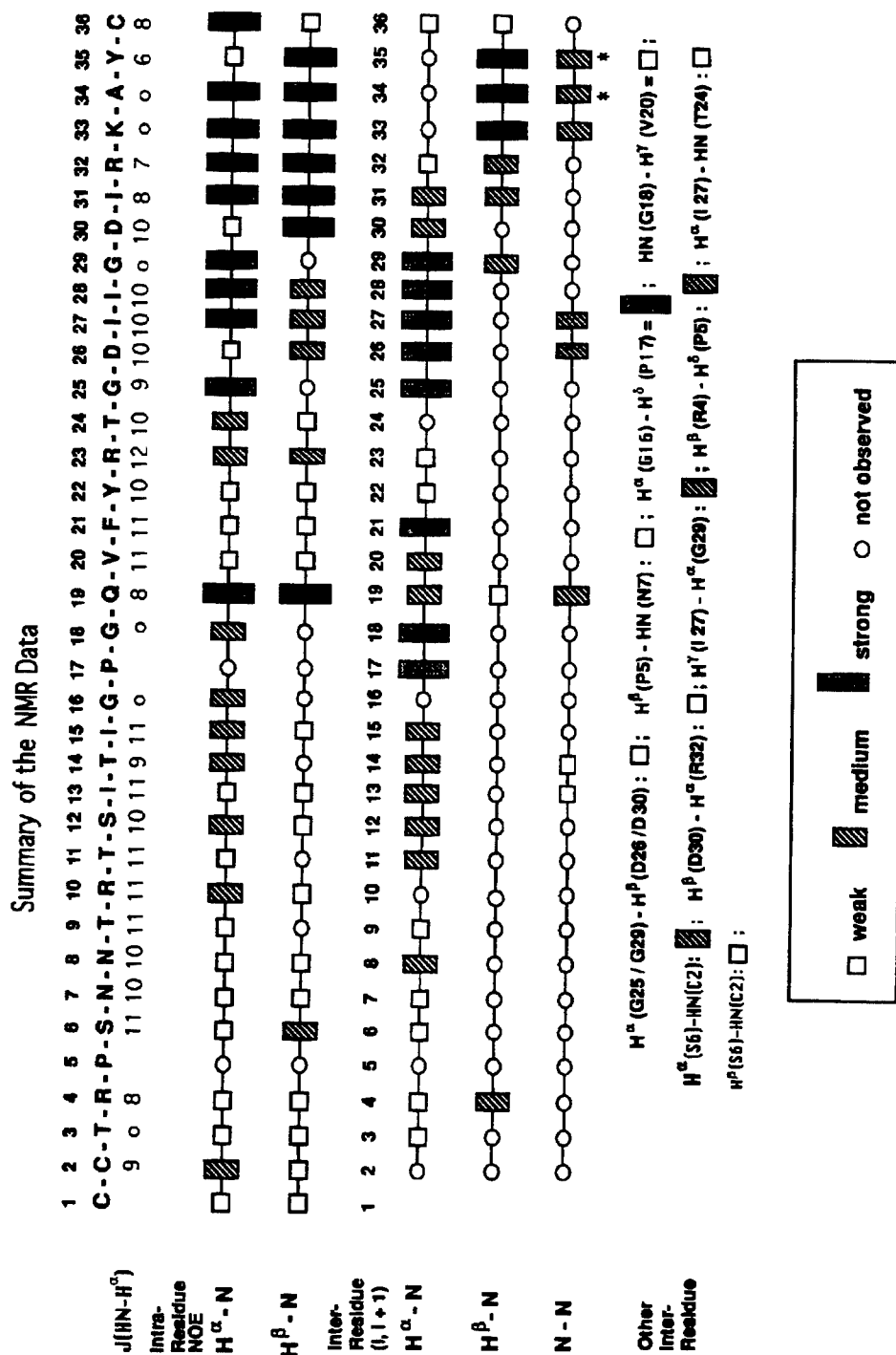


Figure 7: Summary of the DQF-COSY, the NOESY data at mixing times of 100 & 400 ms, and the ROESY data at mixing times of 60 and 200 ms reveal the shown interaction pattern for intra- and inter-residue pair-wise contacts in the *Chiang Mai* V3 loop.

observed NOESY/ROESY peaks and (2) the lower distance limits for the unobserved NOESY/ROESY peaks. These constraints are discussed in the *Materials Methods* and in Table II of the supplementary material. Only inter-residue distances are used as distance constraints. The force constants, k , for the distance constraints corresponding to the observed peaks are set at 5 KCal/Mole. During MC simulated annealing and energy minimization if these constraints are violated by more than 1 Å, then k is increased to 50 KCal/Mole to bring the distances close to the constrained values. The force constant, k , for the unobserved distances are set at 0.5 Kcal/Mole so that the conformational search is not biased by these sets of constraints. A soft force constant, k , of 1 KCal/Mole is used for the distances that correspond to the observed (i) and (i+2/3/4) NOEs/ROEs (that are weak in intensity even at 400 ms mixing time).

MC simulated annealing, followed by energy minimization subject to 100 distance constraints, led to about 100 low energy structures in agreement with the NMR data [see step (iii) of *Materials and Methods*]. Out of these, 37 models with lowest energies and rms deviations of 0.40 ± 10 Å with respect to the 100 distance constraints are chosen as structural solutions. This rms deviation is on top of the allowed distance range stipulated for agreement with the NMR data. Hypothetically, if this deviation were zero then we would have a perfect agreement with the NMR data. Although several starting structures are chosen, the simulated annealing protocol used in this work is by no means an exhaustive search in the conformational space of the V3 loop. This method allows us to search for the local minima on the sampled space of the V3 loop. The sampling is guided by the force-field and the set of inter-residue distance constraints that satisfy the NMR data. The inherent nature of the flexibility of this V3 loop is visualized by making the following observations. Although in the set of energy-minimized structures the standard deviations in the (ϕ, ψ) -values are within 30° , the average rms deviations in the spatial position of the C^α -atoms among 37 structures (Figure 8B) are as large as 1.5 Å. This shows that even small changes in (ϕ, ψ) -values when appropriately correlated causes a large overall change. Data presented here shows the nature of the conformational flexibility in the *Chiang Mai* V3 loop arising due mainly to correlated motions around the single bond torsions, i.e., peptide backbone angles (ϕ, ψ) , with little cost of energy. The extent of flexibility shown here merely reflects the lowest possible values because the thermal motions are filtered out after energy minimization.

Figure 8A shows the average folding pattern of the V3 loop; different secondary structural elements are color coded. Although most of the residues are in the beta/extended state (shown in green), the constituent secondary structural elements are so assembled that in all 37 structures the GPG-crest at the center of the V3 loop forms a protruding surface with Q19 exposed to the environment. Also the turn at R4-P5-S6-N7* exposes N7* to the environment such that N7* is accessible for glycosylation (Figure 8B). It may be noted that the amino acid site of glycosylation often forms a beta-turn and resides on the exposed surface of the protein [21].

As shown in Figure 9, the extent of the flexibility of the *Chiang Mai* V3 loop is best described by analyzing the average and the corresponding standard deviations of

Table II

Inter-residue distance constraints for 101 proton-pairs. The constraints belong to two categories : (1) constraints with specified ranges; (2) constraints specifying that the target distances should be above the specified upper limits. DIST = average distance for a proton-pair; standard deviation is shown inside the parenthesis. DEQ1 and DEQ are the lower and upper distance limits. DEV = average deviation from the specified constraint. Note that only about 10 constraints are violated by more than 1 Å.

ATOM1	RES1	ATOM2	RES2	DIST	DEQ1	DEQ	DEV
HA	CYS 2	HN	THR 3	3.66 (.02)	4.00	5.00	.34
HN	CYS 2	HA	SER 6	4.41 (.17)	3.00	4.00	.41
HN	CYS 2	HB	SER 6	5.04 (.19)	4.00	5.00	.10
HA	THR 3	HN	ARG 4	3.71 (.05)	3.00	4.00	.00
HA	PRO 5	HN	SER 6	2.27 (.08)	4.00	5.00	1.73
HA	SER 6	HN	ASN 7	2.68 (.07)	3.50	4.50	.82
HA	ASN 7	HN	ASN 8	2.60 (.07)	2.50	3.00	.00
HA	ASN 8	HN	THR 9	2.64 (.08)	4.50	5.50	1.86
HA	THR 9	HN	ARG10	2.24 (.13)	5.00	5.00	2.76
HA	ARG10	HN	THR11	2.35 (.04)	2.50	3.00	.15
HA	THR11	HN	SER12	2.52 (.07)	2.50	3.00	.02
HA	SER12	HN	ISO13	2.34 (.07)	2.50	3.00	.16
HA	ISO13	HN	THR14	2.19 (.01)	2.50	3.00	.31
HA	THR14	HN	ISO15	2.82 (.05)	2.50	3.00	.00
HA	ISO15	HN	GLY16	2.18 (.01)	3.50	4.50	1.32
HA	PRO17	HN	GLY18	2.49 (.07)	2.00	2.50	.02
HA	GLY18	HN	GLN19	3.49 (.03)	3.00	3.75	.00
HA	GLN19	HN	VAL20	3.42 (.05)	3.00	3.75	.00
HA	VAL20	HN	PHE21	2.46 (.05)	2.20	2.70	.00
HA	PHE21	HN	TYR22	2.30 (.11)	4.00	5.00	1.70
HA	TYR22	HN	ARG23	2.24 (.05)	4.50	5.50	2.26
HA	ARG23	HN	THR24	2.30 (.05)	3.50	4.50	1.20
HA	THR24	HN	GLY25	2.21 (.02)	2.95	2.95	.74
HA	GLY25	HN	ASP26	2.83 (.28)	2.20	2.70	.19
HA	ASP26	HN	ISO27	2.47 (.04)	2.20	2.70	.00
HA	ISO27	HN	ISO28	2.27 (.02)	2.20	2.70	.00
HA	ISO28	HN	GLY29	2.31 (.03)	2.20	2.70	.00
HA	GLY29	HN	ASP30	2.74 (.21)	2.50	3.00	.01
HA	ASP30	HN	ISO31	2.39 (.04)	2.50	3.00	.11
HA	ISO31	HN	ARG32	2.52 (.04)	2.50	3.00	.01
HA	ARG32	HN	LYS33	2.95 (.27)	4.00	4.00	1.05
HA	LYS33	HN	ALA34	2.93 (.16)	4.00	4.00	1.07
HA	ALA34	HN	TYR35	3.17 (.16)	4.00	4.00	.83
HA	TYR35	HN	CYS36	2.94 (.07)	4.00	4.00	1.06
HB	CYS 2	HN	THR 3	3.78 (.15)	4.50	4.50	.72
HB	THR 3	HN	ARG 4	3.06 (.20)	2.50	3.50	.00
HB	PRO 5	HN	SER 6	4.21 (.17)	4.50	4.50	.29
HB	SER 6	HN	ASN 7	3.64 (.10)	4.50	4.50	.86
HB	ASN 7	HN	ASN 8	4.73 (.01)	4.50	4.50	.23
HB	ASN 8	HN	THR 9	3.76 (.09)	4.50	4.50	.74
HB	THR 9	HN	ARG10	4.21 (.36)	4.50	4.50	.29
HB	ARG10	HN	THR11	4.31 (.05)	4.50	4.50	.19
HB	THR11	HN	SER12	3.91 (.07)	4.50	4.50	.59
HB	SER12	HN	ISO13	4.30 (.10)	4.50	4.50	.20
HB	ISO13	HN	THR14	4.35 (.05)	4.50	4.50	.15
HB	THR14	HN	ISO15	3.55 (.05)	4.50	4.50	.95
HB	ISO15	HN	GLY16	4.33 (.02)	4.50	4.50	.17
HB	GLN19	HN	VAL20	4.17 (.07)	4.00	5.00	.00
HB	VAL20	HN	PHE21	4.00 (.07)	4.50	4.50	.50
HB	PHE21	HN	TYR22	4.20 (.12)	4.50	4.50	.30
HB	TYR22	HN	ARG23	4.25 (.14)	4.50	4.50	.25
HB	ARG23	HN	THR24	4.43 (.12)	4.50	4.50	.10
HB	THR24	HN	GLY25	4.57 (.02)	4.50	4.50	.07
HB	ASP26	HN	ISO27	3.76 (.08)	4.50	4.50	.74

Table II continued

ATOM1	RES1	ATOM2	RES2	DIST	DEQ1	DEQ	DEV
HB	ISO27	HN	ISO28	4.55 (.06)	4.50	4.50	.06
HB	ISO28	HN	GLY29	3.84 (.05)	2.80	3.50	.34
HB	ASP30	HN	ISO31	2.85 (.15)	2.80	3.50	.04
HB	ISO31	HN	ARG32	3.88 (.06)	2.80	3.50	.38
HB	ARG32	HN	LYS33	4.10 (.34)	2.50	3.25	.85
HB	LYS33	HN	ALA34	2.21 (.06)	2.50	3.25	.29
HB	ALA34	HN	TYR35	4.06 (.29)	2.50	3.25	.81
HB	TYR35	HN	CYS36	3.60 (.11)	4.50	5.50	.90
HN	CYS 2	HN	THR 3	3.57 (.15)	4.50	4.50	.93
HN	THR 3	HN	ARG 4	2.94 (.26)	2.50	3.50	.02
HN	SER 6	HN	ASN 7	4.37 (.03)	4.50	4.50	.13
HN	ASN 7	HN	ASN 8	3.65 (.11)	4.50	4.50	.85
HN	ASN 8	HN	THR 9	4.35 (.01)	4.50	4.50	.15
HN	THR 9	HN	ARG10	4.46 (.09)	4.50	4.50	.09
HN	ARG10	HN	THR11	4.27 (.05)	4.50	4.50	.23
HN	THR11	HN	SER12	4.41 (.01)	4.50	4.50	.09
HN	SER12	HN	ISO13	4.16 (.13)	3.00	3.50	.66
HN	ISO13	HN	THR14	3.93 (.14)	4.00	4.55	.08
HN	THR14	HN	ISO15	4.32 (.02)	4.50	4.50	.18
HN	ISO15	HN	GLY16	4.14 (.05)	4.50	4.50	.36
HN	GLY18	HN	GLN19	2.44 (.08)	2.70	3.25	.26
HN	GLN19	HN	VAL20	2.51 (.05)	2.70	3.20	.19
HN	VAL20	HN	PHE21	4.40 (.02)	4.50	4.50	.10
HN	PHE21	HN	TYR22	4.61 (.03)	4.50	4.50	.11
HN	TYR22	HN	ARG23	4.10 (.07)	4.50	4.50	.40
HN	ARG23	HN	THR24	4.60 (.06)	4.50	4.50	.10
HN	THR24	HN	GLY25	3.53 (.04)	4.50	4.50	.97
HN	GLY25	HN	ASP26	4.36 (.04)	3.00	4.50	.00
HN	ASP26	HN	ISO27	4.41 (.04)	3.00	4.50	.00
HN	ISO27	HN	ISO28	4.49 (.05)	4.50	4.50	.04
HN	ISO28	HN	GLY29	4.58 (.04)	3.00	3.75	.83
HN	GLY29	HN	ASP30	4.38 (.05)	4.50	4.50	.12
HN	ASP30	HN	ISO31	4.31 (.04)	4.50	4.50	.19
HN	ISO31	HN	ARG32	4.38 (.01)	4.50	4.50	.12
HN	ARG32	HN	LYS33	2.78 (.36)	2.50	3.25	.05
HN	LYS33	HN	ALA34	4.37 (.08)	2.50	3.25	1.12
HN	ALA34	HN	TYR35	1.90 (.08)	2.50	3.25	.60
HN	TYR35	HN	CYS36	4.28 (.07)	4.50	5.50	.22
HA	GLY16	HD	PRO17	2.18 (.04)	2.20	2.70	.03
HN	GLY18	HG1	VAL20	5.31 (.37)	5.00	6.00	.03
HA	GLY25	HB	ASP26	4.87 (.22)	5.00	6.00	.18
HB	ASP30	HA	ARG32	5.91 (.19)	4.50	5.50	.41
HG2	ISO27	HA	GLY29	4.83 (.29)	3.00	4.00	.83
HB	THR 3	HN	ARG 4	3.06 (.20)	3.00	4.00	.06
HN	THR24	HA	ISO27	8.32 (1.75)	4.50	5.50	2.82
HA	GLY25	HN	ISO27	6.71 (.11)	4.50	5.50	1.21

the (ϕ , ψ)-values for different residues. The average is taken over the 37 models that are finally selected as structural solutions. The residues close to the C-terminal are most flexible in terms of their (ϕ , ψ)-values (Figure 9) and, consequently, in terms of the spatial locations of their C $^{\alpha}$ -atoms (Figure 8B). Although we located two turns (one close to the N-terminal and the other close to the C-terminal), the presence of two Gs and intrinsically flexible/dynamic (R32-K33-A34-Y35) segment (as discussed above) make the C-terminal half of the V3 loop intrinsically more flexible part of the molecule. Table III containing the average values and the standard

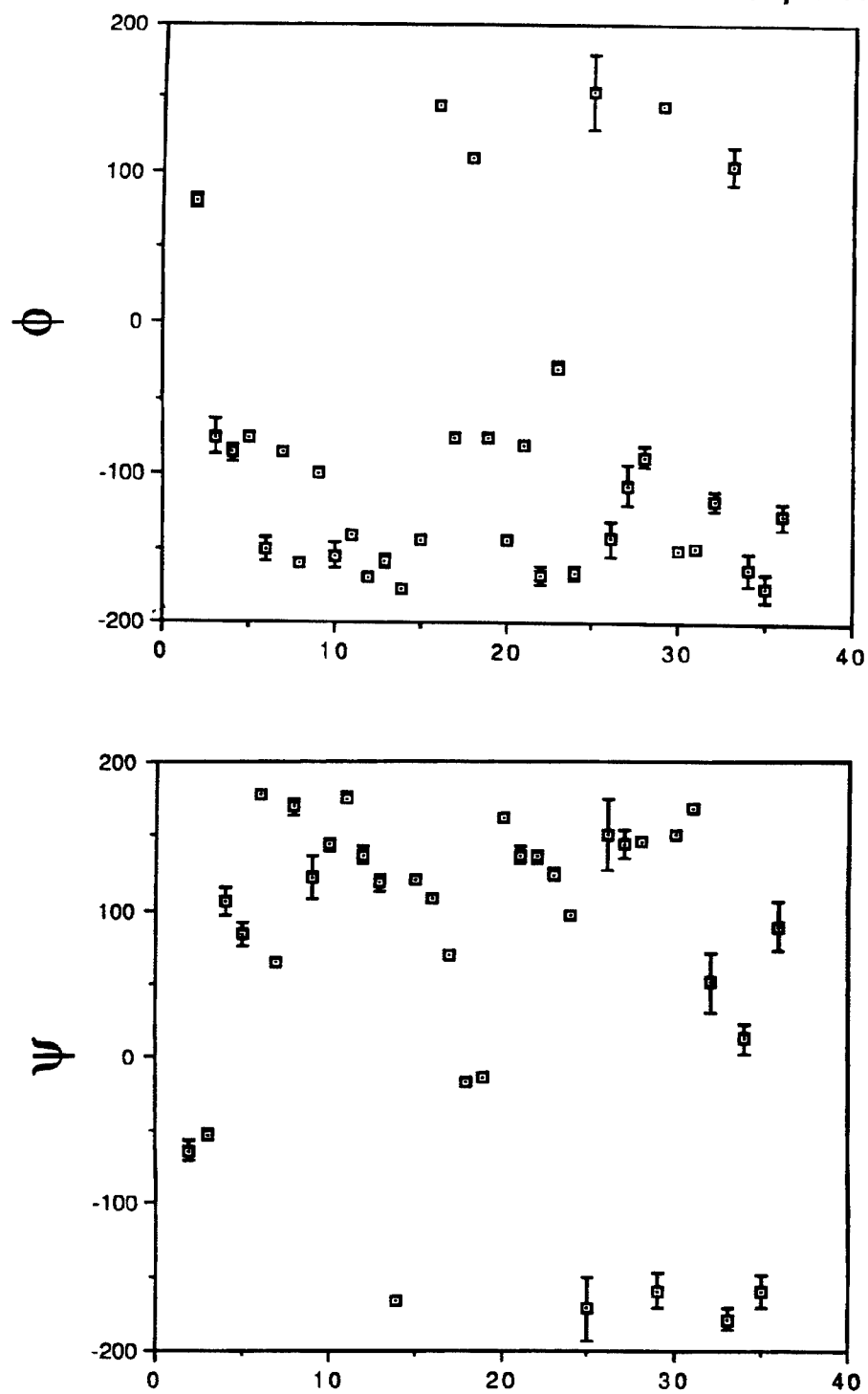
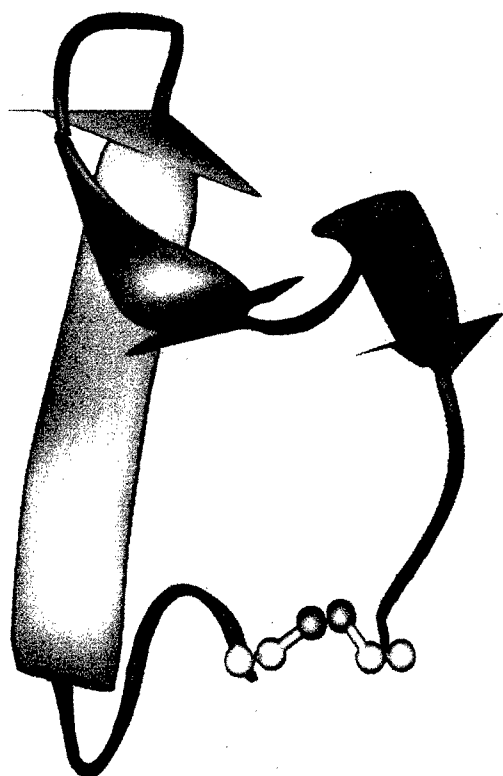
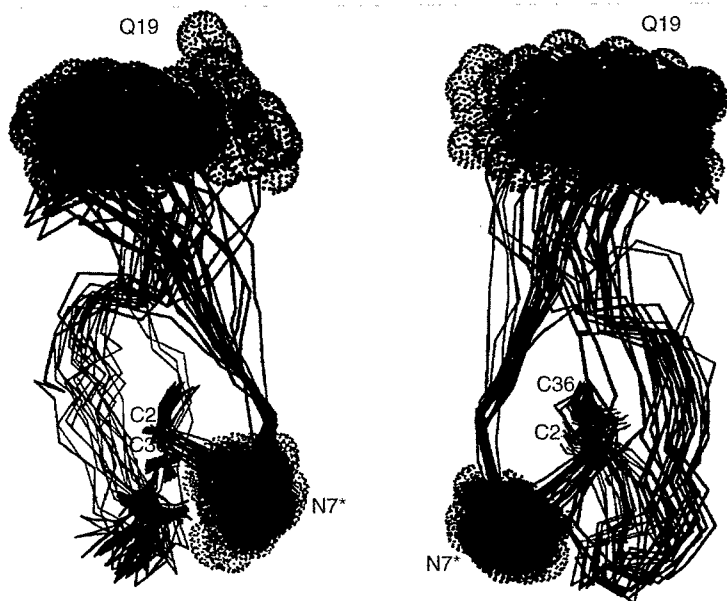


Figure 9: The average values and the corresponding standard deviations of the (ϕ, ψ) -values of different residues in the *Chiang Mai* V3 loop. 37 structures shown in Figure 7B are considered. The residues close to the C-terminal show the largest standard deviations. The Cartesian coordinates of all 100 V3 loop models can be obtained from the authors on request.

Figure 8: Description of the *Thailand* TN243 V3 loop structure



(A) A ribbon diagram of the average folding pattern in agreement with the 2D NMR data presented in *Figures 2-7* and in Table II of the supplementary material. The secondary structural elements are color coded: β -strand/extended = green, turn and coil = blue. The sulfur atoms of C2 & C36 involved in the (S-S) bridge are shown in magenta. The residue C2 is on the left and C36 is on the right. The immunogenic tip containing GPG at the center forms a protruding surface. The first turn close to the N-terminal contains the residue N7* (the site of N-linked glycosylation).



(B) Superposition of 22 lowest energy structures in agreement with the 2D NMR data in two different orientations. The rms deviations of these structures are 40 ± 10 Å with respect to the 100 distances obtained as experimental structural constraints. Note that only C^α -atoms are shown for clarity. The residues close to the C-terminal show greater flexibility. The *van der Waal* surfaces of N7* and Q19 are shown. Note that both functionally determinant sites are exposed to the environment. Out of 37 finally selected structures, only 22 models that do not show close overlap are shown here. In different structures the relative positions of Q19 are different with respect to C2 and C36 but not the local structure of G16-P17-G18-Q19 and the associated solvent accessibility of Q19.

deviations of all torsion angles [i.e., ($\phi, \psi, \chi^1, \chi^2, \chi^3$, etc.)] is also included in the supplementary material.

The molecular diagrams in Figures 8A & 8B and (ϕ, ψ)-values in Figure 9 show that most of the individual amino acids reside in the extended conformation. This conclusion is based upon the observation of typical sequential $H^{\alpha}(i)$ -HN($i+1$) NOESY/ROESY connectivities and values of the $J(H^{\alpha}$ -HN) coupling values shown in Figures 3 and 7. The $H^{\alpha}(i)$ -HN($i+1$) NOESY/ROESY connectivities alone cannot justify an extended conformation. For example, such connectivities are often observed also in a "random coil" in which each residue samples several (ϕ, ψ)-values. For some of the (ϕ, ψ)-values, the $H^{\alpha}(i)$ -HN($i+1$) distance can be short enough to show NOE/ROE. However, for a random coil, an average value of $J(H^{\alpha}$ -HN) coupling should be observed, i.e., a value of about 6-7 Hz. The ϕ -value corresponding to an average value of $J(H^{\alpha}$ -HN) coupling may never be adopted by a residue. This is especially true when a residue samples around two distinctly separated ϕ -values (perhaps separated by a barrier), i.e., (1) the alpha-helical region ($\phi = 50^{\circ} + 30^{\circ}$) giving rise to an average $J(H^{\alpha}$ -HN) coupling of 5 Hz and (2) the region of extended conformation ($\phi = 150^{\circ} \pm 30^{\circ}$) giving rise to an average $J(H^{\alpha}$ -HN) coupling of 9 Hz. We have determined $J(H^{\alpha}$ -HN) couplings for various residues from the DQF-COSY data (Figures 3 and 7). We ascribe extended conformation only to those residues that simultaneously show strong or medium $H^{\alpha}(i)$ -HN($i+1$) NOESY/ROESY connectivities and $J(H^{\alpha}$ -HN) couplings greater than or equal to 10 Hz.

Discussion

As explained in the *Introduction*, accurate determinations of (1) the global tertiary fold and (2) the local structure of the PND are needed for explaining the antigenic specificity of the V3 loop. In addition, the fact that the hypervariability of the amino acid sequence flanking the conserved GPG-sequence results in the variability in the antigenic specificity of the V3 loop, makes it necessary to perform a systematic study to quantify the effect of the amino acid sequence variability on the structure/antigenicity of the V3 loop. We show in this article that the residues flanking the GPG-sequence of the Thailand TN243 V3 loop adopt extended conformation. The protruding PND surface, which contains 5-7 amino acids contiguous in space and in sequence, can be recognized by an antibody via "direct reading mechanism." However, if the amino acids in the regions flanking the GPG-sequence are folded into an α - or 3_{10} -helix, the protruding PND surface can still be formed by 5-7 amino acids contiguous in space but not necessarily in sequence. In such an "indirect reading mechanism," residues in the interior of the antigenic tip also play an important role in deciding the size and shape of the cavity (and hence also the size and shape of the surface). We have identified a more folded conformation at the GPG-crest of the HIV-III_B V3 loop theoretically in presence of a monoclonal antibody, and have subsequently verified the presence of such a conformation by site-specific mutations [22,23]. The goal of such studies is to classify V3 loop sequences in terms of their (1) global tertiary fold and (2) local structure of the PND, rather than by their amino acid sequence. Determination and classification of structural/antigenic properties of the V3 loops will be useful in vaccine development.

Table III

The average values of all torsion angles and the corresponding standard deviations. 37 structures are used. Note that the residues close to the C-terminal cystine (C36) show the largest deviations in their torsion angles. NOESY/ROESY data contains information about (ϕ, ψ, χ^1).

	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
C 2	81	-64	-161	178						
	5	7	7	10						
T 3	-75	-53	150	50	165	-43				
	12	4	10	18	17	28				
R 4	-86	106	-159	-168	177	176	-173	1	179	0
	5	10	9	21	3	18	26	46	2	3
P 5	-75	84	-176							
	0	8	3							
S 6	-151	177	175	-84	-54					
	8	3	6	24	55					
N 7	-85	64	176	-171	-104	0				
	2	3	3	1	1	0				
N 8	-160	169	-179	-90	104	0				
	2	5	2	5	1	1				
T 9	-99	122	174	-59	71	176				
	3	14	2	2	2	4				
R 10	-155	144	-178	-109	-77	-176	-82	3	-177	0
	9	4	4	6	4	5	3	4	2	2
T 11	-141	175	175	-171	83	-173				
	3	4	2	1	13	1				
S 12	-170	137	-165	-83	56					
	2	5	9	3	6					
I 13	-158	118	-177	-176	168	-44	67			
	5	5	2	2	2	3	1			
T 14	-177	-166	-175	178	97	55				
	3	2	3	24	18	8				
I 15	-145	120	-177	-174	168	-45	68			
	2	2	1	1	1	2	1			
G 16	145	108	178							
	2	2	2							
P 17	-75	69	179							
	0	3	2							
G 18	109	-16	177							
	2	2	1							
E 19	-75	-13	-172	-160	176	98	-179			
	2	3	4	3	2	0	1			
V 20	-144	161	-178	78	63	40				
	3	3	2	8	4	5				
F 21	-81	137	-178	-178	81					
	3	6	4	2	3					
Y 22	-168	136	172	173	79	-179				
	7	4	3	11	13	3				
R 23	-29	124	-167	-111	-73	-63	156	-168	3	4
	4	4	3	14	7	4	9	5	29	1
T 24	-166	96	-178	174	81	-36				
	5	3	6	4	21	1				
G 25	154	-172	172							
	25	22	5							
D 26	-143	150	179	-56	-82					
	12	24	3	17	4					

Table III continued

	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
I 27	-107	144	153	-131	179	-51	-45			
	14	9	2	19	18	18	19			
I 28	-88	146	-175	-46	177	-173	-42			
	7	3	2	4	4	17	1			
G 29	145	-160	178							
	3	12	3							
D 30	-151	151	177	54	99					
	3	3	3	2	4					
I 31	-149	168	-178	-159	153	-169	62			
	2	3	4	15	21	20	8			
R 32	-117	51	169	-44	-70	172	-171	174	179	178
	7	20	6	5	3	11	4	3	1	1
K 33	-104	-179	-156	75	157	-170	157	56		
	13	7	5	5	14	8	30	7		
A 34	-163	13	-151	-55						
	11	10	8	2						
Y 35	-176	-160	168	-147	-123	0				
	9	11	6	4	4	41				
C 36	-127	89	179	-173						
	9	17	1	16						

Acknowledgement

This work is supported by a US Army Grant MIPR 92MM2581 and an NIH Grant R01 AI32891-01A2 (G.G). G.G wishes to thank Drs. A. E. Garcia, Christian Burks, and B. Korber for many useful discussions. G.G also thanks Dr. A. Gifford for preparing Figures 1 and 7 and Dr. Patricia Reitemeier for proof-reading the manuscript. G.G. and G.M. thank Dr. Francine McCutchan for providing the Thailand TN243 sequence prior to publication.

References and Footnotes

1. Geyer, H., Holschbach, C., Hunsmann, G. and Scheider, J., *J. Biol. Chem.* 263, 11760-11767 (1988).
2. Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N. and Gregory, T.J., *J. Biol. Chemistry* 265, 10373-10382 (1990).
3. Putney, S.D., Mathews, T.J., Robey, W.G., Lynn, D.L., Robert-Guroff, M., Mueller, W.T., Langlois, A.J., Ghrayeb, J., Petteway, S.R., Weinhold, K.J., Fischinger, P.J., Wong-Staal, F., Gallo, R.C. and Bolognesi, D.P., *Science* 234, 1392-1395 (1986).
4. Goudsmit, J., Debouck, C., Meleon, R.H., Smit, L., Bakker, M., Asher, D.M., Wolff, A.V., Gibbs, C.J. and Gajdusck, D.C., *Proc. Natl. Acad. Sci. U.S.A.* 85, 4478-4482 (1988).
5. Javaherian, K., Langlois, A.J., McDanal, C., Ross, K.L., Eckler, L.I., Jellis, C.L., Profy, A.T., Rusche, J.R., Bolognesi, D.P., Herlihy, W.C., Putney, S.D. and Mathews, T.J., *Proc. Natl. Acad. Sci. U.S.A.* 86, 6768-6772 (1989).
6. Rusche, J.R., Javaherian, K., McDanal, C., Petro, J., Lynn, D.L., Grimaldi, R., Langlois, A.J., Gallo, R.C., Arthur, L.O., Fischinger, P.J., Bolognesi, D.P., Putney, S.D. and Mathews, T.J., *Proc. Natl. Acad. Sci. U.S.A.* 85, 3198-3202 (1988).
7. Korber, B., Wolinsky, S., and Myers, G., In *Vaccines92: Modern approaches to new vaccines including prevention of AIDS*. Cold Spring Harbor Press, pp 75-79 (1992).

8. LaRosa, G.J., Davide, J.P., Weinhold, K., Waterbury, J.A., Profy, A.T., Lewis, J.A., Langlois, A.J., Dressman, G.R., Boswell, R.N., Shaddock, P., Holley, L.H., Karplus, M., Bolognesi, D.P., Mathews, T.J., Emini, E.A. and Putney, S.D., *Science* 249, 932-935 (1990).
9. Davies, D., Padlan, E.A. and Sheriff, S., *Ann. Rev. Biochem.* 59, 439-473 (1990).
10. Ramachandran, G.N. and Sasisekharan, V., *Adv. Protein Chem.* 23, 283-438 (1968).
11. Fletcher, R., *Practical Methods of Optimization 1*, (NY: John Wiley & Sons, 1984).
12. Sippl, M.J., Nmethy, G. and Scheraga, H.A., *J. Phys. Chem.* 88, 6231 (1984).
13. Roterman, I.K., Gibson, K.D. and Scheraga, H.A., *J. Biomol. Str. & Dyn.* 7, 1-25 (1989).
14. Kirkpatrick, S., Gelatt, C.D., Jr. and Vecchi, M.P., *Science* 220, 671-680 (1983).
15. Gupta, G. and Myers, G., *Cinquieme colloque des cent gardes*: (Paris:Pasteur), pp 99-105 (1990).
16. Sklenar, V. and Bax, A., *J. Magn. Reson.* 74, 469-479 (1987).
17. Neuhaus, D., Wagner, G., Vasak, M., Kagi, J.H.R. and Wuthrich, K., *Eur. J. Biochem.* 151, 257-273 (1985).
18. Wuthrich, K., Billeter, M. and Braun W., *J. Mol. Biol.* 180, 715-740 (1984).
19. Chandrasekhar, K., Profy, A.T. and Dyson, H.J., *Biochemistry* 30, 9187-9194 (1991).
20. Zvi, A., Hiller, R. and Anglister, J., *Biochemistry* 31, 6972-6279 (1992).
21. Rose, G.D., Gierasch, L.M. and Smith, J.A., *Adv. Protein Chem.* 37, 1-106 (1985).
22. Veronese, F.D., Reitz, M.S., Gupta, G., Robert-Guroff, M., Boyer-Thompson, C., Louie, A., Gallo, R. and Lusso, P., *J. Biol. Chem.*, in press (1993).
23. Gorny, M.K., Xu, J., Gianakakos, V., Karwowska, S., Williams, C., Sheppard, H.Y., Hanson, C.V. and Zolla-Pazner, S., *Proc. Natl. Acad. Sci. U.S.A.* 88, 3238-3242 (1991).
24. Single letter amino acid codes are as follows: A=alanine, C=cystine, D=aspartic acid, E=glutamic acid, F=phenylalanine, G=glycine, H=histidine, I=isoleucine, K=lysine, L=leucine, M=methionine, N=asparagine, P=proline, Q=glutamine, R=arginine, S=serine, T=threonine, V=valine, W=tryptophan.
25. Louwagie, J., McCutchan, F.E., Peeters, M., Brennan, T.P., Sanders-Buell, E., Eddy, G.A., Groen, G.V.A., Franssen, K., Guy-Michel, G.-D., Deleys, R. and Burke, D.S., *AIDS* 7, 769-780 (1993).
26. Myers, G., *AIDS Res. & Human Retroviruses* 9, 697-702 (1993).

Date Received: July 8, 1993

Communicated by the Editor V. Sasisekhavan

Structure of a Tumor Associated Antigen Containing a Tandemly Repeated Immunodominant Epitope[†]

J. Darrell. Fontenot¹, S.V. Santhana Mariappan¹, Paolo Catasti^{1,2},
Neives Domenech³, Olivera J. Finn³ and Goutam Gupta^{1*}

¹Theoretical Biology and Biophysics,

²Life Sciences Division, LS-2, MS 880

Los Alamos National Laboratory

T-10, MS-K710

Los Alamos, NM 87545

³ Department of Molecular Genetics and Biochemistry,
University of Pittsburgh, School of Medicine
Pittsburgh PA, 15216

Abstract

Human mucins are T or S glycosylated tandem repeat proteins. In breast cancer, mucins become under or unglycosylated. Two-dimensional nuclear magnetic resonance experiments are performed on chemically synthesized mucin tandem repeat polypeptides, (PDTRPAGST-APPAHGVTSAn) in the unglycosylated form for n=1,3 where (APDTR) constitutes the antigenic sites for the antibodies isolated from the tumors in the breast cancer patients. These studies demonstrate how the tandem repeats assemble in space giving rise to the overall tertiary structure, and the local structure and presentation of the antigenic site (APDTR) at the junction of two neighboring repeats. The NMR data reveal repeating knob-like structures connected by extended spacers. The knobs protrude away from the long-axis of Muc-1 and the predominant antigenic site (APDTR) forms the accessible tip of the knob. Multiple tandem repeats enhance the rigidity and presentation of the knob-like structures.

Introduction

Human mucins are a family of high molecular weight, heavily glycosylated proteins which are dominated by large tandem repeat (TR) domains (1-4). Mucin tandem

* Author to whom correspondence should be addressed.

[†] This work was supported by the LANL grant XL 77.

Abbreviations Three Muc-1 tandem repeats (3TR), one Muc-1 tandem repeats (1TR)3-trimethylsilylpropionate (TSP), Monte Carlo (MC), parts per million (ppm), Rapid Multiple Peptide Synthesizer (RaMPS), trifluoroacetic acid (TFA), 9-fluorenylmethyloxycarbonyl (Fmoc), total correlation spectroscopy (TOCSY), rotating frame Overhauser effect spectroscopy (ROESY), nuclear Overhauser effect spectroscopy (NOESY), double quantum filtered correlated spectroscopy (DQF-COSY), end-to-end length of protein (Re), major histocompatibility complex (MHC), β -D-N-acetylgalactosamine (GalNac), Single letter amino acid codes are as follows: A=alanine, D=aspartic acid, G=glycine, H=histidine, I=isoleucine, P=proline, R=arginine, S=serine, T=threonine, V=valine

repeat domains vary in size, proline content, and potential extent of glycosylation. Underglycosylation of the human mucin Muc-1 tandem repeat domain in certain breast, pancreatic and ovarian tumors results in the unmasking of protein core epitopes (5-8). Tumor reactive, mucin-specific monoclonal antibodies reveal differences between the surface of Muc-1 derived from tumors and normal tissues (6,9,10). Synthetic peptide studies show that most tumor specific antibodies recognize an epitope within the tandem repeat protein core of Muc-1 (11,12).

Humoral immune responses and the epitope specificity of antibodies to the tandem repeats of Muc-1 raise interesting structural questions (the sequence in each repeat is PDTRPAPGSTAPPAHGVTS_A). For example, every monoclonal antibody that has been shown to be specific for the protein core of the TR domain of Muc-1, and in which the fine specificity has been mapped, recognizes some or all of the sequence A0-P1-D2-T3-R4-P5-A6 (A0 is the last residue from the previous repeat). This includes 19/19 different monoclonal antibodies to the protein core reviewed by Devine and McKenzie (13) and Taylor-Papadimitriou (10). In addition, mucin specific IgM humoral immunity which has been detected in sera from patients with breast and ovarian cancer has been determined to be specific for the A0-P1-D2-T3-R4-P5 as well (14,15). The induction of exclusively IgM antibodies in cancer patients suggests a mechanism of cross-linking of Ig receptors by multiple epitopes including the sequence A0-P1-D2-T3-R4-P5. The questions that immediately arises are: Why is the antibody specificity always centered on the sequence A0-P1-D2-T3-R4-P5? Is there a structural explanation for the immunodominance of this epitope?

Tumor reactive Muc-1 specific cytotoxic T-cell lines have been described from breast and pancreatic cancer patients in which it was indicated that the antigen specificity was the intact protein core of Muc-1 expressed by these tumors, and not the processed and presented forms (7,16). These α/β T-cells showed a lack of MHC-restriction, that is they were able to kill both breast and pancreatic tumor cell lines from individuals with different MHC alleles (16). Additional studies demonstrated that tumor reactive non-MHC restricted CTL lines and clones could be blocked by monoclonal antibodies to the same immunodominant epitope (A0-P1-D2-T3-R4-P5) of the Muc-1 tandem repeat, and not by an antibody to HLA class I (W632) (17,18). The multivalent nature of the TR domain suggests a possible mechanism for cross-linking of T-cell receptors in the process of MHC unrestricted T cell activation (19).

We have undertaken the current study to gain insight into the structural basis of the observed antibody immunodominance and MHC-unrestricted immunogenicity of Muc-1 tandem repeats. The primary structural/functional question raised by Muc-1, concerns the relationship between numbers of tandem repeats and the formation of the immunodominant antigenic determinant site, A0-P1-D2-T3-R4-P5. Two-dimensional nuclear magnetic resonance (2D NMR) spectroscopy on chemically synthesized mucin tandem repeat polypeptides, (PDTRPAPGSTAPPAHGVTS_A) and (PDTRPAPGSTAPPAHGVTS_A)₃, provide information about the global tertiary folding (and associated packaging) of the TR domain and the local structure of the antigenic determinant. These structural studies on Muc-1 TR with various repeat numbers, *n*, reveal that larger number of repeats (*n* > 2) enhance the definition of

the immunodominant site. In addition, we learn how these sites are spatially located and oriented with respect to each other.

Materials and Methods

Peptide Synthesis

Tandem repeat peptides 1TR and 3TR corresponding to 1 tandem repeat (20 residues) and 3 tandem repeats (60 residues) are synthesized by a manual solid-phase strategy using 9-fluorenylmethyloxycarbonyl protected amino acids. The final products are peptide amides. The procedures for synthesis, purification, and characterization of the peptide products are described in detail elsewhere (20). Briefly, 20, and 60 amino acid peptides are synthesized using the Rapid Multiple Peptide Synthesizer (RaMPS) apparatus from Dupont (Boston, MA). For 3TR, once the peptide chain reached 30 residues, the resin was split in half and separated into two reaction cartridges to allow space for the peptide chains to expand in the cartridge. Once the resins were divided, the concentration of input amino acid was maintained at .5 mM in order to drive the coupling reaction to completion with high efficiency. The products of the synthesis were deprotected and cleaved from the resin support in concentrated trifluoroacetic acid (TFA) in the presence of the appropriate scavengers. The TFA soluble products were extracted sequentially in organic solvents and then transferred to water and lyophilized. The peptides were purified by conventional gel filtration and reverse-phase high pressure liquid chromatography (HPLC). Molecular weight characterizations of the peptide products are performed with an electrospray mass spectrometer which verified that the desired products were obtained with high yield. The correct molecular weight product is purified by high pressure liquid chromatography (HPLC) prior to NMR analysis.

NMR Experiments and Sequential Assignments

All the spectra were recorded in a Bruker 500 MHz AMX instrument at 10 C with 5 mM peptide concentration in .01 M phosphate buffer (pH 5.5). All 2D data were acquired in the phase-sensitive mode with the saturation of the HDO signal during the relaxation delay. DQF-COSY data were collected with the following acquisition parameters: data matrix ($t_2=2048, t_1=1024$); relaxation delay = 1.5 s; number of transients = 32. TOCSY data were collected with the data matrix ($t_2=2048, t_1=1024$); relaxation delay = 1.5 s; number of transients = 32; isotropic mixing = 60ms; MLEV-64 pulse sequence for spin lock. NOESY data were collected with similar acquisition parameters and for 500 and 200 ms of mixing. ROESY data were collected in the phase-sensitive mode using the CW-spinlock pulse sequence with the following acquisition parameters: data matrix ($t_2=2048, t_1=512$); mixing time = 150 ms; relaxation delay = 1.5 s; number of transients = 32. The sequential assignment of 3TR is obtained by combining TOCSY, DQF-COSY, and NOESY data.

Extraction of Inter-proton Distances as Structural Constraints from the NMR Data

This step involves obtaining an energetically stable structure given the secondary

structural states of the constituent amino acid residues as obtained by analyzing the sequential NOESY and ROESY pattern. Appropriate ranges of (ϕ , ψ) values are assigned to all amino acids. For example,

$$\begin{array}{llll} \phi & = -55^\circ \pm 25^\circ, & \psi & = -55^\circ \pm 25^\circ & \text{for residues in a helix;} \\ \phi & = -140^\circ \pm 50^\circ, & \psi & = 140^\circ \pm 50^\circ & \text{for residues in a beta strand or} \\ & & & & \text{in an extended conformation;} \\ \phi_{i+1} & = -65^\circ \pm 20^\circ, & \psi_{i+1} & = -50^\circ \pm 20^\circ & \\ \phi_{i+2} & = -90^\circ \pm 20^\circ, & \psi_{i+2} & = 0^\circ \pm 20^\circ & \text{for residues in a type-I turn;} \\ \phi_{i+1} & = -65^\circ \pm 20^\circ, & \psi_{i+1} & = 120^\circ \pm 20^\circ & \\ \phi_{i+2} & = 90^\circ \pm 20^\circ, & \psi_{i+2} & = 0^\circ \pm 20^\circ & \text{for residues in a type-II turn.} \end{array}$$

(ϕ , ψ) of residues in the coil state are set free to choose any values in the allowed space (for definitions of different secondary structures and corresponding (ϕ , ψ)-values, see Ramachandran & Sasisekharan, 1968 (21)).

We obtain starting mucin structures by a linked-atom-least-square refinement by minimizing a function, F , only in the (ϕ , ψ) space.

$$\begin{aligned} F &= \text{R-Factor} + \sum_{ij} (d_{ij}^{mn} - D^{mn})^2 \\ \text{R-Factor} &= \frac{\sum |I_o - I_c|}{\sum I_o} \end{aligned} \quad [1]$$

I_o = observed NOESY intensity and I_c = calculated NOESY intensity by full-matrix NOESY simulations; d_{ij}^{mn} = the actual distance between atom i (type m) and atom j (type n) and D^{mn} = corresponding contact limits between atom types m and n . The sum extends over all pairs (i, j) of spins for which NOESY cross-peaks are observed. Full-Matrix NOESY simulations with respect to experimental data at two mixing times (one low and another high) enable us to include both primary and higher orders of NOEs. Thus, the complications in the distance estimate using a two-spin model often encountered at a high mixing time due to spin-diffusion (i.e., higher order NOEs) are avoided in the Full-Matrix NOESY simulations where all spins are considered in the relaxation (22). Such a simulation at two mixing improves the rigor of distance estimates.

In this refinement, first the (ϕ , ψ) values of various residues are only elastically varied (i.e., variables with weights) such that by appropriate choice of weights the experimentally determined secondary structural states of residues are minimally altered (23). Finally the potential energy of the 1TR/3TR sequence is minimized in the (ϕ , ψ , ω , χ) space using the force-field of Scheraga and co-workers (24). The minimization of the function, F , in the (ϕ , ψ) space followed by the energy-minimization in the (ϕ , ψ , ω , χ) space is repeated for 30 different starting structures (25). The set of 30 different structures are chosen such that they are conformationally different; by considering the positions of the Ca atoms the rms deviations among the chosen 30 structures are greater than 2.5 Å. As discussed later, such a large deviation is possible because of the flexibility in the mucin spacer. Although

the residues in the spacer remain in the extended conformation, the correlated motions in the backbone torsion angles of these residues can cause a large change without altering the sequential NOEs. The choice of 30 different starting structures duly account for the whole range of spacer flexibility. At the end of this procedure, we obtain a set of 30 models in agreement with the NOESY and ROESY data. From these models, a set of inter-proton distances are extracted as structural constraints required for agreement with the NOESY/ROESY data. Each pairwise distance representing a structural constraint provides an upper and a lower limit of the distance. Two types of constraints are identified.

Type I

This is given as

$$\begin{aligned} \text{EDIST} &= 0 \text{ if the distance } r \text{ is within a specified range } r_1 \text{ \& } r_2 \\ &= k(r-r_1)^2 \text{ if } r < r_1 \\ &= k(r-r_2)^2 \text{ if } r > r_2. \text{ } k : \text{ force constant.} \end{aligned}$$

Type II

This is given as

$$\begin{aligned} \text{EDIST} &= 0 \text{ if } r \geq r_1 \\ &= k(r-r_1)^2 \text{ if } r < r_1. \end{aligned}$$

This type is particularly useful for an unobserved NOE where we can set a lowest allowable distance limit for the corresponding proton pair. Analyses of the 2D NMR data of 3TR produced 220 inter-proton distance constraints.

Examination of the Conformational Flexibility Subject to the NMR Data

The energy term, EDIST, is added to the force-field as in Scheraga and co-workers (24). The simulated annealing is performed (26) in the following manner. *First*, a starting energy-minimized structure is chosen and Monte Carlo (MC) simulations are performed for 50,000 steps at 1000K in the (ϕ , ψ , ω , χ)-space; the last accepted conformation is stored to be subsequently used as a starting conformation in the next lower temperature-cycle. *Second*, 50,000 MC steps are repeated in several cycles of gradually decreasing temperature until a temperature of 100K is reached. *Third*, the lowest energy configuration at 100K is further energy-minimized to a low energy gradient. This is the "temperature quenching" step in which thermally excited single bond rotations around the equilibrium positions are quenched. *Finally*, the first three steps are repeated for 120 different starting conformations which comprise of the 30 starting structures obtained after the step discussed in the previous paragraph and 90 other structures as obtained after carrying out first through third steps of the simulated annealing protocol. The maximum step size of the torsion angles was set at 15 degrees which produced acceptance ratios of .20-.50 for the 50,000 step MC cycle at each temperature. Full-Matrix NOESY calculations were re-computed for the final 120 Muc-1 3TR structures. Details of the simulated annealing procedure and its application in efficient conformational search have already been described elsewhere (22, 27).

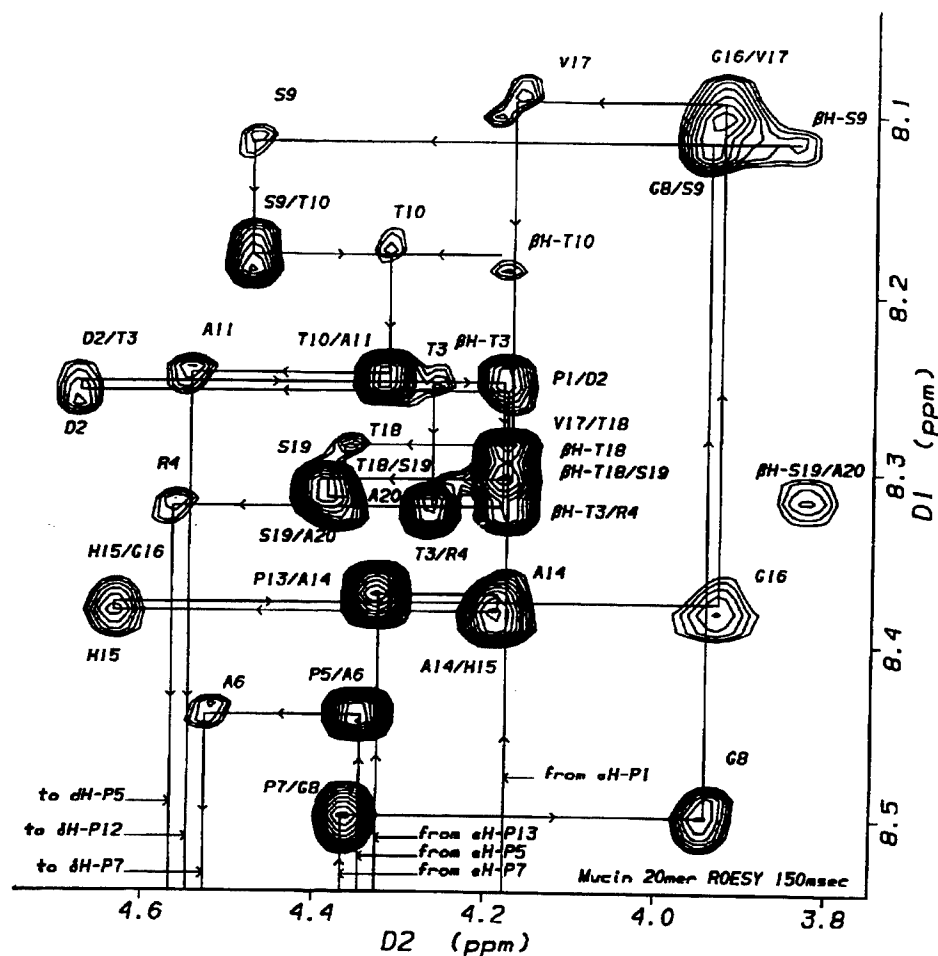


Figure 1: ROESY cross-section of the finger print HN-H α region of 5 mM 1TR in 90% H₂O + 10% D₂O at 10°C (10 mM phosphate buffer with pH 5.5).

Results

NMR Data of a Single Muc-1 Tandem Repeat (1TR)

For 1TR, the total correlation spectroscopy (TOCSY) and rotating frame Overhauser effect spectroscopy (ROESY) data are used to arrive at the sequential assignment of the protons belonging to the constituent amino acids; then ROESY data are used to estimate various inter-proton distances. 2D NMR studies on 1TR, show a flexible extended conformation of the molecule. The flexibility and the dynamics of the molecule is such that the effective correlation time for most of the pair-wise inter-proton interactions are fast enough to elude the observation of the corresponding nuclear Overhauser experiment (NOESY) cross-peaks in the laboratory-frame NOESY experiments. However, several inter-residue interactions are identified in the ROESY experiment (28, 29). Figure 1 shows the HN-H α or the finger-print

region of the ROESY spectrum of 1TR. The prominent sequential $H^a(i)$ -HN($i+1$) and the absence of HN(i)-HN($i+1$) NOEs indicate an extended conformation. The absence of most of the inter-proton interactions in the NOESY (30, 31) indicates that the extended structure is flexible. In addition, it is also evident from the NMR data of 1TR that the (A)PDTR-turn is not formed in the absence of A from the previous repeat.

Previous studies have shown that an unglycosylated 11 residue Muc-1 peptide can contain residual β -turn structure (32, 33). Scanlon et al. found a type I turn through the residues D2-T3-R4 using one complete tandem repeat p(1-20) for (34) and in three nine residue mucin or substituted mucin peptides (35). The difference in the type of turn and the one residue shift could be attributed to a preferential stabilization of the type I β -turn conformation in the organic solvent, dimethyl-sulfoxide used in this work. However, these studies reveal a tendency toward the formation of β -turns within this region.

NMR Data of Three Muc-1 Tandem Repeats (3TR)

Structural analyses of 3TR, (PDTRPAPGSTAPPAHGVTS_A)₃, involve a more extensive data set that includes the TOCSY and NOESY data for sequential assignment, double quantum filtered correlated spectroscopy (DQF-COSY) and the NOESY data at 500 and 200 ms of mixing times for structure derivation. Table I lists the chemical shift values of the protons belonging to the residues in 3TR. The observation of interproton interactions in the 500 and 200 ms of mixing time NOESY experiments indicates that 3TR is structurally more ordered than 1TR. Figure 2A and 2B show

Table I
Chemical Shift Values for the Central Tandem Repeat

	NH	H ^a	H ^b	H ^c	other
A20/40	8.43	4.63	1.39		
P21/41		4.42	2.30	1.92	3.81/3.68 (H ^b)
D22/42	8.51	4.64	2.74/2.67		
T23/43	8.14	4.33	4.27	1.20	
R24/44	8.32	4.62	1.80	1.71	3.22 (H ^b)
P5/25/45		4.41	2.28	1.88	3.83/3.62 (H ^b)
A6/26/46	8.54	4.59	1.38		
P7/27/47		4.43	2.32	1.96	3.85/3.68 (H ^b)
G8/28/48	8.60	3.99/3.95			
S9/29/49	8.22	4.54	3.95/2.90		
T10/30/50	8.28	4.39	4.25	1.21	
A11/31/51	8.34	4.61	1.36		
P12/32/52		4.71	2.36	1.90	3.86/3.64 (H ^b)
P13/33/53		4.39	2.28	1.86	3.83/3.63 (H ^b)
A14/34/54	8.47	4.25	1.35		
H15/35/55	8.47	4.69	3.26/3.20		7.28 (4H)
G16/36/56	8.47	4.00/3.97			
V17/37/57	8.20	4.24	2.13		
T18/38	8.39	4.43	4.25	1.22	
S19/39	8.37	4.47	3.85		

Chemical shift values in ppm of the protons belong to 20 central amino acids in 3TR at 10°C and at pH 5.5. The chemical shift values are given with respect to 3-trimethylsilylpropionate (TSP) as the internal standard. The sequential assignment is obtained by combining TOCSY and NOESY data.

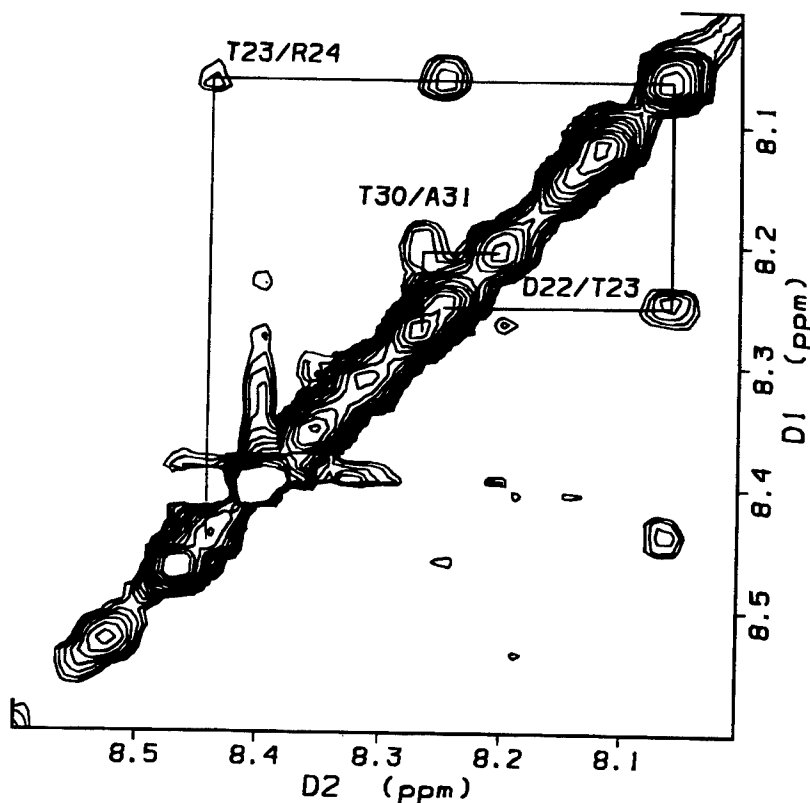
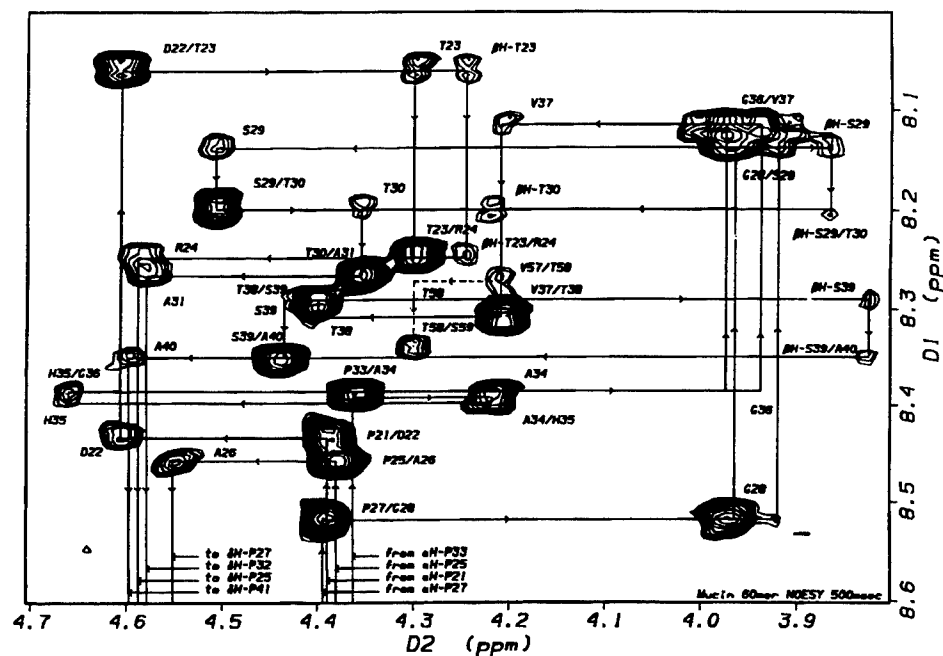
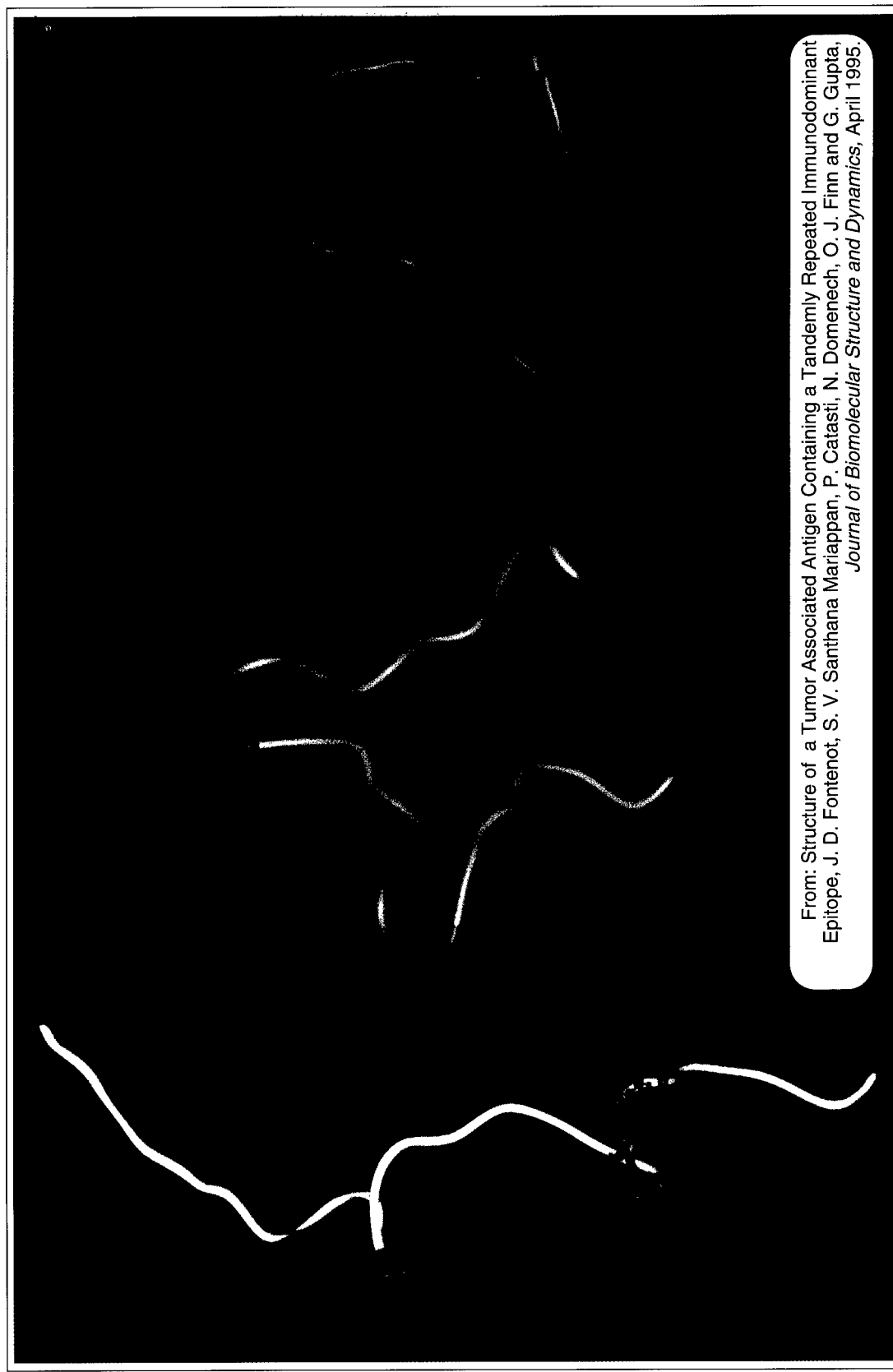


Figure 2: NOESY cross-section of (A) the finger print HN vs. H^a region and (B) HN vs. HN region of 5 mM 3TR in 90% H_2O +10% D_2O at 10°C (10 mM phosphate buffer with pH 5.5).



From: Structure of a Tumor Associated Antigen Containing a Tandemly Repeated Immunodominant Epitope, J. D. Fontenot, S. V. Santhana Mariappan, P. Catasti, N. Domenech, O. J. Finn and G. Gupta, *Journal of Biomolecular Structure and Dynamics*, April 1995.

Figure 5: Examples of possible low-energy NMR structures of 3TR Muc-1 that vary primarily in *end-to-end* length (Re) and the orientation of the knobs relative to one another. (A) (left) The elongated structure with the knobs protruding from the same face of the molecule. $Re = 82\text{\AA}$ in this structure. (B) (middle) A representative structure with intermediate $Re (= 46\text{\AA})$ with the knobs facing opposite directions. (C) (right) A compactly folded form of Muc-1 with $Re = 5\text{\AA}$ have the knobs oriented toward two corners of an equilateral triangle.

the NOESY cross-sections (mixing time = 500 ms) for HN vs. H^{α} and HN vs. HN regions. Several features of an ordered structure are evident from Figure 2. For example, excepting a few residues at the N- and C-terminal, the same residues in different repeats experience the same chemical shift environment and also show equivalent inter-residue NOE patterns as would be expected for a tandemly repeating structure.

Local Structure of the Immunodominant Epitope

The diagnostic NOE pattern of importance is the one involving the immunodominant epitopes, i.e., A20-P21-D22-T23-R24 and A40-P41-D42-T43-R44. The sequential $H^{\alpha}(i)$ -HN($i+1$) and HN(i)-HN($i+1$) NOEs are indicative of a type-II β -turn of APDTR centered around P and D (36). Figure 2A shows the strong $H^{\delta}(P)$ - $H^{\alpha}(A)$ and $H^{\alpha}(P)$ -HN(D) NOEs that are characteristics of a tight type-II β -turn; in addition, Figure 2B shows the inter-residue HN(D)-HN(T) NOEs are consistent with a type-II β -turn. In contrast, a type I β -turn at this position should produce weak sequential $H^{\alpha}(i)$ -HN($i+1$) and $H^{\alpha}(i+1)$ -HN($i+2$) NOEs (37). This pattern of NOEs for the type II β -turn in the immunodominant epitopes A20-P21-D22-T23-R24 and A40-P41-D42-T43-R44 of 3TR are the same as described for the principal neutralizing determinant (PND) with sequence GPGR that is located inside the third variable (V3) loop of the human immunodeficiency virus type I (HIV-1). Incidentally, HIV GPGR is favored to form a stable type II β -turn (23, 38). For example, the NOE pattern observed for the PND fragment of HIV (GPGR) is strong $H^{\delta}(P(i+1))$ - $H^{\alpha}(G(i))$, $H^{\alpha}(P(i+1))$ -HN($G(i+2)$), medium/strong $H^{\alpha}(G(i+2))$ -HN($R(i+3)$) and medium/weak HN($G(i+2)$)-HN($R(i+3)$) (22, 39). The same pattern of NOEs are observed for the mucin immunodominant epitope involving residues APDTR as shown Figures 2A and 2B. The NMR observations of a type II β -turn in the HIV-PND sequence (GPGR) were verified subsequently with crystallographic evidence (40). Wilmot and Thornton, using a set of 59 protein crystal structures, found a sequence preference for type II β -turns that included proline in the $i+1$ position followed by either G or N in the $i+2$ position (38). When N is present at the $i+2$ position, the (ϕ_{i+2}, ψ_{i+2}) values shift to the region 3 of the Ramachandran plot with $\phi_{i+2}=60\pm30$ and $\psi_{i+2}=60\pm30$ from the ideal values of (90, 0), which is stereochemically only possible for G. Such a departure from the type II β -turn helps to accommodate the heavier sidechain of N. This departure also weakens the $C=O(i)\cdots HN(i+3)$ H-bond. But the energetic loss due to the weakening of the H-bond in the type II β -turn with N at the ($i+2$) position is always compensated by the electrostatic interactions of the N with neighboring residues (38). The type II β -turns (APDTR) in 3TR also show the experimental evidence for the formation of a salt bridge through the side chains of D22/R24 and D44/R44. We could observe R side chain NH_2 group signals that were broad but low-field shifted to 10 ppm in the spectrum of 3TR in water as measured by the "jump and return" method (41) (absent in the spectrum of 1TR). This low-field shift of the R side chain NH_2 proton signals demonstrates partial burial of these groups from the solvent. Salt bridge formation with the COO^- group of D and $NH_2+-C-NH_2+$ group of R is indeed possible when the immunodominant epitope APDTR takes a tight turn as described above. Scanlon et al. also found evidence for salt bridge formation between the side chains of D2 and R4 using a 20 residue peptide dissolved in dimethyl sulfoxide (34). Although, Scanlon et al. propose a

pattern observed for the PND fragment of HIV (GPGR) is strong $H^{\delta}(P(i+1))$ - $H^{\alpha}(G(i))$, $H^{\alpha}(P(i+1))$ - $HN(G(i+2))$, medium/strong $H^{\alpha}(G(i+2))$ - $HN(R(i+3))$ and medium/weak $HN(G(i+2))$ - $HN(R(i+3))$ (22, 39). The same pattern of NOEs are observed for the mucin immunodominant epitope involving residues APDTR as shown Figures 2A and 2B. The NMR observations of a type II β -turn in the HIV-PND sequence (GPGR) were verified subsequently with crystallographic evidence (40). Wilmot and Thornton, using a set of 59 protein crystal structures, found a sequence preference for type II β -turns that included proline in the $i+1$ position followed by either G or N in the $i+2$ position (38). When N is present at the $i+2$ position, the (ϕ_{i+2}, ψ_{i+2}) values shift to the region 3 of the Ramachandran plot with $\phi_{i+2} = 60 \pm 30$ and $\psi_{i+2} = 60 \pm 30$ from the ideal values of (90, 0), which is stereochemically only possible for G. Such a departure from the type II β -turn helps to accommodate the heavier sidechain of N. This departure also weakens the $C=O(i) \cdots HN(i+3)$ H-bond. But the energetic loss due to the weakening of the H-bond in the type II β -turn with N at the $(i+2)$ position is always compensated by the electrostatic interactions of the N with neighboring residues (38). The type II β -turns (APDTR) in 3TR also show the experimental evidence for the formation of a salt bridge through the side chains of D22/R24 and D44/R44. We could observe R side chain NH_2 group signals that were broad but low-field shifted to 10 ppm in the spectrum of 3TR in water as measured by the "jump and return" method (41) (absent in the spectrum of 1TR). This low-field shift of the R side chain NH_2 proton signals demonstrates partial burial of these groups from the solvent. Salt bridge formation with the COO^- group of D and $NH_2^+-C-NH_2^+$ group of R is indeed possible when the immunodominant epitope APDTR takes a tight turn as described above. Scanlon et al. also found evidence for salt bridge formation between the side chains of D2 and R4 using a 20 residue peptide dissolved in dimethyl sulfoxide (34). Although, Scanlon et al. propose a type I turn at (P1-D2-T3-R4) with D2 and T3 as the central residues of the turn. The presence of a strong $H^{\alpha}(D2)$ - $HN(T3)$ disfavors such a turn. However, a type I (P1-D2-T3-R4) could conceivably co-exist with the type II turn at A0-P1-D2-T3. Regardless, the presence of a turn of any kind creates a protruding knob at A0-P1-D2-T3-R4. The absence of this tight turn in 1TR clearly indicates that residues at the TR interface contribute to the formation of a well defined APDTR turn. Finally, when all three APDTR segments in 3TR were replaced by (GPGR) of the HIV-1 PND we observed that the latter was structurally isomorphous with same sequence inside the V3 loop and with the APDTR of 3TR. All these data are consistent with a type II β -turn of the APDTR sequence (42).

Figure 3 shows the finger-print region of the DQF-COSY data; the $J(HN-H^{\alpha})$ coupling constants in the figure reveal ϕ -values in the range of ± 80 to 180. The $J(HN-H^{\alpha})$ coupling constants for the internal tandem repeat of Muc-1 3TR are as follows D(9.3), T(8.9), R(13.5), A(8.3), G(11.0), S(8.7), T(9.2), A(13.5), A(7.7), H(9.9), G(12.1), V(8.9), T(8.9), S(8.3), A(9.8). The $J(HN-H^{\alpha})$ coupling constants for the terminal residues are as follows R4(9.8), T58(8.4), S59(11.0), A60(8.1). The D2 cross peak was not observed in this cross section. The J coupling constants are estimated using FELIX software; the J coupling values are therefore approximate and only provide the upper bounds.

Knob-Like Structures of Muc-1 3TR

As explained in *Methods*, NOESY data at 200 and 500 ms of mixing are used for distance

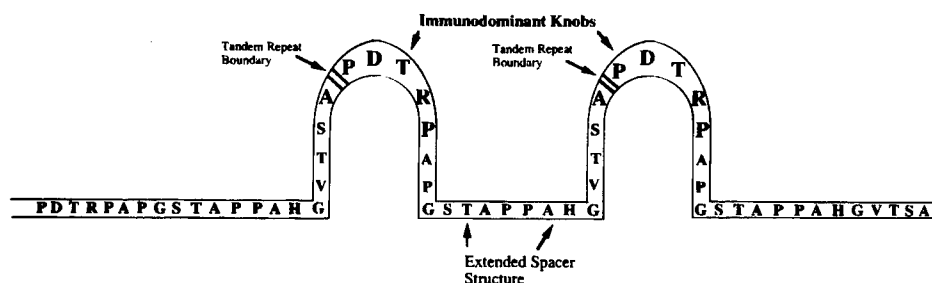


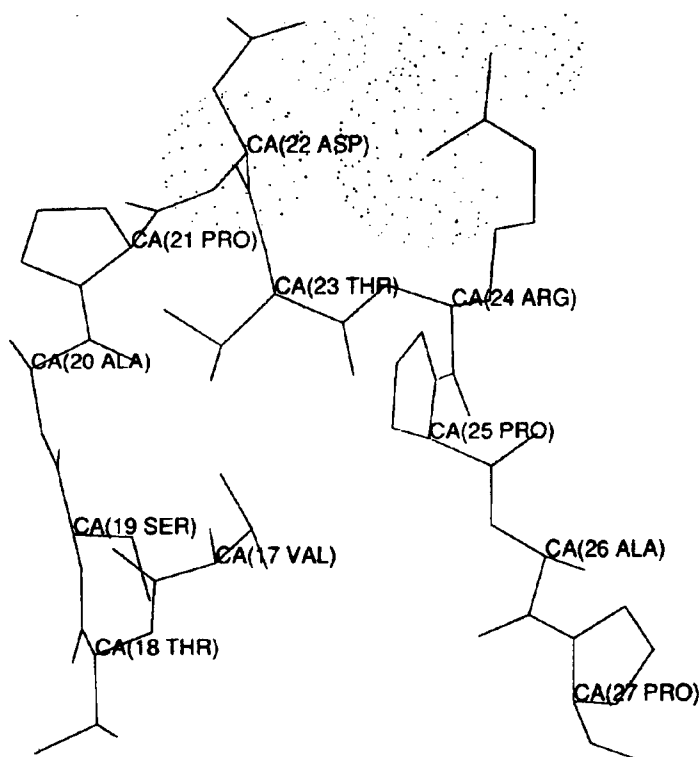
Figure 4: Conserved structural features of the tandem repeat domain of Muc-1 include; Repeating and protruding knob-like structures consisting of sequential reverse turns that span tandem repeat interfaces (residues 17-27, 37-47), an extended region composed of polypoline II and β -strand structure (residues 10-15, 30-35, 50-55). The N- and C- terminal 2-3 residues are unordered due to the absence of adjoining tandem repeats.

are used for determining the $J(\text{HN}-^a\text{H})$ coupling constants required for estimating various ϕ angles. A total of 220 distance constraints are derived using a full-matrix NOESY simulation (22); in this method all orders of NOE are considered as data not merely the primary NOEs. Therefore, the spin-diffusion problem (i.e. the presence of higher order NOEs) is not a problem but are included in the data set. The 45 torsional and the 220 distance constraints are then incorporated in a Monte Carlo (MC) simulated annealing procedure to sample different 3TR conformations that agree with the NMR data (26, 43). Simulated annealing constrained to the NMR data results in 120 energy-minimized structures for 3TR. All of the structures share critical features as shown in Figure 4 — i.e., (1) a solvent exposed protruding knob with the immunodominant APDTR turn at the tip, (2) the protruding knob spanned residues V17-P27/V37-P47 thereby including the last 4 residues from the previous repeat, (3) the knobs are connected by an extended spacer structure (comprising of β -strand and poly-proline conformations) for residues 30-35, and (4) extended conformations are also detected for the residues 10-15 and 50-55. The conformational difference among the sampled structures is the difference in the end-to-end length (R_e) and the relative spatial orientation of the two knobs. The predominant portion of the sampled structures show $R_e = 70\text{-}90 \text{ \AA}$ (data not shown). The large extent of variation in the R_e ($= 30\text{-}90 \text{ \AA}$) prompted us to examine the nature of variation in the backbone torsion angles of the sampled structures of 3TR that finally produce the variation in the R_e . Table II lists the average values and the standard deviations of the backbone and side chain torsion angles for the 20 central residues in 3TR. It is seen that the standard deviations in the backbone torsion angles are below 10 degrees excepting for a few residues. Therefore, small changes in the backbone torsion angles correlated over several residues gives rise to large variations in R_e . The sampled structures all show similar agreement with the NMR data. Figure 5 shows three such structures of 3TR for long, intermediate, and short R_e . Conserved structural features (1)-(4), described above, are retained in all three structures. However, tertiary chain-folding patterns are different in these structures. The structure of 3TR with the long R_e (Figure 5A) has the protruding immunodominant knobs arranged on a linear extended chain; the structure with the intermediate R_e has the knobs in opposite orientations (Figure 5B); the structure with the short R_e has the knobs pointing out from the two corners of a triangle (Figure 5C). These structures illustrate an unusual characteristic of Muc-1, the ability to present multiple copies of an antigenic determinant with flexibility in the intervening sequences.

Table II
Torsion Angles and Standard Deviations for the Internal Tandem Repeat

Residue	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
A 20	-64 (25)	153 (9)	172 (8)	-55 (5)						
P 21	-75 (0)	161 (1)	177 (6)							
D 22	53 (3)	54 (7)	178 (5)	-171 (2)	57 (1)					
T 23	-123 (5)	30 (4)	-177 (3)	-48 (1)	-66 (1)	168 (1)				
R 24	48 (4)	72 (2)	-175 (4)	-63 (0)	179 (0)	176 (30)	94 (34)	-2 (2)	-1 (1)	0 (0)
P 25	-75 (0)	169 (5)	177 (4)							
A 26	-146 (24)	161 (1)	157 (2)	-64 (4)						
P 27	-75 (0)	131 (26)	178 (3)							
G 28	-86 (21)	66 (6)	177 (3)							
S 29	-117 (12)	9 (6)	-177 (1)	-177 (1)	-50 (1)					
T 30	-94 (8)	68 (3)	177 (1)	55 (2)	71 (1)	64 (1)				
A 31	-150 (0)	81 (20)	178 (4)	-179 (0)						
P 32	-75 (0)	171 (0)	163 (1)							
P 33	-75 (0)	81 (17)	178 (2)							
A 34	-75 (5)	151 (2)	-179 (1)	61 (0)						
H 35	-137 (2)	37 (4)	178 (2)	-53 (3)	119 (4)					
G 36	-67 (6)	140 (11)	-173 (3)							
V 37	-157 (3)	148 (4)	174 (5)	63 (2)	-38 (2)	-67 (2)				
T 38	-78 (4)	78 (7)	178 (6)	-54 (2)	160 (5)	179 (2)				
S 39	-162 (4)	164 (6)	175 (3)	-63 (30)	-23 (27)					

The average values and the standard deviations (in degree) of the backbone and sidechain torsion angles of the 20 central residues as obtained by analyzing 120 sampled structures Muc-1 3TR. 220 inter-proton distance constraints used in the MC simulated annealing are included in supplementary materials. The rms deviations with respect to these constraints were 0.18 ± 0.01 . The deviations in energy (KCal/Mole) in different contributions were as follows: electrostatic energy (dielectric constant = 80) = 4 ± 1 ; non-bonding energy = -210 ± 10 ; torsional energy = 25 ± 5 . Therefore, the variation in the total energy was -181 ± 10 . Note that the central P21 and D22 of the (A20-P21-D22-T23) turn show less flexibility than the central residues P27 and G28 of the A26-P27-G28-S29 turn.



Antigenic tip of Muc-1

Figure 6: Structure of the immunogenic knob comprising of the central eleven residues, (V17-T18-S19-A20-P21-D22-T23-R24-P25-A26-P27). Note that residues 37-47 in the same molecule show almost identical structure. All the low-energy NMR-derived structures show similar structures for residues 17-27 or 37-47.

Figure 6 shows the structure of the immunogenic knob comprising of the central eleven residues, (V17-T18-S19-A20-P21-D22-T23-R24-P25-A26-P27). Such a knob configuration is common to all the low energy structures of Muc-1 3TR. The local structure of the residues 18-26 is virtually indistinguishable to that of residues 38-46. A close examination of the knob reveals a double bend at the tip. Two bends are centered around P21-D22 and D22-T23. These two consecutive bends bring the sidechain of R24 in favorable electrostatic contacts with the backbone C=O and sidechain COO- of D22. This also results in the burial of the T23 sidechain. Such a conformation is consistent with the NH-NH NOEs between (D22&T23) and (T23&R24) and strong H^α-NH NOES between (P21&D22). One might argue that such NOEs may also due to a mixture of type II turn centered at P21-D22 and a type I turn centered at D22-T23. If that is true, the conformational equilibrium must be fast on the NMR time-scale because we do not observe resonance doubling for the residues 21-24. The immunogenic knob shown in Figure 6 retains a type II turn at P21-D22 and T23 shows the same conformation as in a type I turn (see Table II). Nevertheless, what is most important in the structure is the surface exposure of the critical APDTR residues.

Discussion

The results of the present 2D NMR investigations of 1TR and 3TR reveal that no significant structure exists in one copy of the mucin tandem repeat (PDTRPAPGS-TAPPAHGVTS), and that the peptide with three tandem repeats exhibits flexible structures with well defined knob-like structure centered around the antigenic (APDTR) site. Previously, we compared peptides containing one, two, and three-tandem repeats and showed changes occurring throughout the one-dimensional NMR spectrum (44). We concluded that the secondary structure develops as additional tandem repeats are added. The NMR investigations were further supported by results with circular dichroism in which the ratio of the molar ellipticity at 198 nm of one, three or five tandem repeats to that of a control peptide was 3.2 for one tandem repeat, 13.6 for three tandem repeats and 21.0 for 5.25 tandem repeats (20). This non-linear increase in molar ellipticity can be attributed to the development of structure as the number tandem repeats increases. Furthermore, we presented a model for the structure of Muc-1 tandem repeats that was suggested by Matsushima et al. for proline rich tandem repeats (44, 45). We predicted an elongated structure composed of repeating type I turns. The actual structure for the Muc-1 tandem repeat domain is shown in Figure 5, and differs significantly from the earlier model in that it contains repeating solvent exposed protruding knobs spanning the tandem repeat interface. The knob structures (V17-P27/V37-P47) are crested by the immunodominant APDTR turn at the tips, and sequential knobs are connected by extended spacer structures for residues T30-A35, and T10-A15 and T50-55A. The physical dominance of the knob structure at the tandem repeat interface was surprising and yet serves to explain much of the observed immunoreactivity (10, 13, 16).

The stable turn structure and protruding nature of the repeating epitope APDTR could obscure access to the remainder of the molecule by B and T cell antigen receptors and help explain the observed antibody immunodominance of APDTR and MHC unrestricted activation of T cells by unglycosylated mucin on tumors (10, 16). A parallel MHC restricted T cell response could be generated to peptides from the processed Muc-1 tandem repeats (17). A restricted T cell response would depend on the presence of proteolytic processing sites within Muc-1 and the ability of processed peptides to bind to class I MHC proteins, and finding restricted T cells would not be incompatible with the interpretation of these data (46).

The enzymatic o-glycosylation pattern of the T residues of 3TR is particularly relevant to the knob-like structures shown in Figures 4 and 5. 2D NMR studies recently carried out in our laboratory in collaboration with Dr. Henrick Clausen shows that enzymatic glycosylation of 3TR produces T residues o-glycosylated with β -D-N-acetylgalactosamine (GalNAc) in TS and ST sequence doublets and not at the T of the APDTR turn. The tight turn at APDTR where the OH group of T is buried inside the turn could explain the protection of this T from glycosylation. An alternative explanation for the lack of glycosylation at the T within the PDTRP is that the primary sequence may not signal for glycosylation at this position (47). In general, the elongated structures shown in Figure 5A can easily accommodate glycosylation at the potential T and S sites without steric problems. This implies that a mucin molecule containing at least thirty-six 20-residue tandem repeats (when glycosylated)

could span more than 1,000 Å from the cell surface and thereby function as an anti-adhesion molecule (48). This is the first study that shows the protein core of the TR domain intrinsically contains an elongated structure studded with immunodominant knobs that protect the T residue in the APDTR but allows o-glycosylation of the remaining solvent exposed T and S residues and that this could lead to further stabilization of elongated structures. Muc-1 in the extended state cannot simply be accommodated inside the cell. Therefore, the types of compaction shown in Figures 5B and 5C become relevant in the context of packaging mucin inside the cell.

The structural studies presented here assume a special importance because underglycosylated mucin repeats are tumor associated antigens in breast (7), pancreatic (16), and ovarian (17) cancers. NMR data on 1 and 3 Muc-1 tandem repeats clearly show how the immunodominant knob structures are preserved and presented only when multiple tandem repeats are assembled in space. The knob-like structures on intact Muc-1 may obscure access to the extended portions of the molecule by antibody receptors on the surface of B-lymphocytes during the induction of an immune response. This may explain the immunodominance of the accessible tip (APDTR) of the knobs. Preliminary work further indicates that higher numbers of tandem repeats ($n=5$) enhance the rigidity and presentation of the knobs. Hence it is not surprising that previous studies have shown that increasing the number of tandem repeats from 3 to 5 results in a nonlinear enhancement in the binding of antibodies in the serum of cancer patients (15). Therefore, the combination of structural and immunological studies of this type will be important for the rational design of vaccines and immunotherapies for Muc-1 positive tumors

Acknowledgments

We thank Dr. Cliff Unkefer and Dr. Jill Trewella of Los Alamos National Laboratory, and Dr. Dave Scott of the University of Iowa for use of their 500 MHz NMR spectrometers. We also thank Dr. Ron Montelaro for support during the early phase of this work. This work was supported by the US Army Grant RH03 to GG.

References and Footnotes

1. S. Gendler, J. Taylor-Papadimitriou, T. Duhlig, J. Rothbard and J.A. Burchell, *J. Biol. Chem.* 263, 12820-12823 (1988).
2. J.R. Gum, J.C. Byrd, J.W. Hicks, N.W. Toribara, D.T.A. Lampert and T.S. Kim, *J. Biol. Chem.* 264, 6480-6487 (1989).
3. J.R. Gum, J.W. Hicks, D.M. Swallow, R.L. Lagace, J.C. Byrd, D.T.A. Lampert, B. Siddiki and Y.S. Kim, *Biochem. Biophys. Res. Commun.* 171, 407-415 (1990).
4. N. Porchet, N.V. Cong, J. Dufosse, J.P. Audie, V. Guyonnet-Duperat, M.S. Gross, C. Denis, P. Degand, A. Bernheim and J.P. Aubert, *Biochem. Biophys. Res. Commun.* 175, 414-422 (1991).
5. J. Hilkens, F. Buijs and M. Ligtenberg, *Cancer Res.* 49, 786-793 (1989).
6. A. Girling, J. Bartkova, J. Burchell, S. Gendler, C. Gillet and J. Taylor-Papadimitriou, *Int. J. Cancer.* 43, 1072-1076. (1989).
7. K.R. Jerome, D.L. Barnd, K.M. Bendt, C.M. Boyer, J. Taylor-Papadimitriou, I.F.C. McKenzie, R.C.B. Jr. and O.J. Finn, *Cancer Res.* 51, 2908-2916 (1991).
8. K.R. Jerome, D. Bu and O.J. Finn, *Cancer Res.* 52, 5985-5990. (1992).
9. S. Sell, *Progress Path.* 21, 1003-1019. (1990).
10. J. Taylor-Papadimitriou, *Int. J. Cancer* 49, 1-5. (1991).
11. P.X. Xing, J.J. Tjandra, S.A. Stacker, J.G. Teh, C.H. Thompson, P.J. McLaughlin and I.F.C. McKenzie, *Immunol. Cell. Biol.* 67, 183-195 (1989).

12. J. Burchell, S. Gendler, J. Taylor-Papadimitriou, A. Girling, A. Lewis, R. Millis and D. Lampion, *Cancer Res.* 47, 5476-5482 (1987).
13. P.L. Devine and I.F.C. McKenzie, *BioEssays* 14, 619-625 (1992).
14. A. Rughetti, V. Turchi, C.A. Ghetti, G. Scambia, P.B. Panici, G. Ronucci, S. Mancuso, L. Frati and M. Nuti, *Cancer Res.* 53, 2457-2459 (1993).
15. Y. Kotera, J.D. Fontenot, G. Pecher, R.S. Metzgar and O.J. Finn, *Cancer Res.* 54, 2856-2860 (1994).
16. D.L. Barnd, M. Lan, R. Metzgar and O.J. Finn, *Proc. Natl. Acad. Sci. USA*, 86, 7159-7163 (1989).
17. C.G. Ioannides, B. Fisk, K.R. Jerome, T. Irimura, J.T. Wharton and O.J. Finn, *J. Immunol.* 151, 3693-3703 (1993).
18. K.R. Jerome, N. Domenech and O.J. Finn, *J. Immunol.* 151, 1654-1662 (1993).
19. O.J. Finn, *Biotherapy* 4, 239-249 (1992).
20. J.D. Fontenot, O.J. Finn, N. Dales, P.C. Andrews and R.C. Montelaro, *Pept. Res.* 6, 330-336 (1993).
21. G.N. Ramachandran and V. Sasisekharan, *Adv. Prot. Chem.* 23, 283-437 (1968).
22. G. Gupta, G.M. Anantharamaiah, D.R. Scott, J.H. Eldridge and G. Meyers, *J. Biomol. Struct. and Dynam.* 11, 345-366 (1993).
23. G. Gupta and G. Meyers, *Computer analysis of HIV epitope sequences*, 1-99-105 Pasteur Vaccins, Paris, 1990.
24. G. Nemethy, M.S. Pottle and H.A. Scheraga, *J. Phys. Chem.* 87, 1883-1887 (1983).
25. R. Fletcher, *Practical methods in optimization 1*, John Wiley & Sons, New York, 1984.
26. S. Kirkpatrick, C.D.G. Jr. and M.P. Vecchi, *Science* 220, 671-680 (1983).
27. G. Gupta and G. Meyers, *Analyses of various folding patterns of the HIV-1 loop*, (Birkhauser, Boston, 1994).
28. A. Bax and D.G. Davis, *J. Magn. Res.* 63, 207-213 (1985).
29. A.A. Bothner-BY, R.L. Stephens, J.-m. Lee, C.D. Warren and R.W. Jeanloz, *J. Am. Chem. Soc.* 106, 811-813 (1984).
30. J. Jeener, B.H. Meier, P. Bachman and R.R. Ernst, *J. Chem. Phys.* 71, 4546-4553 (1979).
31. S. Macura and R.R. Ernst, *Mol. Phys.* 41, 95-117 (1980).
32. M.R. Price, F. Hudecz, C. O'Sullivan, R.W. Baldwin, P.M. Edwards and S.J.B. Tendler, *Molec. Immunol.* 27, 795-802 (1990).
33. S. Tendler, *Biochem. J.* 267, 733-737 (1990).
34. M.J. Scanlon, S.D. Morley, D.E. Jackson, M.R. Price and S.J.B. Tendler, *Biochem. J.* 284, 137-144 (1992).
35. S.J.B. Tendler, M.J. Scanlon and M.R. Price, *Protein and Peptide Letters* 1, 39-43 (1994).
36. D.R. Rose, L.M. Gierasch and J.A. Smith, *Adv. Prot. Chem.* 37, 1-109 (1985).
37. K. Wuthrich, *NMR of proteins and nucleic acids*, John Wiley & Sons, New York, 1986.
38. C.M. Wilmot and J.M. Thornton, *J. Mol. Biol.* 203, 221-232 (1988).
39. K. Chandrasekhar, A.T. Profy and H.J. Dyson, *Biochemistry* 30, 9187-9194 (1991).
40. J.B. Ghiara, E.A. Stura, R.L. Stanfield, A.T. Profy and I.A. Wilson, *Science* 264, 82-85 (1994).
41. P. Plateau and M. Gueron, *J. Am. Chem. Soc.* 104, 7310-7311 (1982).
42. J.D. Fontenot, J.M. Gatewood, S.V.S. Mariappan, C.P. Pau, B.S. Parekh, J.R. George and G. Gupta, *Proc. Natl. Acad. Sci. USA*, In Press (1994).
43. F.D.M. Veronese, M.S.R. Jr., G. Gupta, M. Robert-Guroff, C. Boyer-Thompson, R.C. Gallo and P. Lusso, *J. Biol. Chem.* 268, 25894-25901 (1993).
44. J.D. Fontenot, N. Tjandra, D. Bu, C. Ho, R.C. Montelaro and O.J. Finn, *Cancer Res.* 53, 5386-5394 (1993).
45. N. Matsushima, C.E. Creutz and R.H. Kretsinger, *Proteins: Structure, Function and Genetics.* 7, 125-155 (1990).
46. E.E. Secarz, P.V. Lehmann, A. Ametani, G. Benichou, A. Miller and K. Moudgil, *Annu. Rev. Immunol.* 11, 729-766 (1993).
47. I. Nishimori, F. Perini, K.P. Mountjoy, S.D. Sanderson, N. Johnson, R.L. Cerny, M.L. Gross, J.D. Fontenot and M.A. Hollingsworth, *Cancer Res.* 54, 3738-3744 (1994).
48. M.J.L. Ligtenberg, F. Buijs, H.L. Vos and J. Hilken, *Cancer Res.* 52, 2318-2324 (1992).

Date Received: December 10, 1994

Communicated by the Editor Ramaswamy H. Sarma

Loss of a Neutralizing Epitope by a Spontaneous Point Mutation in the V3 Loop of HIV-1 Isolated from an Infected Laboratory Worker*

(Received for publication, May 20, 1993, and in revised form, August 16, 1993)

Fulvia di Marzo Veronese^{‡§}, Marvin S. Reitz, Jr.[¶], Goutam Gupta[¶], Marjorie Robert-Guroff[¶], Cynthia Boyer-Thompson[‡], Audrey Louie[¶], Robert C. Gallo[¶], and Paolo Lusso[¶]

From the Department of Cell Biology, [‡]Advanced BioScience Laboratories, Inc., Kensington, Maryland 20895, the [¶]Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, and the [§]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

The third hypervariable region, or V3 loop, represents the principal neutralizing domain of the gp120 envelope glycoprotein of human immunodeficiency virus type 1 (HIV-1). Sequential viral isolates from a laboratory worker (LW) accidentally infected with HIV-1_{IIIB} in 1985 were analyzed using type-specific neutralizing monoclonal antibodies directed to the V3 loop. A single amino acid substitution, Ala → Thr at position 21 in the V3 loop of HIV-1_{LW} isolated in 1987, was shown to determine the loss of the neutralizing epitope recognized by one of the monoclonal antibodies (M77). However, this antibody efficiently recognized linear V3 loop peptides containing either the Ala or Thr residue at position 21, indicating that a local change in conformation was responsible for the epitope loss in the native gp120. Molecular modeling studies, experimentally supported by different amino acid replacements at position 21, indicated that the Ala → Thr substitution leads to a drastic change in the domain of the V3 loop, which contains the complementary surface for antibody binding. These results provide evidence for the first time that a conformation-dependent epitope within the V3 loop of HIV-1 is involved in the generation of neutralization escape mutants *in vivo*.

The primary translational product of the envelope gene of the human immunodeficiency virus type 1 (HIV-1)¹ is a gp160 which is processed to gp120 and gp41 as the external and transmembrane glycoproteins, respectively (Veronese *et al.*, 1985; Allan *et al.*, 1985). Because of its location on the surface of the virion and of the infected cell, gp120 contains epitopes naturally accessible to the immune system and thus represents a prime candidate for the development of vaccine strategies. Indeed, a domain of gp120, situated in the third hypervariable region, has been identified as dominant for development of high titered type-specific neutralizing antibodies against HIV-1 (Jahavarian *et al.*, 1989; Rusche *et al.*, 1988; Palker *et al.*, 1988). This principal neutralizing domain lies within a loop formed by a disulfide bridge between 2 conserved cysteines. A distinctive characteristic of this loop is the extensive degree of genetic variability found among individual isolates of HIV-1 (Robert-Guroff, 1990). This hypervariability, a likely consequence of

strong immune selective pressures *in vivo*, suggests that mutations in the primary amino acid sequence of the neutralizing loop become prevalent as viruses which escape neutralization are selected.

Different approaches have been followed in order to identify amino acid changes which are critical for neutralization. One approach involved the *in vitro* generation of neutralization escape mutant viruses under selective pressure of human sera from HIV-infected individuals (Robert-Guroff *et al.*, 1986) or monoclonal antibodies to the V3 loop (McKeating, 1989). *In vitro* studies with human neutralizing antibodies were the first to demonstrate that HIV mutants are selected by immune pressures (Robert-Guroff *et al.*, 1986). In addition, these studies identified a point mutation in gp41 (Ala → Thr replacement at position 582) which resulted in neutralization resistance (Reitz *et al.*, 1988) but was apparently not part of the antibody binding site (Wilson *et al.*, 1990), indicating the involvement of discontinuous epitopes in HIV neutralization. Molecular analysis of mutant viruses obtained by *in vitro* selection with monoclonal antibodies demonstrated amino acid changes within the V3 loop itself (McKeating *et al.*, 1989), but more interestingly also in regions distantly located from the loop, suggesting an interaction of this domain with other regions of gp120 (McKeating *et al.*, 1989). Another approach involved the analysis of sequential viral isolates obtained from chimpanzees experimentally infected with a laboratory strain of HIV-1 (HIV-1_{IIIB}). The *in vivo* passage of this isolate resulted in the generation of variant viruses which were resistant to neutralization by V3 loop-specific monoclonal antibodies (Nara *et al.*, 1990). Sequence analysis of the V3 loops from these mutants did not reveal any amino acid difference between the neutralization-resistant and -sensitive viruses, suggesting that the crucial changes did not directly involve the antibody binding site (Nara *et al.*, 1990).

In this report, we studied the immunological reactivity and neutralizing capability of two monoclonal antibodies to the V3 domain of HIV-1_{IIIB} with biologically active molecular clones from sequential viral isolates obtained from a laboratory worker who became accidentally infected with HIV-1_{IIIB} (Weiss *et al.*, 1988). This unique *in vivo* human model allowed us to identify a single change in the V3 loop primary amino acid sequence which is crucial for neutralization resistance by V3 loop-specific monoclonal antibodies. Moreover, we have performed molecular modeling studies, experimentally supported by amino acid replacements, to show that this single amino acid substitution may cause a drastic change in the domain of the V3 loop which provides the complementary surface for antibody binding.

EXPERIMENTAL PROCEDURES

Neutralization Assay of Cell-free Virus—Ascitic fluids from the two monoclonal antibodies were filtered through a 0.2-μm filter and di-

* This work was supported in part by Contracts NO1-CP-73723 and NO1-CP-73722 from the National Cancer Institute. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Laboratory of Tumor Cell Biology, Bldg. 37, Rm. 6B23, National Institutes of Health, Bethesda, MD 20892. Tel.: 301-496-4850; Fax: 301-496-8394.

¹ The abbreviations used are: HIV-1, human immunodeficiency virus type 1; LW, laboratory worker; HIV-1_{LW}, HIV-1 laboratory worker strain; kbp, kilobase pair(s); PCR, polymerase chain reaction.

luted in complete RPMI 1640 medium. Aliquots of the respective antibody were incubated with HIV-1_{IIIIB} virus stock containing 25,000 cpm reverse transcriptase activity in 100 μ l final volume for 60 min at 4 °C and then for 15 min at room temperature. CEM cells (1×10^6) were then added to the virus/antibody mixture and incubated for 60 min at 37 °C. The cells, still in the presence of virus and antibody, were supplemented with 2 ml of complete RPMI medium and transferred to six-well plates. Two ml of RPMI medium were added to each well after 24 h, and the reverse transcriptase activity of the supernatant was determined on the first day when syncytia were clearly visible by microscopic examination.

Syncytium Inhibition Assay—The syncytia assay was performed in 96-well plates by coculturing 1×10^6 uninfected cells with 5×10^3 virus-infected cells. Monoclonal antibodies were added at different dilutions to the mixture of cells, and the total volume of medium was adjusted to 0.2 ml. The plates were incubated for 40 h at 37 °C, and the number of multinucleated giant cells was determined by microscopic examination.

Radiolabeling of Cells and Immunoprecipitation of HIV-1 Proteins—Transfected HeLa-tat cells or infected SupT1 were suspended for 1 h at 37 °C in cysteine-free media. [³⁵S]Cysteine was then added to a final concentration of 200 μ Ci/ml, and the cells were incubated for 18 h at 37 °C.

After labeling the cells were washed in phosphate-buffered saline and disrupted in phosphate-buffered saline containing 0.5% NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate (SDS). The lysates were absorbed for 3 h at 4 °C with protein A-Sepharose and protein A-Sepharose bound to rabbit anti-mouse light chain antiserum and then clarified by ultracentrifugation. Radioimmunoprecipitation analyses were performed by the addition to 0.5 ml of labeled extract of either 10 μ l of serum from an HIV-infected individual

and 0.2 ml of a 10% suspension of protein A-Sepharose or 1 μ l of ascites containing the appropriate antibody and 0.2 ml of a 10% suspension of protein A-Sepharose bound to rabbit anti-mouse light chain antiserum. The samples were incubated for 18 h at 4 °C. Immunoprecipitates were collected by centrifugation, washed repeatedly, resuspended in Laemmli sample buffer (Laemmli, 1970), heated for 2 min at 90 °C, and analyzed by SDS-polyacrylamide gel electrophoresis.

Indirect Immunofluorescence and Fluorocytometric Analysis—Indirect immunofluorescence analysis was performed on live cells with primary antibodies at 5 μ g/ml and with fluorescein isothiocyanate-conjugated goat antiserum to murine IgG (Sigma) as a second layer antibody. Controls were incubated with a mouse IgG₁ irrelevant antibody, then stained with the same fluorescein isothiocyanate-conjugated antiserum. After staining, the cells were fixed with 2% paraformaldehyde and analyzed with a Facsan analyzer (Becton Dickinson Immunocytometry). The results are expressed as arbitrarily normalized histograms (i.e. relative number of cells versus fluorescence intensity). At least 10,000 events were accumulated in all the experiments.

Construction of V3 Loop Cassette and env Chimerae—A 2.7-kbp *SalI*-*Bam*HI restriction enzyme fragment from pHXB2gpt, a molecular clone of HIV-1_{IIIIB}, containing the V3 loop region, was subcloned into the cognate sites of gGEM4 (Promega, Madison, WI). A 0.6-kbp V3-containing *Bgl*II fragment from this subclone was further subcloned into pSP72 (Promega). Two PCR amplifications were performed on this fragment, such that one amplified fragment consisted of the region from the 5' *Bgl*II site to the 5' terminus of the V3 loop and contained *Mlu*I and *Pst*I sites introduced at the downstream end of the fragment by a tail on the antisense PCR primer. The other PCR fragment consisted of the region from the 3' end of the V3 loop, where *Pst*I and *Hpa*I sites were introduced at the upstream end of the fragment by a tail on the sense PCR primer. The PCR fragments were purified, digested with *Bgl*II plus *Pst*I, and coligated into pSP72. The 0.5-kbp *Bgl*II fragment from the resultant clone was ligated with the 5.1-kbp *Bgl*II fragment of the *SalI*-*Bam*HI *env* subclone of pHXB2gpt. The resulting plasmid has unique *Mlu*I and *Hpa*I restriction sites, whose introduction did not alter the amino acid sequence, forming a cassette into which oligonucleotides representing V3 coding regions and containing an *Mlu*I 4-base overhang at one end and a blunt end at the other end can be ligated. This construction is summarized in Fig. 4. The V3 loop coding region of LW12.3 (Lori *et al.*, 1992), an infectious molecular clone from the virus isolated in 1987 from a laboratory worker accidentally infected with HIV-1_{IIIIB} and hereinafter designated LW87, was synthesized as a set of four partially overlapping nucleotides, phosphorylated with T4 kinase, annealed, and ligated into the *Mlu*I and *Hpa*I sites of the cassette plasmid. The resultant 2.7-kbp *SalI*-*Bam*HI fragment, now containing

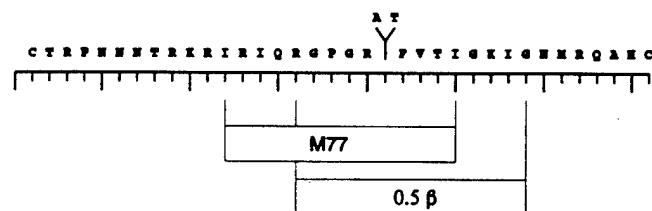


FIG. 1. Scheme depicting the sequence for the V3 loop of the pHXB2 molecular clone of HIV-1_{IIIIB} and the location of the Thr (LW virus from 1987) for A (HXB2) substitution. Also designated are the binding sites for M77 and 0.5 β monoclonal antibodies.

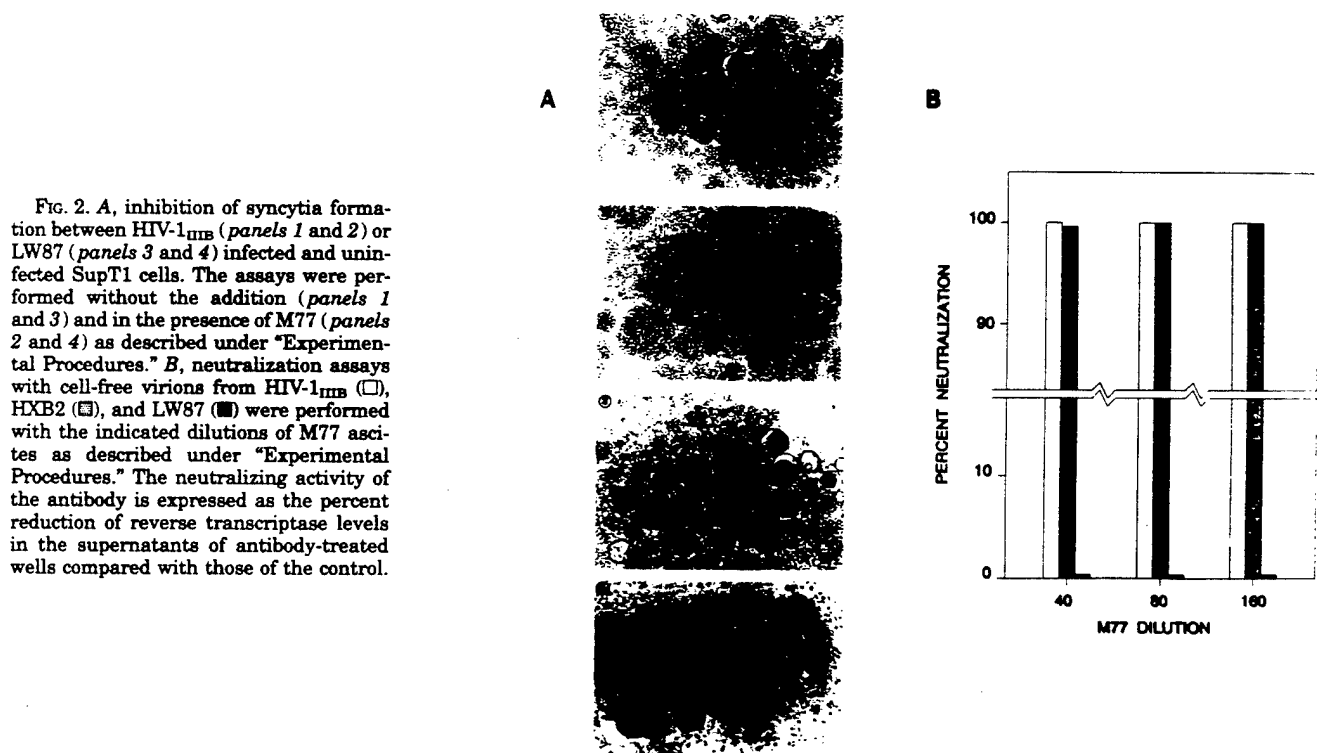


FIG. 2. A, inhibition of syncytia formation between HIV-1_{IIIIB} (panels 1 and 2) or LW87 (panels 3 and 4) infected and uninfected SupT1 cells. The assays were performed without the addition (panels 1 and 3) and in the presence of M77 (panels 2 and 4) as described under "Experimental Procedures." B, neutralization assays with cell-free virions from HIV-1_{IIIIB} (□), HXB2 (▤), and LW87 (■) were performed with the indicated dilutions of M77 ascites as described under "Experimental Procedures." The neutralizing activity of the antibody is expressed as the percent reduction of reverse transcriptase levels in the supernatants of antibody-treated wells compared with those of the control.

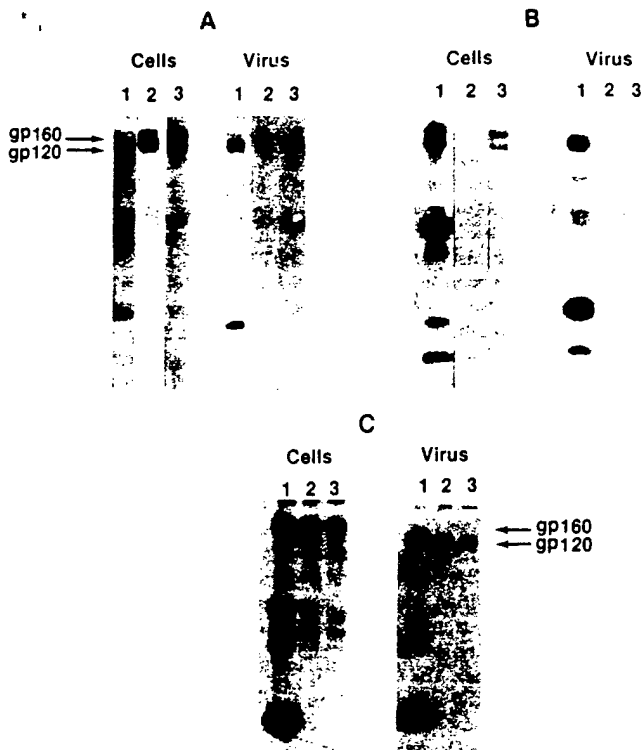


FIG. 3. Reactivity of M77 and 0.5 β monoclonal antibodies with gp120. SupT1 cells infected with HIV-1_{III}B (A) or LW87 (B) and HeLa-tat cells transfected with pHXB2 (env LW85-2) (C) were labeled with [³⁵S]cysteine. The clarified lysates from cells and virus containing supernatants were then immunoprecipitated with an HIV-1 positive human serum (lane 1), M77 (lane 2), and 0.5 β (lane 3).

the V3 loop of LW87, was reintroduced into pHXB2gpt, which was then used as a source for infectious chimeras of HXB2 with the LW87 V3 loop (HXB2 [V3LW87]). Infectious clones of HXB2 containing V3 loops into which Ser or Ile replaced A in the 21 position were constructed in the same way as HXB2 (V3LW87) except for the use of slightly different oligonucleotides bearing the appropriate changes.

DNA Transfection.—The infectious molecular clones pLW87, pHXB2, the chimeric construct of pHXB2 with the LW85-2 env (HXB2 (env LW85-2)), and pHXB2 with the LW87 loop (HXB2 [V3LW87]) were transfected into HeLa-tat cells by calcium phosphate coprecipitation (Chen, 1987). Briefly, an appropriate mixture of DNA, sterile water and CaCl₂ was added dropwise, to a 2 × solution of 274 mM NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄·7H₂O, 12 mM dextrose, and 42 mM Hepes, pH 7.1, while applying an airstream on the surface. After 10 min the mixture was pipetted gently into a flask of log phase-growing HeLa-tat and incubated with the cells overnight at 37 °C. The next morning the cells were washed two times in serum-free medium, then fed with complete medium. Twenty-four hours after transfection, the cells were starved in cysteine-free medium in preparation for the metabolic labeling.

Peptide Binding.—A peptide enzyme-linked immunosorbent assay (Robert-Guroff *et al.*, 1992) was used to monitor antibody binding to V3 loop peptides. The peptides were obtained from Multiple Peptide Systems, Richmond, CA and included the central 24 amino acids of LW87 (NNTRKRIRIQRGPGRTFVTIGKIG(C)) and BH10 (HIV-1_{III}B) (NNTRKSIRIQRGPGRAFVTIGKIG(C)) V3 loop sequences. A peptide representing amino acids 88–115 of the HIV-1_{III}B gag sequence (VHQRIEIKDKTEALDKIEEQNKSKKKA) served as negative control. Peptides were bound at 2 μ g/well to 96-well Immulon I microtiter plates (Dynatech, Chantilly, VA) pretreated with polylysine (5 μ g/well) and 0.1% glutaraldehyde. Specific binding of serial dilutions of monoclonal antibodies were detected using a goat anti-mouse IgG peroxidase conjugate.

Molecular Modeling.—The methodology used for the molecular modeling involved the following steps. First, the amino acid sequence of the V3 loop was converted into secondary structural state using a prediction algorithm (Deleage and Roux, 1989; Gupta and Myers, 1990) so that each amino acid was assigned one of the four secondary states, *i.e.* helix

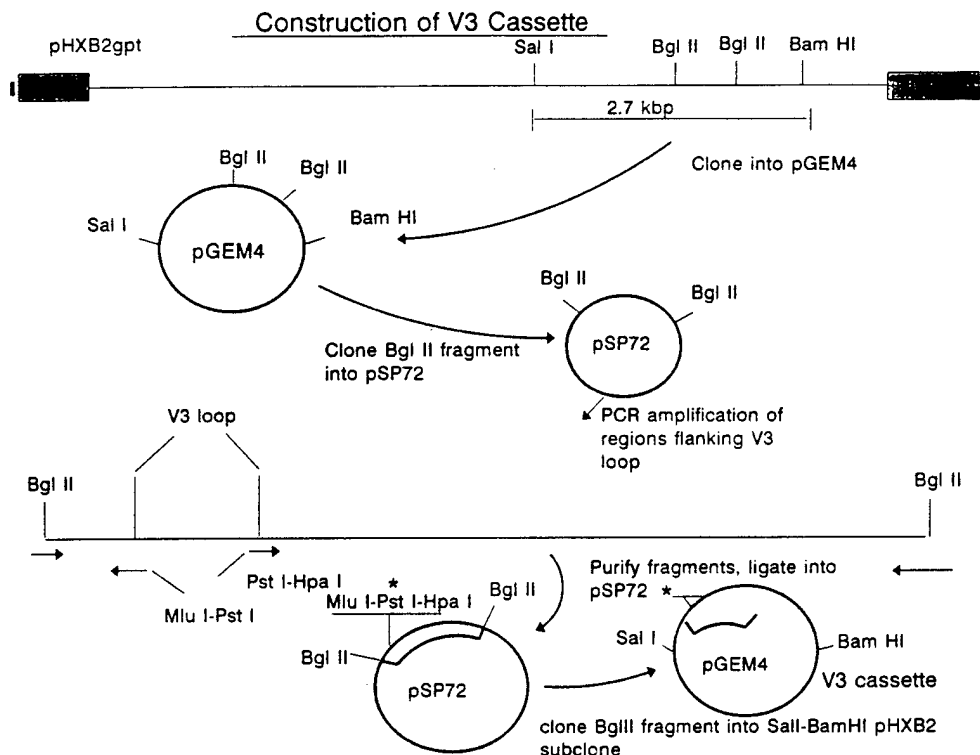


FIG. 4. Derivation of chimeric viruses. The construction of the cassette for insertion of foreign V3 loops into the infectious clone pHXB2gpt, described under "Experimental Procedures," is shown above. The Sal I-Bam HI 2.7-kbp fragment, containing most of the env gene, was subcloned. A Bgl II fragment containing the V3 region was subcloned, altered by PCR to replace the V3 loop coding region with cloning sites, and put back into the Sal I-Bam HI fragment. The resultant Sal I-Bam HI construct was used as the cassette and inserted back into pHXB2gpt after introduction of the LW87 V3 loop.

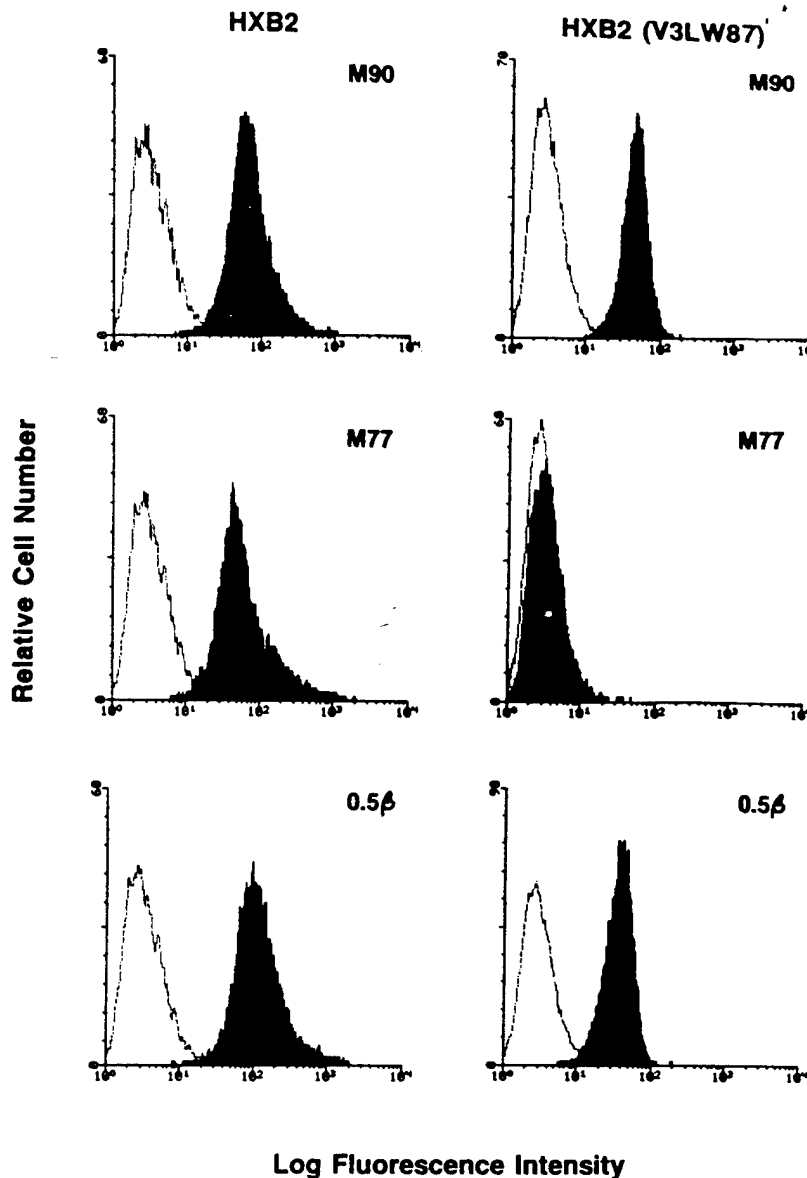


FIG. 5. Fluorocytometric analysis of CEM cells infected with HXB2 or HXB2 (V3LW87). All the monoclonal antibodies used are directed toward the gp120 external env glycoprotein; M90 has a group-specific reactivity, whereas both M77 and 0.5 β are type-specific for the HIV-1_{IIIB} isolate. The empty profiles represent the reactivity of the cells with an irrelevant isotype-matched monoclonal antibody.

(h), beta (β), coil (c), or turn (t). Second, appropriate conformation domains in the (Φ , Ψ)-space were then assigned to each of the 36 amino acids corresponding to their secondary structural states. Third, the local energy minima of the V3 loop were sampled under the constraint of the S-S bridge between the 2 invariant cysteines in the V3 loop (Fig. 1). Sampling was done by combining Monte Carlo-simulated annealing (Kirkpatrick *et al.*, 1983) and energy minimization (Fletcher, 1984). The force field of Sippl *et al.* (1984) was used for Monte Carlo and energy minimization calculations.

RESULTS

Failure of an Anti-HIV-1_{IIIB} gp120 Monoclonal Antibody to Neutralize Infection by LW87—The monoclonal antibodies chosen as probes are designated M77 and 0.5 β . They react with HIV-1_{IIIB} gp120 in both its glycosylated and unglycosylated form and were shown to inhibit syncytia formation and infection by cell-free HIV-1_{IIIB} virions in a type-specific manner (Pal *et al.*, 1992; Matsushita *et al.*, 1988). By examining a number of gp120-derived peptides, the epitope recognized by M77 was found to map between amino acids 12 and 25 (Pal *et al.*, 1992) of the sequence for the V3 loop of the HXB2 molecular clone of HIV-1_{IIIB} as depicted in Fig. 1. 0.5 β was reported previously to bind between amino acids 16–29 of the same sequence (Skinner *et al.*, 1988). Thus, the epitopes for the two antibodies are partially overlapping.

We compared the ability of these antibodies to neutralize infection by HIV-1_{IIIB}, HXB2, and LW12.3 (Lori *et al.*, 1992), an infectious molecular clone derived from a biological clone (no. 12) of the virus isolated in 1987 from a laboratory worker accidentally infected with HIV-1_{IIIB} in 1985 (Weiss *et al.*, 1988) and hereinafter designated LW87. The only amino acid difference within the M77 and 0.5 β binding sites between LW87 and the HXB2 resides at position 21 in the sequence of the V3 loop, where an alanine has been substituted for a threonine (Fig. 1). This substitution was initially detected 1 year after infection and persisted in all subsequent virus isolations.² As expected, M77 completely blocked the formation of syncytia and neutralized the infectivity of both the IIIB and HXB2 viruses (Fig. 2). However, M77 did not inhibit the formation of syncytia between SupT1 cells infected with LW87 and uninfected SupT1 (Fig. 2A). Moreover, M77 failed to neutralize cell-free infection of CEM cells by the same virus (Fig. 2B). In contrast, 0.5 β inhibited infection by LW87 in both experiments (data not shown).

Failure of M77 to Recognize Native gp120 from LW87—Since lack of neutralization is not necessarily associated with lack of

² M. S. Reitz, Jr., L. Hall, M. Robert-Guroff, J. Lautenberger, B. M. Hahn, G. M. Shaw, L. I. Kong, S. H. Weiss, D. Waters, R. C. Gallo, and W. Blattner, submitted for publication.

immunological recognition, we performed radioimmunoprecipitation assays to determine if M77 was able to react with LW87 gp120 in its native form. SupT1 cells infected with HIV-1_{IIIIB} and LW87 were metabolically labeled with [³⁵S]cysteine overnight. Labeled proteins from cellular and viral lysates were then immunoprecipitated with M77, 0.5 β , and an HIV-1 antibody-positive human serum as a positive control.

M77 efficiently precipitated gp120 from both cellular and viral extracts of HIV-1_{IIIIB}-infected cells, but failed to do so with both cellular and viral extracts of LW87 infected cells (Fig. 3, A and B). In contrast, 0.5 β immunoprecipitated gp120 from extracts of both types of infected cells (Fig. 3, A and B). Similar results were also obtained by live cell immunofluorescence labeling followed by fluorocytometric analysis (data not shown). Thus, lack of neutralization of LW87 by M77 is explained by the inability of the antibody to react with native gp120 from this virus. These results also suggest that antibodies 0.5 β and M77, despite partially overlapping epitopes in the V3 loop, differ in their binding specificity since 0.5 β reacted with native gp120 from the LW virus isolated in 1987. However, a decrease in the efficiency of immunoprecipitation of gp120 from LW87 virus was noticeable for 0.5 β , suggesting a lowered binding affinity for its epitope.

To substantiate and extend the observation obtained with the molecular clone LW87, we repeated the previously described experiments with a biological clone, designated LW87 no. 17, obtained from the same 1987 LW isolate grown in H9 cells. Again in contrast to 0.5 β , M77 failed to neutralize infection or to recognize gp120 from LW87 no. 17-infected cells in both radioimmunoprecipitation and live cell immunofluorescence labeling followed by fluorocytometric analysis (data not shown), demonstrating that an independent biological clone had lost the same neutralizing epitope.

To determine whether M77 recognized gp120 from an earlier LW virus isolated shortly after infection in 1985 (LW85), which has the same V3 loop as the molecular clone HXB2, we tested a chimeric clone with the complete *env* of a clone of LW85 inserted in HXB2 (HXB2(*env* LW85-2), a kind gift by Drs. Beatrice Hahn and George Shaw, University of Alabama, Birmingham, AL). HeLa-tat cells were transfected with HXB2(*env* LW85-2) and labeled with [³⁵S]cysteine. Both M77 and 0.5 β efficiently immunoprecipitated HXB2(*env* LW85-2) gp120 from labeled cellular and viral extracts (Fig 3C). Moreover, both monoclonal antibodies were able to completely neutralize the infectivity of this chimeric virus (data not shown).

The Ala $\rightarrow$ Thr Change Is Crucial for Neutralization Resistance and Recognition of Native gp120 by M77—In addition to the Ala $\rightarrow$ Thr replacement in the hypervariable loop, the LW virus isolated in 1987 is divergent from HXB2 in more than 10 positions throughout the gp120 sequence. To determine whether the Ala $\rightarrow$ Thr substitution was critical for the immunological reactivity and neutralizing capability of M77, we inserted the loop sequence of LW87 into HXB2, as described in detail under "Experimental Procedures" and shown in Fig. 4. The resulting infectious chimera was designated HXB2(V3LW87). HeLa-tat cells were transfected with HXB2(V3LW87) or with HXB2 as wild type control and cocultivated with uninfected CEM cells 36 hours after transfection. Approximately 3 days later, when formation of syncytia was readily visible, CEM cells were collected and subjected to live cell immunofluorescence labeling followed by fluorocytometric analysis. The 0.5 β antibody stained the surface of both HXB2(V3LW87) and HXB2-infected cells as did the positive control M90, another monoclonal antibody which reacts with an exposed conserved epitope in gp120 unrelated to the V3 loop (Fig. 5). In contrast, the M77 antibody positively reacted with wild type HXB2-infected cells, but failed to stain the surface of

HXB2(V3LW87) infected cells. These results clearly indicated that the Thr for Ala change was sufficient to determine the loss of the M77 epitope, ruling out that other amino acid changes between HXB2 and LW87 outside the V3 loop were necessary for loss of recognition of the epitope by M77.

Recognition of Linear V3 Loop Peptides by M77—As mentioned above, the only amino acid difference between the LW87 and HXB2 V3 loops consists of the substitution of a Thr for an Ala at position 21. When we assayed the ability of M77 and 0.5 β to bind in enzyme-linked immunosorbent assay to V3 linear peptides containing either Ala or Thr at that position, both antibodies bound to the two peptides (Fig. 6). These results indicated that M77 was still able to bind to the linear form of the epitope and suggested that local change in conformation of the epitope in the functionally folded protein resulted in the loss of neutralization.

Molecular Modeling and Experimental Validation of Theoretical Predictions—Molecular modeling studies using a simulated annealing approach were performed in order to understand the effect of the Ala $\rightarrow$ Thr substitution on M77 antibody binding. The folded forms of the V3 loop for the Ala and Thr analogs, as well as the extended forms, were computed using the methodology outlined under "Experimental Procedures." Molecular modeling studies indicated that the Ala $\rightarrow$ Thr substitution can disrupt the folded epitope at the tip of the V3 loop (Figs. 7 and 8). Fig. 7, A and B, shows ribbon and skeleton models of the V3 loop for the Ala analog. In this folded motif residues 16–28 of the V3 loop, which contain the crucial sequence for M77 binding, form a compact protruding surface in which the Ala at position 21 resides in the interior of the contact surface and makes close contact with atoms of the neighboring residues (Fig. 7B). Thus, the Ala $\rightarrow$ Thr substitution should result in an enlargement of the interior to accommodate the bulkier side chain. This, in turn, should enlarge the contact surface of the 16–28 motif with a resultant loss of M77 binding.

The relative stability of the folded form of the V3 loop was

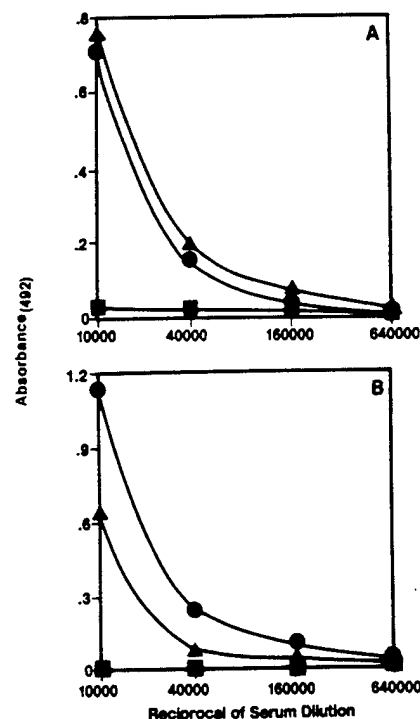


FIG. 6. Antibody binding to V3 loop peptides. Peptides from BH10 (▲), LW 1987 (●) V3 loops and an irrelevant peptide (■) were coated on plates and reacted with M77 (A) and 0.5 β (B). Sequences of the peptides are given under "Experimental Procedures."

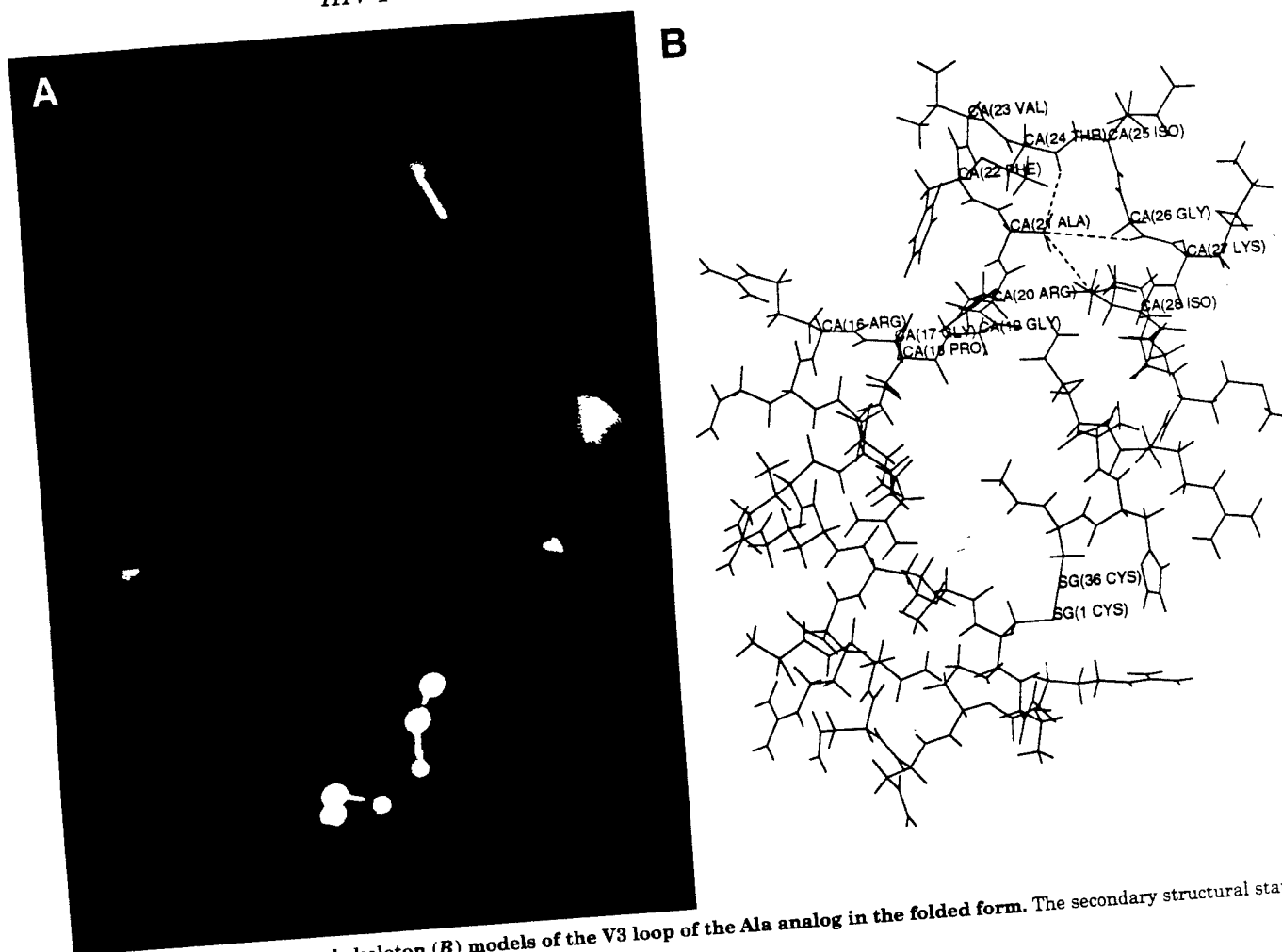


FIG. 7. The ribbon (A) and skeleton (B) models of the V3 loop of the Ala analog in the folded form. The secondary structural states of various amino acids in the sequence are as follows.

Amino acid sequence: CTRPKNNTRKRIRIQRGPGRAVFTIGKIGNMRQAHG
 Secondary structure: ccttttttctchhhhhctttttccbbccchhhhhhh

This conformation is energetically more stable than the corresponding extended conformation. The Ala at position 21 is located in the interior of the protruding surface at the center of the V3 loop. The residues 16–28 make up this protruding surface. In this folded conformation Ala²¹ shows proximity with the backbone and side chain atoms of residues Thr²⁴, Gly²⁶, and Ile²⁸. The short distances are: CB(Ala²¹)—O(Thr²⁴) = 3.7 Å, CB(Ala²¹)—O(Gly²⁶) = 4.1 Å, CB(Ala²¹)—O(Arg²⁰) = 3 Å and CB(Ala²¹)—CG2(Ile²⁸) = 3.7 Å. The Ala → Thr substitution would result in unfavorable van der Waals repulsions. These repulsive interactions would have to be relieved by enlarging the interior. This, in turn, would increase the surface area of residues 16–28 that define the M77 antibody contact domain. Secondary structural elements are color coded as follows: green, β ; red, helix; blue, coil and turn. The CB and SG atoms of C1 and C36 are shown in white and magenta, respectively.

estimated by comparing its conformational energy with that of the extended V3 loop structure where GPGR forms a type II turn (LaRosa *et al.*, 1990; Gupta and Myers, 1990), and the remaining amino acid residues adopt extended conformations, as shown in Fig. 8, A and B. The V3 loop structure of Fig. 8 is as stereochemically possible for both the Ala and the Thr analogs, but in this structure residues 16–28 (presumably forming the contact domain) extend over 21 Å. This should be too large to fit the M77 antibody binding pocket, judging from the geometry of the binding pocket of the reported crystal structures of various antibody-antigen complexes (Davies *et al.*, 1990). For the Ala analog, but not for the Thr analog the folded form shown in Fig. 7 is more stable than the extended form. Therefore, the V3 loop Ala analog does not require any loss of stabilization energy in order to bind the antibody in the folded form.

If the conformation of amino acids 16–28 defines the specificity of the contact domain of M77, and residue 21 forms part of a cavity determining the proper orientation of other residues in the protruding surface of the V3 loop to make specific contacts at the antibody binding pocket, the following predic-

tions can be made: 1) the Ala → Ser substitution at position 21 should not result in the loss of binding by M77, because Ala and Ser have a similar size and thus the small side chains of Ala and Ser can be accommodated in the specific cavity. 2) The Ala → Ile substitution should result in the loss of M77 binding because the longer side chains of Ile or Thr should not fit inside the cavity.

To support these theoretical predictions, we constructed infectious clones of HXB2 containing V3 loops into which Ser or Ile replaced Ala at position 21. HeLa-tat cells were transfected with HXB2 bearing Ala (Ala²¹), Ser (Ala²¹ → Ser) or Ile (Ala²¹ → Ile), or with HXB2(V3LW87) bearing Thr (Ala²¹ → Thr) at that position. Transfected cells were then cocultivated with uninfected CEM cells 36 h after transfection. Three to four days later, when formation of syncytia was readily visible, CEM cells were collected and subjected to live cell immunofluorescence were collected and subjected to live cell immunofluorescence labeling followed by fluorocytometric analysis. The 0.5 β antibody stained the surface of CEM cells infected by all these viruses (Fig. 9). In contrast, the M77 antibody positively reacted with cells infected by wild type HXB2 (Ala²¹) and HXB2 containing Ser at position 21 (Ala²¹ → Ser) in the V3 loop, but

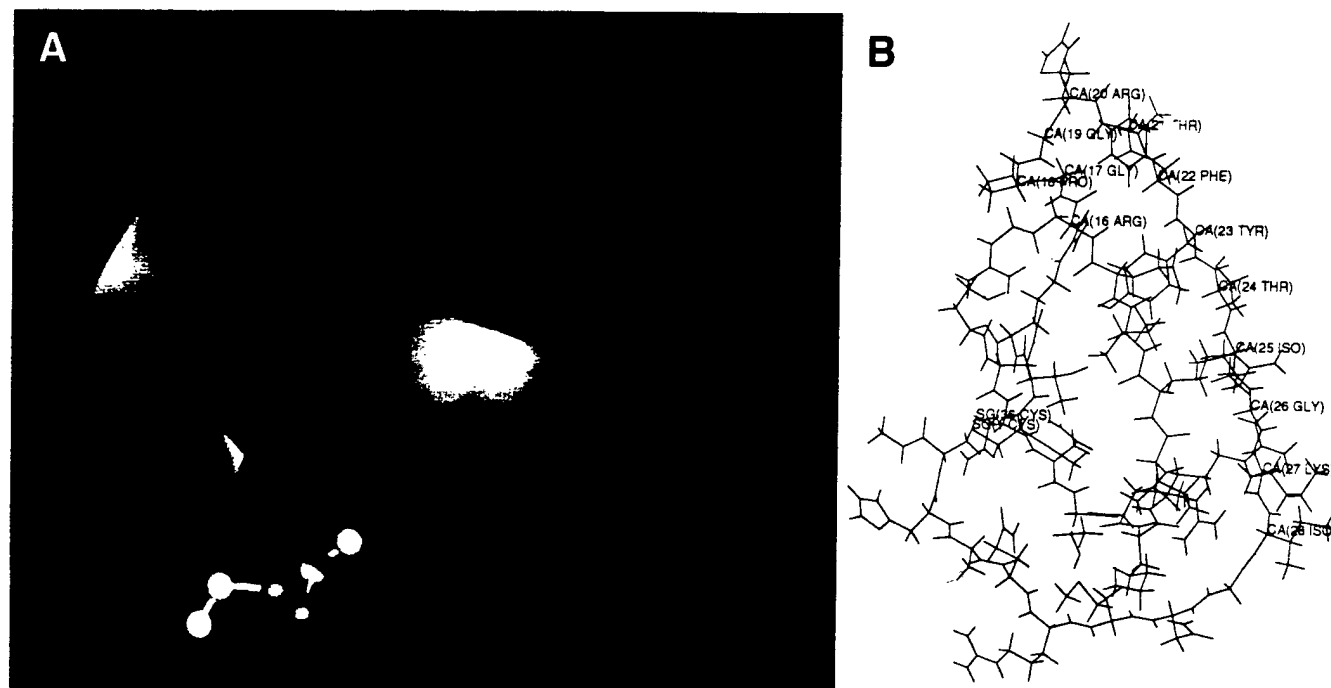


FIG. 8. The ribbon (A) and skeleton (B) models of the V3 loop of the Thr analog in the extended conformation. This conformation can also be adopted by the Ala analog. This conformation is taken as the reference (unfolded state). The relative stabilities of the folded states of the Ala and Thr analogs are measured from this reference state. In the extended conformation, residues 16–28 occupy a larger surface area than in the folded one.

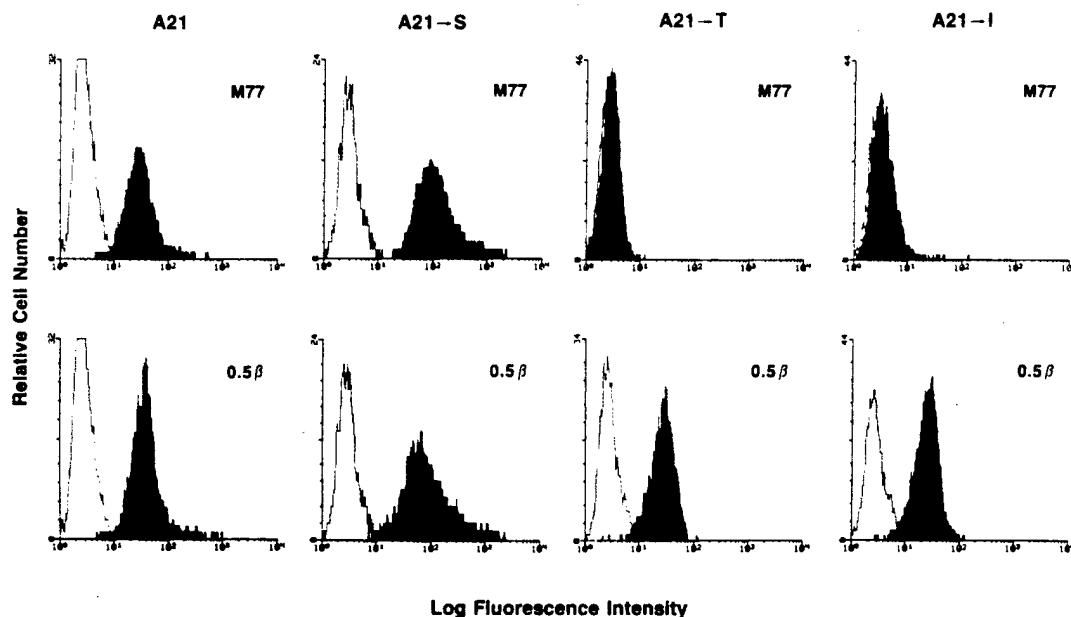


FIG. 9. Fluorocytometric analysis of CEM cells infected with HXB2 mutants. These are as follows: HXB2 bearing Ala (Ala²¹), Ser (Ala²¹ → Ser), Thr (Ala²¹ → Thr), or Ile (Ala²¹ → Ile) at position 21 in the V3 loop. The empty profiles represent the reactivity of the cells with an irrelevant isotype-matched monoclonal antibody.

failed to react with cells infected by HXB2 containing Thr (Ala²¹ → Thr) or Ile (Ala²¹ → Ile) at that position (Fig. 9). Thus, the M77 reactivity perfectly correlated with the modeling predictions.

DISCUSSION

The availability of sequential viral isolates from a laboratory worker accidentally infected with the prototype HIV-1 strain IIB gave us the unique opportunity of studying the influence of naturally occurring mutations on V3 loop-dependent neutralization. Different molecular clones of LW obtained shortly after infection were genetically almost indistinguishable among

themselves and from HIV-1_{IIB}-derived molecular clones.² Indeed, within the V3 region, they completely retained the sequence of the HXB2 clone from HIV-1_{IIB}. An Ala to Thr substitution within the GPGRF motif at the tip of the loop was detected 1 year after infection and persisted in all subsequent isolations.²

We have observed that the HIV-1_{IIB}-neutralizing monoclonal antibody M77 was unable to neutralize infection by a biologically active molecular clone of LW isolated in 1987 and bearing the Thr for Ala substitution at position 21 within the V3 loop. Lack of neutralization does not always correlate with lack of immunological recognition, since it has been shown that

monoclonal antibodies can bind the V3 loop without neutralization (Nara *et al.*, 1992). In this case, however, lack of neutralization by M77 was a direct consequence of the inability of the antibody to bind native gp120, as indicated by immunoprecipitation and fluorocytometric analysis.

The antibody M77 did not react with the chimeric virus obtained after insertion of LW87 V3 loop in HXB2, clearly indicating that the Ala → Thr substitution is sufficient for the abrogation of antibody binding to native gp120. Thus far, only one neutralization resistant escape mutant selected *in vitro* with monoclonal antibodies to the V3 loop showed an amino acid change within the binding site of the selecting antibody (McKeating *et al.*, 1989). This change involved the substitution of a G for an R at position 13 in the sequence for the V3 loop and clearly affected the binding of the antibody to the linear epitope. In contrast, the majority of the neutralization-resistant mutants which were selected *in vitro* with monoclonal antibodies to the V3 loop did not show any amino acid sequence variation within the loop itself, suggesting that critical changes affecting virus neutralization reside outside the V3 domain (McKeating *et al.*, 1989). This, in fact, has been proven in the *in vitro* immune escape system with human antisera (Robert-Guroff *et al.*, 1986), in which a point mutation in the transmembrane portion of the *env* gene resulted in neutralization resistance (Reitz *et al.*, 1988). Subsequent studies showed that this mutation affected neutralization by conformational changes induced by an Ala → Thr change at position 582 in gp41 (Wilson *et al.*, 1990).

These *in vitro* findings were also supported by results generated *in vivo* in chimpanzees (Nara *et al.*, 1990). Experimental infection of chimpanzees with HIV-1_{III}B resulted in the rapid emergence of mutant viruses resistant to neutralization by V3-specific neutralizing antibodies. Sequencing of the envelope gene from these viruses revealed the complete identity of the V3 amino acid sequence between the neutralization resistant mutant viruses and the neutralization-sensitive molecular clones contained in the original inoculum. However, several amino acid substitutions occurred throughout gp120, indicating again that the changes affecting virus neutralization resided outside the V3 loop. In contrast, our data show that the mutation of a single residue within the V3 loop occurring *in vivo* in a human results in the loss of the binding site for a V3-specific neutralizing antibody.

The antibody M77 was able to bind to the linear form of its epitope, implying that a local change in conformation was actually responsible for the loss of the epitope itself in the functionally folded protein. It is conceivable that M77 reacts with both isolated V3 peptides because these are flexible, do not have structural constraints, and thus have more freedom to adapt to the antibody complementarity region. However, the M77 antibody can only bind the (S-S)-bridged V3 loop Ala analog but not the V3 loop Thr analog. This means that neither Ala nor Thr at the same position of the V3 loop make direct contact with the antibody. If this were the case, the M77 antibody would have discriminated between the Ala and the Thr analogs in the linear peptide epitopes. Therefore, it appears that the (S-S)-bridged V3 loop imposes certain stereochemical constraints such that only the Ala analog, but not the Thr analog, can present the antibody-binding domain of the V3 loop in an effective manner to the M77 antibody.

Our molecular modeling studies, experimentally supported by different amino acid replacements, suggested that the Thr for Ala substitution can disrupt the folded antigenic epitope at the tip of the V3 loop (amino acids 16–28). In the folded form the amino acids forming the epitope adopt a protruding surface, in which the Ala²¹ residue resides in the interior of the contact surface, while the Ala → Thr substitution requires enlargement

of the contact surface resulting in the loss of antibody binding.

This report demonstrates that the V3 loop can adopt a specific structure in the presence of the antibody. Both from secondary structure prediction algorithms and two-dimensional NMR studies on free V3 loops, it has been suggested that the GPGR sequence at the center of the immunogenic tip can adopt a type II β -turn (LaRosa *et al.*, 1990; Chandrasekhar *et al.*, 1991). However, no evidence has been presented to date that the GPGR sequence together with the flanking amino acids can form a well defined combining site for antibodies. Amino acids on the surface of this combining site provide direct contact with the antibody binding pocket, whereas the amino acids in the cavity determine the size and shape of the cavity itself while maintaining a specific geometry of the antibody contact domain of the V3 loop.

Acknowledgments—The following reagent was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: 0.5 β monoclonal antibody from Dr. Shuzo Matsushita. We thank Richard Crowley for excellent technical assistance and Sharon Orndorff and Joan Jackson for editorial assistance.

REFERENCES

- Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H., and Essex, M. (1985) *Science* **228**, 1091–1094.
- Chandrasekhar, K., Profy, A. T., and Dyson, H. J. (1991) *Biochemistry* **30**, 9187–9194.
- Chen, C., and Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 2745–2752.
- Davies, D., Padlan, E. A., and Sheriff, S. (1990) *Annu. Rev. Biochem.* **59**, 439–473.
- Deleage, G., and Roux, B. (1989) in *Prediction of Protein Structure and Principles of Protein Conformation* (Fasman, G. D., ed) pp. 587–597, Plenum Press, New York.
- Fletcher, R. (1984) in *Practical Methods of Optimization*, First Ed., John Wiley & Sons, New York.
- Gupta, G., and Myers, G. (1990) in *Proceedings of the Cinquieme Colloque des Cent Gardes* (Girard, M., and Valette, L., eds) pp. 99–104, Pasteur Vaccines, Paris.
- Javaherian, K., Langlois, A. J., McDaniel, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Putney, S. D., and Matthews, T. J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6768–6772.
- Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. (1983) *Science* **220**, 671–680.
- LaRosa, G. J., Davide, J. P., Weinhold, J. A., Waterbury, A. T., Profy, J. A., Lewis, A. J., Langlois, G. R., Dreesman, R., Boswell, N., Shaddock, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E. A., and Putney, S. D. (1990) *Science* **249**, 932–935.
- Laemmli, U. K. (1970) *Nature* **227**, 680–685.
- Lori, F., Hall, L., Lusso, P., Popovic, M., Markham, P., Franchini, G., and Reitz, M. S., Jr. (1992) *J. Virol.* **66**, 5553–5560.
- Matsushita, S., Robert-Guroff, M., Rusche, J., Koito, A., Hattori, T., Hoshino, H., Javaherian, K., Takatsuki, K., and Putney, S. (1988) *J. Virol.* **62**, 2107–2114.
- McKeating, J. A., Gow, J., Goudsmit, J., Pearl, L. H., Mulder, C., and Weiss, R. A. (1989) *AIDS (Phila.)* **3**, 777–784.
- Nara, P. L., Smit, L., Dunlop, N., Hatch, W., Merges, M., Waters, D., Kelliher, J., Gallo, R. C., Fischinger, P. J., and Goudsmit, J. (1990) *J. Virol.* **64**, 3779–3791.
- Nara, P. L., Minassian, A., Garrity, R., and Goudsmit, J. (1992) in *Vaccines 92* (Brown, R., Binsberg, H., and Lerner, R., eds) pp. 175–181, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Pal, R., Veronese, F. D., Nair, B. C., Rahman, R., Hoke, G., Mumbauer, S. W., and Sarngadharan, M. G. (1992) *Intervirology* **86**, 86–93.
- Palker, T. J., Clark, M. E., Langlois, A. J., Matthews, T. J., Weinhold, K. J., Randall, R. R., Bolognesi, D. P., and Haynes, B. F. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1932–1936.
- Reitz, M. S., Jr., Wilson, C., Naugle, C., Gallo, R. C., and Robert-Guroff, M. (1988) *Cell* **54**, 57–63.
- Robert-Guroff, M. (1990) *Int. Rev. Immunol.* **7**, 15–30.
- Robert-Guroff, M., Reitz, M. S., Jr., Robey, W. G., and Gallo, R. C. (1986) *J. Immunol.* **137**, 3306–3309.
- Robert-Guroff, M., Aldrich, K., Muldoon, R., Stern, T. L., Bansal, G. P., Matthews, T. J., Markham, P. D., Gallo, R. C., and Franchini, G. (1992) *J. Virol.* **66**, 3602–3608.
- Rusche, J. R., Javaherian, K., McDaniel, C., Petro, J., Lynn, D. L., Grimala, R., Langlois, A., Gallo, R. C., Arthur, L. O., Fischinger, P. J., Bolognesi, D. P., Putney, S. D., and Matthews, T. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3198–3202.
- Sippl, M. J., Némethy, G., and Scheraga, H. A. (1984) *J. Phys. Chem.* **88**, 6231.
- Skinner, M. A., Ting, R., Langlois, A. J., Weinhold, K. J., Lyster, H. K., Javaherian, K., and Matthews, T. J. (1988) *AIDS Res. Hum. Retroviruses* **4**, 187–197.
- Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G. (1985) *Science* **229**, 1402–1405.
- Weiss, S. H., Goedert, J. J., Gartner, S., Popovic, M., Waters, D., Markham, P., di Marzo Veronese, F., Gail, M. H., Barkley, W. E., Gibbons, J., Gill, F. A., Leuther, M., Shaw, G. M., Gallo, R. C., and Blattner, W. A. (1988) *Science* **239**, 68–71.
- Wilson, C., Reitz, M. S., Jr., Aldrich, K., Klasse, P. J., Blomberg, J., Gallo, R. C., and Robert-Guroff, M. (1990) *J. Virol.* **64**, 3240–3248.

Presentation of HIV V3 Loop Epitopes for Enhanced Antigenicity, Immunogenicity and Diagnostic Potential

Running Head: Immunological Properties of MUC1/V3 Loops

J. Darrell Fontenot^{1,2}, Tom. C. VanCott³, Bharat S. Parekh⁴, Chou-Pong Pau⁴,
J. Richard George⁴, Debra L. Birx³, Susan Zolla-Pazner⁵, M. K. Gorny⁵
and Joe M. Gatewood^{2*}

¹Theoretical Biology and Biophysics and ²Life Sciences Division
Los Alamos National Laboratory
Los Alamos, New Mexico 87545, USA

³Department of Retroviral Research
Walter Reed Army Institute of Research
13 Taft Court, Suite 200
Rockville, Maryland 20850, USA

⁴Division of HIV/AIDS
National Center for Infectious Diseases
Centers for Disease Control and Prevention
1600 Clifton Road
Atlanta, Georgia 30330, USA

⁵Veterans Administration Medical Center
423 East 23rd Street
New York, NY 10010

* Corresponding Author

This work was supported by NIH grant R01 AI32891-01A2, and Los Alamos National Laboratory grant XL 77.

Abstract

Objectives: The immunological properties of a panel of human mucin MUC1/HIV V3 loop chimeras is evaluated.

Design: The immunodominant epitope of MUC1 (APDTR) was found to be structurally isomorphous with the tip of the principle neutralizing determinant (PND) of HIV-1 (MN) (GPGRA). A panel of 120 residue, 6 tandem repeat and 60 residue, 3 tandem repeat antigens chimeric antigens were constructed in which the repeating MUC1 epitope is replaced by the principle neutralizing determinant of HIV-1. Each 20 residue tandem repeat contains one PND epitope. The PND of HIV-1 is presented in the native β -turn conformation at the crest of each repeating knob structure of the mucin-like protein.

Methods: The antigenicity of the chimeric antigens are compared using ELISA and HIV infected patient sera. Structural effects of antibody -antigen interactions are determined using Surface Plasmon Resonance (SPR), with human monoclonal antibodies, chimeric antigens and the cyclic and linear V3 loops. Immunogenicity of 3 versus 6 tandem repeats is measured in mice.

Results: Nine residues of the HIV PND substituted into the mucin backbone were equivalent to the 36 residue cyclic V3 loop in ELISA. The 120 residue antigens induced high titer, IgM and IgG, HIV specific antibodies in mice.

Conclusions: MUC1/V3 chimeras efficiently detect HIV specific antibodies in patient sera. Multivalent presentation of the PND is advantageous for higher affinity antibody-antigen interactions and for inducing HIV specific IgM and IgG antibodies.

KEY WORDS: Human Mucin MUC1, HIV-1, V3 Loop, tandem repeats, Immunogenicity, Antigens, Surface Plasmon Resonance

Introduction

The external surface unit glycoprotein (gp120) of the human immunodeficiency virus type 1 contains multiple disulfide bonded loops [1]. The principal neutralizing determinant (PND) is located inside the third variable loop (V3 loop) of gp120 [2-4]. The common structural element of the V3 loop PND is the type-II reverse turn (GPGR) near the midpoint of the 35-residue long [5-10]. PND specific neutralizing antibodies generally recognize the GPGR turn and two or three flanking residues, either toward the C- or N-terminus [11]. Recent studies show that this relatively conserved structural feature of the HIV-1, PNDs, is further characterized by the formation of a solvent accessible protruding motif (designated knob), with the type II turn at the crest [7, 8, 12]. However, during an immune response, the structural conservation of the PND may be masked by the variable flanking sequences in the V3 loop of gp120.

Our work with the structure of the human mucin MUC1 tandem repeat domain supported the design of protein constructs in which HIV-1, PNDs could be presented by this multivalent antigen in their "correct" conformations [13-16]. Human mucin MUC1 is a cell surface glycoprotein with a large tandem repeat (TR) domain [17, 18]. The number of tandem repeats is variable, but individuals may contain (20-100) perfect copies of a twenty residue proline, serine, threonine, glycine and alanine rich tandem repeat (the tandem repeat sequence is GVT SAPDTRPAPGSTAPPAH)[19, 20]. In breast, pancreatic, and ovarian cancers the MUC1 tandem repeat sequence APDTR, appears to be immunodominant for antibody binding specificity, similar to the GPGR of HIV-1, (MN) [21-23]. The high resolution structure of MUC1 revealed repeating (every 20 residues) solvent exposed knobs, with an immunodominant APDTR type II turn at the crest [15]. We have also shown that isomorphous replacement of the mucin knobs, with the HIV-1,

PND knobs, results in mucin-like HIV-1 antigens that retain the structural purity of the HIV-1 epitope, the global extended structure of mucin, and are recognized by polyclonal sera from HIV-1 infected individuals [16].

In this paper, we describe how the immunologic properties of six, seven, and nine residue sequences, which include the β -turn (GPGR) of the HIV-1 V3 (MN), are enhanced by presentation within the mucin backbone. In addition, we show that large multivalent antigens ($N = 3, 6$) are advantageous over monovalent linear peptides for detecting HIV-1, PND antibodies in the serum of HIV infected individuals. By studying the kinetics of HIV specific monoclonal antibodies, interacting with chimeric mucin-V3 loop (mucV3) multivalent antigens, we determine how the larger 120 residue multivalent antigens containing six antigenic structural knobs increase the affinity of the antigen-antibody interaction over the 60 residue antigens with three structural knobs.

Materials and Methods

Antigens

The principles of design and NMR structural information of the mucV3 chimeric polypeptides were described earlier [16]. These include both the 60 and 120 residue synthetic peptides of the mucV3 chimeras (Table I). The chimeras mucV3-6, mucV3-7, and mucV3-9 contain either six (IGPGRA), seven (HIGPGRA), or nine (IHIGPGRAF) residues, respectively, of the HIV-1 V3 (MN) loop, per twenty residue tandem repeat. In each antigen, as shown in Table I, mucin residues were removed and replaced by HIV-1 residues, and the length of the repetitive element was maintained at 20 residues. The intention was to maintain the overall mucin structure, but to replace the immunodominant knobs of mucin with the immunodominant knobs of the V3 loop. The sequence of one tandem repeat for human mucin MUC1, mucV3-6 containing 6 PND residues per 20 residue tandem repeat, mucV3-7 containing 7 PND residues per 20 residue tandem repeat, and mucV3-9 containing 9 PND residues per 20 residue tandem repeat are shown in Table I. Notice that there are two sizes of each mucV3 chimeric peptide, one with three tandem repeats or 60 residues and one with six tandem repeats or 120 residues (Table I). Additional V3 (MN) antigens evaluated include the full length cyclic (oxidized) and linear (reduced) forms of the V3 loop of HIV-1 (MN), and a 15 residue linear peptide, designated 1143D, (DKRIHIGPGRFYTT). The N-terminal aspartic acid was added for enhanced peptide binding to microwells.

Peptide Synthesis, Purification and Mass Characterization

All peptides, except the 15 residue 1143D, were peptide amides and were synthesized by a manual solid-phase strategy using 9-fluorenylmethyloxycarbonyl protected amino acids as described in detail elsewhere [14]. Peptide molecular weights were characterized by electrospray ionization mass spectroscopy at the

Protein and Carbohydrate Structure Facility, University of Michigan, Medical School under the direction of Dr. Phil Andrews. In each case the difference between expected and observed peptide molecular weights was within the experimental error of the spectrometer. For example, the expected molecular weights of the mucV3-6, mucV3-7 and mucV3-9, 120 residue peptides are 10,922, 10994 and 11,380 Daltons, and the observed molecular weights are 10,916, 10,988 and 11,370 Daltons.

Patient Sera, Monoclonal Antibodies and ELISA Methods

A panel of HIV-1 sera from patients living in Honduras was collected, characterized and maintained at the Centers for Disease Control by the Division of HIV/AIDS in Atlanta, GA. These patients are infected with HIV-1 genotypes characteristic of clade B viruses. The V3 loop sequence of the predominant virus population for 22 of these individuals has been determined [24]. The HIV-1 (MN) isolate, on which the mucin V3 chimeric antigens are based, also belongs to genotype B, as do the majority of HIV-1 isolates found in North America [25]. In addition, we used serum samples from eleven individuals known to be infected with diverse non-clade B HIV-1 subtypes, these included 2 from Rwanda, 4 from Brazil, 3 from Thailand, and 2 from Uganda [24]. The predominant V3 loop sequence from all of these individuals was determined to be non clade B, both by sequence and serology [24]. V3 loop specific monoclonal antibodies 447, 412, 453, 386, 268, 782, 257, 391, 838, 419, 181, 908, 537 were derived from AIDS patients at the New York Veterans Affairs Medical Center, New York, NY as described previously [11, 26]. Additional V3 monoclonal antibodies used in this study 9205 and 9284 [27] were obtained from Dupont (Boston, MA) and mAB 50.1 [28] was obtained from Repligen (Cambridge, MA). The ELISA method using patient sera was described in detail elsewhere [29].

BIAcore Measurement of Antibody Binding and Affinity

The mucV3-9, 60 and 120 residue peptide antigens, were selected for affinity measurements with V3 mAbs using surface plasmon resonance (SPR), (BIAcore™, Pharmacia Biosensor, Piscataway, NJ) because these antigens showed maximum binding to V3 mAb. Monoclonal antibodies 412, 447, and 453 were chosen for this experiment because their dissociation rates are faster than 1×10^{-5} as required for accurate measurements. The SPR technology is described in detail elsewhere [30-32]. Both mucV3-9, 60 and 120 residue peptides were covalently attached to a dextran matrix by EDC/NHS chemical activation and coupling of free amines on the peptide N-termini to activated carboxyl groups on the resin. The optimal immobilization buffer was 10mM 2-(N-morpholino)ethanesulfonic acid (MES) at pH 6.0. The eluent buffer is 10mM HEPES, 150mM NaCl, 3.4mM EDTA, .05% BIAcore surfactant P20, pH 7.4. A flowrate of 5 ul/min and 2 ul/min was used for the immobilizations and the binding experiments respectively. The unreacted EDC-esters were blocked with an injection of 1M ethanolamine. The amount of protein was determined by the increase in baseline level of plasmon resonance.

Kinetic analysis was performed as described [30-33]. Serial 2-fold dilutions of the antibodies in HBS buffer were injected across the mucV3 matrices at 5 ul/min. Following the injection of the antibody the dissociation was monitored for 15 minutes by flowing buffer over the biosensor matrix. The apparent dissociation and association rates were determined using the equation $dR/dt = -(k_1C + k_{-1})R + k_1CR_{max}$. k_1 is the apparent association rate constant and k_{-1} is the apparent dissociation rate constant. R_{max} is the maximum binding capacity of the peptide immobilized on the dextran matrix, R is the amount of bound antibody (in Response Units RU) at time t and concentration C . A linear plot of R versus dR/dt yields a slope of $-(k_1C + k_{-1})$ and y-intercept of k_1CR_{max} . dR/dt is determined by measuring the slope at several points along the antibody-peptide association curve.

The slopes were plotted versus the antibody concentration to obtain k_1 . k_1 was obtained from the equation $\ln R_{t1}/R_{t0} = k_1(t_0 - t_1)$. R_{t0} is the response immediately following the completion of the antibody injection and R_{t1} is the response at a later time. A plot of $\ln R_{t1}/R_{t0}$ versus $(t_0 - t_1)$ yields apparent dissociation rate, k_1 .

Immunogenicity of the Antigens in Mice

Peptide immunogenicity studies were performed using 4 week old female Balb/c mice and groups of 4 mice/antigen. One set of immunizations used an initial intraperitoneal injection that consisted of 100 ug antigen /50 ul of phosphate buffer emulsified with 50 ul Freund's complete adjuvant, for a total volume of 100ul/ mouse. The initial injection was followed by 4 subsequent boosts of 100 ug of antigen emulsified in Freund's incomplete adjuvant (100ul total volume) at approximately 3 week intervals. To assess the role of antigen valency and dose on the antibody response, groups of 4 mice/antigen were immunized intraperitoneally using 2, 10, and 20 nmoles/immunization of mucV3-9 (120) or mucV3-9 (60) residue peptides. In this experiment, the initial immunization was followed by 3 boosts at approximately 4 week intervals. Blood was collected by tail bleeds 5-6 days after immunizations.

Results

The Importance of Antigen Structure

A panel of human V3 mAbs from infected patients were used to assess the importance of V3 structure on binding affinity [11, 34]. The ability of these human mAbs to distinguish between the linear reduced V3 loop and the cyclic oxidized forms of the HIV-MN V3 loop was evaluated using SPR. The uncertainty in determining the amount of bound antigen, that is a limitation in ELISA, was eliminated in the current experiment. Since the level of each peptide bound to the dextran matrix can be determined. Therefore, the binding data are scaled by this amount and errors due to differences in binding are eliminated. For example, the cyclic V3 MN peptide coupled to the dextran matrix with an Re (831), and the linear V3 MN peptide coupled to the dextran matrix with Re (1124). The mAb 447 bound to the matrix attached cyclic V3 MN with Re (3355) and to the matrix attached linear V3 MN peptide with Re (3591). The binding reactivity ratio for mAb 447 for the cyclic and linear forms of the V3 loop are $3355/831=4.0$ for the cyclic peptide and $3591/1124=3.2$ for the linear peptide. Therefore, mAb 447 preferentially bound to the cyclic form of the loop by the following ratio $4.0/3.2=1.25$, which is a 25% preference. In Table II, 15 out of 15 monoclonal antibodies showed a higher binding reactivity ratio to the cyclic V3 loop (13-27%). These data support other observations showing that antibodies derived from HIV-1 infected patients have a higher binding affinity for the cyclic form of the V3 loop [12].

Enhanced Antigenicity and Diagnostic Potential

The ELISA reactivity of a panel of HIV-1 sera was used to assess the antigenicity of the V3 peptide antigens in Table I. The antigenicity of the mucin V3 loop chimeric antigens was compared to the antigenicity of the cyclic and linear forms of the 36 residue V3 loop, the 15 residue linear HIV-MN peptide, and a 105

residue mucin MUC1 peptide. The panel of sera represent 29 patients from the Honduras infected with clade B viruses characteristic of the North American isolates. For comparison, the serological reactivity of 11 sera from patients infected with viral strains whose PND sequence is non-clade B were measured.

Serological reactivity of these HIV-1 sera with the chimeric antigens (1/100 dilution of serum) as measured by ELISA is shown in Figure 1. The data are displayed in order to emphasize the increasing antigenicity of either six (IGPGRA)(Fig. 1A), seven (HIGPGRA)(Fig. 1B), or nine (IHIGPGRAF)(Fig. 1C) residues of the HIV-V3 loop when presented in the context of a multi-valent mucin-like knob. As shown in Figure 1A, when only 6 residues of the PND are used, 65% of the clade B sera tested positive, (24% show a strong reaction and 41% have an intermediate level of serological reactivity). The number of clade B patients who show a strong (52%) to intermediate (24%) level of serological reactivity increases to about 76% when seven residues of the V3 loop are substituted into MUC1 (Figure 1B). The strong and intermediate categories were enhanced by 28% and 15% respectively, simply by adding a single residue of the V3 loop into the mucin chimera. As shown in Figure 1C, substitution of nine residues of the V3 loop into mucin, results in 100% of the clade B sera showing either strong reactivity (90%) or intermediate (10%). Increasing the number of PND residues substituted into the mucin backbone from 7 to 9 results in significant enhancement of the number of sera in the strong category at the expense of the intermediate and weak categories of serological reactivity. The number of non-clade B patients with strong to intermediate levels of cross reactivity to the MUC1/HIV-MN chimeras increased from one to three as the number of HIV residues in the chimera increased from six to nine (Figures 1A-C). Only two sera, one clade B with weak reactivity and one non-clade B with intermediate reactivity, recognized the mucin negative control, (data not shown).

ELISA reactivity of the sera described above are compared to the full HIV-MN cyclic V3 loop (Fig. 2A), the 15 residue linear peptide (Fig. 2B), and the mucV3-9-120 residue multi-valent chimera (Fig. 1C). Comparing figures 2A and 1C shows that a multi-valent mucin-V3 loop chimera with only nine residues of the V3 loop detected an equivalent level of antibody reactivity in patient serum (90% strong, 10% intermediate and 3 cross reactors) as the entire 36 residue cyclic V3 loop. The mucin chimeras with nine residues of the HIV-V3 loop (Fig. 1C), detected 50% more strong serological reactions among the clade B sera than a 15 residue HIV-MN linear peptide (Fig. 2B). Comparing figures 1A and 1B with 2B reveals that mucin chimeras with 6 and 7 residues of HIV sequence detect similar or higher amounts of HIV-PND serological reactivity than the 15 residue linear peptide.

Endpoint ELISA titers of the clade B HIV-1 serum against these antigens reveals several additional features of both the serum and the antigens involved. Figure 3A shows one example of an individual clade B serum titrated with the cyclic and linear V3 loops, the 15 residue linear peptide, and the three mucin-V3 chimeric peptides. The cyclic and linear V3 loops displayed similar antibody reactivity to HIV-1 serum, although the cyclic V3 loop consistently detected slightly higher levels of antibody reactivity in all the titrated sera (a total of 19 clade B sera). The pattern of reactivity seen in Figure 3A for the 120 residue mucV3 chimeras mucV3-6, mucV3-7, and mucV3-9 and the 15 residue linear peptide is also characteristic of the 19 titrated sera. The chimera with nine HIV residues detects the same or slightly less antibody reactivity than the full cyclic V3 loop at all dilutions. Although the mucV3-7 chimera with 7 residues detects consistently less antibody reactivity than mucV3-9, antibody could be detected at all dilutions. The mucV3-6 chimera usually detected the least amount of antibody reactivity of the chimeric proteins, but usually is very close to the 15 residue linear peptide. One trend is consistently observed, that the mucV3-7, and mucV3-9 chimeras detect antibody reactivity at 10

to 50 fold higher dilutions than the 15 residue linear peptide or, at a given dilution, much higher absorbance values are observed.

Multivalent antigens are potentially capable of detecting higher levels of serum antibodies specific for the β -turn at the crest of the V3 loop than even the full length cyclic V3 loop antigen. Some of the antibody reactivity detected by the full length V3 loop peptides must arise from epitopes other than the β -turn at the crest, which could not possibly be present in the chimeras. Figure 3B shows, that for an individual with a V3 loop sequence at the position of the β -turn (IHMGWGRAFY) with critical mutations relative to the MN sequence (I to M; P to W) that eliminate all antibody binding to the chimeric proteins, there can be significant antibody reactivity to other epitopes within the loop. It is conceivable that multivalent presentation of the β -turn at the crest of the V3 loop in the correct conformation by the chimera could offer quantitative advantages for the creation of diagnostic peptides.

Advantages of Multi-Valency

BIAcore SPR can be used to measure binding kinetics of monoclonal antibody and HIV-1 chimeric antigen interactions and to examine the impact of multi-valency in antigen-antibody binding. The mucV3-9 antigens were used in this experiment since these show the maximum reactivity in ELISA of the mucin chimeras with HIV infected patient serum. The association (R_1) and dissociation (R_{-1}) rate constants of the antibody-antigen interaction mABs 447, 412, or 453 and the 60 and 120 residue peptides (3 knobs versus 6 knobs) are summarized in Table III. The R_1 values shown in Table III for the monoclonal antibodies 447 and 412 binding to mucV3-9 (60) and mucV3-9 (120) are similar. However, the dissociation rates for the monoclonal antibodies 447 and 412 is 10.4 and 11.0 times slower for mucV3-9 (120) than for mucV3-9 (60). As a result these antibodies have a ten-fold stronger

affinity constant for the higher valency antigen. The effect is less pronounced but similar for the monoclonal antibody 453 binding to the 60 and 120 residue peptides.

Immunogenicity

Multivalent binding has the effect of changing the overall equilibrium constant for the antibody-antigen interaction. This effect should also be operating when the antigen engages an immunoglobulin molecule on the surface of a lymphocyte during the induction of an immune response. The six knob, 120 residue peptide was expected to induce significant IgM antibodies through a T-cell independent mechanism [22, 23]. Table IV summarizes the results of four 100 ug immunizations of sets of 4 Balb/C mice with the mucV3-6 (120), mucV3-7 (120), mucV3-9 (120), and mucin 105 residue proteins. MucV3-9 (120), which proved to be the most effective chimeric antigen, is also the most immunogenic in Balb/C mice. The mucV3-9, 120 residue antigen induced high levels of IgG antibodies as well as the expected IgM response. Using the ability to induce IgG and IgM antibodies as an indicator, the chimeric antigens displayed the following decreasing immunogenicity in Balb/C mice as shown in Table IV; mucV3-9 (120) > mucV3-6 (120) > mucV3-7 (120). The 105 residue mucin control protein only induced IgM antibodies, which is consistent with the *in vivo* humoral antibody responses to MUC1 reported in cancer patients [22, 23].

We reasoned that if the high levels of IgG induced by mucV3-9, 120 residue peptide were the result of the creation of a T-cell helper epitope in the chimeric sequence, then the response to the 60 residue peptide should be very similar to that seen for the 120 residue immunogen. This is because the 60 residue molecule contains the same sequence and structural information as the 120 residue molecules, the only difference is valency. The mucV3-9 (60) and 120 residue peptides were each used to immunize 3 sets of 4 Balb/c mice. The mice were

immunized with 3 different antigen doses corresponding to 2, 10, and 20 nmoles of antigen. The ELISA serological reactivity of the mouse sera to the immunizing antigens are shown in Figure 4. Mice immunized with mucV3-9 (60), at all doses, did not induce IgG or IgM antibody responses to mucV3-9 (60). In contrast, high levels of IgG antibody are induced by mucV3-9 (120) residue peptide at each dose. We did not observe detectable IgG or IgM antibody responses to mucV3-7 (60) in 100 ug immunization of Balb/c mice. It appears that the 60 residue antigens are non-immunogenic due to either insufficient valency to induce T-independent responses or lack of T-cell helper epitopes.

The specificity of the murine IgG response to the mucV3-9 (120) residue peptide, from mice immunized with 10 nmoles of mucV3-9 (120), is shown in Figure 5. The titer to the immunizing antigen is greater than 1/100,000 after only 4 months. The titer to the cyclic V3 loop titer is 1/50,000 and less than 1/10,000 for the linear V3 loop at 4 months. No substantial anti-mucin reactivity develops during this time. This data strongly supports the idea that these antigens will be useful tools for inducing T-independent HIV specific antibodies.

Discussion

We previously showed by high resolution NMR that substitution of the immunodominant knob of human mucin MUC1 with the corresponding HIV-PND results in structural preservation of the "correct" HIV epitope [16]. The goal of the present study was to examine the antigenic, diagnostic and immunogenic characteristics of these multi-valent mucV3 chimeric antigens. By substituting increasing numbers of HIV-PND residues at the equivalent positions in the repeating mucin knobs, mucin-like molecules with multiple HIV-PND epitopes in their correct conformation were created. This approach offers several theoretical advantages: First, presentation of the "native" conformation of short β -turn segments can be maintained by the constraints of the MUC1 framework. Second, any possibility for the PND epitope to be obscured by flanking hypervariable sequences within the V3 loop is eliminated. Third, the antigenic and immunogenic potential of the epitope is increased by presentation within the multi-valent MUC1. Finally, mucin presentation of foreign β -turns at the tip of the knob assures that the epitope in question is located on a surface that is accessible to antibody molecules in solution or on the surface of lymphocytes. This approach can work as long as the antigen in question forms a β -turn in the original protein [35, 36]. In this example we used the PND of HIV-1 (MN), the residues GPGR have been shown by molecular modeling, crystallography and NMR to form a type II β -turn structure [5, 7-10]. In addition to the PND of HIV, the immunodominant segments of the transmembrane proteins (TM) or the tips of other surface unit (SU) protein loops of lentiviruses would make good candidates for incorporation into mucin chimeras [37-41].

We showed that human monoclonal antibodies derived from HIV infected patients bind more efficiently to the PND in the cyclic form [26]. Our ELISA results suggest that antigens with seven to nine residues of the V3 loop, constrained in a β -

turn conformation and presented in the context of the multi-valent MUC1, efficiently detect the same levels of V3 loop specific antibodies as full length V3 loop peptides.

A possible mechanism for the enhanced ability of mucin/HIV-PND chimeras to detect antibodies involves the process of antigen adsorption to a plastic plate. When a multi-valent antigen is bound to a surface, there are additional sterically available epitopes that can bind antibodies [42, 43]. An alternative mechanism is suggested by considering the kinetic effects of a multivalent antibody interacting with a multivalent antigen. The kinetic analysis allows the relative importance of the association and dissociation rates in the antigen-antibody interaction to be discerned. We observed a 10-fold decrease in the rate of dissociation when the same antibody interacted with the 120 residue or six knob peptide as compared to the 60 residue three knob peptide. When an antibody binds to the PND epitope on one of the knobs of the 120 residue peptide there are 5 additional knobs available for binding the remaining arm of the antibody. Since the minimal distance between two antibody combining sites is 35Å [44] and the maximal distance between two knobs is 32Å [16]. One antibody may bind to two knobs only in the 120 residue (6 knob) peptide. In the 60 residue peptide, the N-terminal knob has a skewed orientation due to the lack of the adjacent sequence, which may interfere with multivalent binding. Differences in multi-valent binding may change the overall equilibrium constant for the antibody-antigen interaction.

Another intriguing idea is the use of multivalent mucin/HIV-PND chimeras as immunogens, either as a component of a multi-subunit vaccine, or possibly as an immunotherapeutic for HIV infected patients. Berberian et al. recently found that the immunoglobulin VH3 gene family, found in a subset of IgM⁺ B cells, contains a conserved idiotope for HIV-gp120 [45]. The authors show that gp120 preferentially binds to EBV-immortalized B cells, and immortalized tonsil mantle zone B cells

that express the V_H3 idiotope. The binding of gp120 could activate this subpopulation of B cells and substitute for the normal Ig ligand. Fractions of IgM from normal individuals bound gp 120 (K_d 8.6 X 10⁻⁹M) and monoclonal IgM antibodies with non-HIV specificity were found which neutralize the virus. In addition, the authors report a correlation between anti-gp120 IgM and the clinical stage of HIV infection [45]. Our results indicate that the 120 residue peptides with 6 knobs are immunogenic and can induce both high titer IgM and IgG antibodies specific for only HIV-1. Unfortunately, the antisera produced to date have not been neutralizing. We are preparing multivalent antigens with added T-cell epitopes in order to assess the effect of T-dependent versus T-independent induction of antibody on the neutralizing ability of the sera.

Acknowledgements

The authors would like to acknowledge that some monoclonal antibodies were supplied by the NIH AIDS Research and Reference Reagent program. We also thank Dr. Goutam Gupta for valuable discussions and for providing the cyclic and linear forms of the V3 loop peptides of HIV-1 (MN) isolate.

References

1. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ: Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* 1990, 265:10373-10382.
2. Javaherian K, Langlois AJ, McDanal *Cet al.*: Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc. Natl. Acad. Sci. USA* 1989, 86:6768-6772.
3. Rusche JR, Javaherian K, McDanal *Cet al.*: Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. *Proc. Natl. Acad. Sci. USA* 1988, 85:3198-3202.
4. Profy AT, Salinas PA, Eckler LI, Dunlop NM, Nara PL, Putney SD: Epitopes recognized by the neutralizing antibodies of an HIV-1-infected individual. *J. Immunol.* 1990, 144:4641-4647.
5. Lorimier Rd, Moody MA, Haynes BF, Spicer LD: NMR-Derived solution conformations of a hybrid synthetic peptide containing multiple epitopes of envelope protein gp120 from the RF strain of human immunodeficiency virus. *Biochemistry* 1994, 33:2055-2062.
6. LaRosa GJ, Davide JP, Weinhold *Ket al.*: Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science* 1990, 249:932-935.
7. Gupta G, Anantharamaiah GM, Scott DR, Eldridge JH, Meyers G: Solution structure of the V3 loop of a Thailand HIV isolate. *J. Biomol. Struct. & Dynam.* 1993, 11:345-366.

8. Gupta G, Meyers G. Analyses of various folding patterns of the HIV-1 loop . In *PEPTIDES: Design, Synthesis, and Biological Activity* . Edited by Basava C, Anantharamaiah GM, Boston:Birkhauser; 1994.
9. Ghiara JB, Stura EA, Stanfield RL, Profy AT, Wilson IA: Crystal Structure of the Principal Neutralization site of HIV-1. *Science* 1994, 264:82-85.
10. Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA: Crystal Structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl. Acad. Sci. USA* 1993,
11. Gorny MK, Xu JY, Gianakakos Vet *al.*: Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* 1991, 88:3238-3242.
12. Catasti P, Fontenot JD, Bradbury EM, Gupta G: Local and global structural properties of the HIV-MN V3 loop. *J. Biol. Chem.* 1995, 270:2224-2232.
13. Fontenot JD, Tjandra N, Bu D, Ho C, Montelaro RC, Finn OJ: Biophysical characterization of one-, two-, and three-tandem repeats of Human mucin (muc-1) protein core. *Cancer Res.* 1993, 53:5386-5394.
14. Fontenot JD, Finn OJ, Dales N, Andrews PC, Montelaro RC: Synthesis of large multi-determinant immunogens using a poly-proline β -turn helix motif. *Pept. Res.* 1993, 6:330-336.
15. Fontenot JD, Mariappan SVS, Catasti P, Domenech N, Finn OJ, Gupta G: Structure of a tumor associated antigen containing a tandemly repeated immunodominant epitope. *In Press Jl. Biomol. Struct. & Dynam.* 1995,
16. Fontenot JD, Gatewood JM, Mariappan SVSet *al.*: Human immunodeficiency virus (HIV) antigens: Structure and serology of multivalent human mucin muc-1 chimeric proteins. *Proc. Natl. Acad. Sci. USA* 1995, 92:315-319.

17. Strouss GJ, Dekker J: Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* 1992, 27(1/2):57-92.
18. Devine PL, McKenzie IFC: Mucins: structure, function, and associations with malignancy. *BioEssays* 1992, 14:619-625.
19. Lan ML, Batra SK, Qi WN, Metzger RS, Hollingsworth. MA: Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J. Biol. Chem.* 1990, 265:15294-15299.
20. Gendler SJ, Burchell JM, Duhig Tet al.: Cloning of a partial cDNA encoding differentiation and tumor associated mucin glycoproteins expressed by human mammary epithelium. *Proc. Natl. Acad. Sci. USA* 1987, 84:6060-6064.
21. Taylor-Papadimitriou J: Report on the first international workshop on carcinoma associated mucins. *Int. J. Cancer* 1991, 49:1-5.
22. Rughetti A, Turchi V, Ghetti CAet al.: Human B-cell immune response to the polymorphic epithelial mucin. *Cancer Res.* 1993, 53:2457-2459.
23. Kotera Y, Fontenot JD, Pecher G, Metzgar RS, Finn OJ: Humoral Immunity against a tandem repeat epitope of human mucin muc-1 in sera from breast, pancreatic, and colon cancer patients. *Cancer Res.* 1994, 54:2856-2860.
24. Pau CP, Kai M, Holloman-Candal DLet al.: Antigenic variation and serotyping of HIV type 1 from four world health organization-sponsored HIV vaccine sites. *AIDS Res. Hum. Retroviruses* 1994, 10:1369-1377.
25. Myers G, Korber B, Berzofsky JA, Smith RF, Pavlakis GN: Human Retroviruses and AIDS, 1992. In *A compilation and analysis of nucleic acid and protein sequences*. Los Alamos National Laboratory, Los Alamos, NM.
26. Gorny MK, Xu JY, Karwowska S, Buchbinder A, Zolla-Pazner S: Repertoire of neutralizing human monoclonal antibodies specific for the V3 domain of HIV-1 gp120. *J. Immunol.* 1993, 150:635-643.

27. Skinner MA, Ting R, Langlois AJ *et al.*: Characteristics of a neutralizing monoclonal antibody to the HIV envelope glycoprotein.
AIDS Res. Hum. Retroviruses 1988, 4:187-197.
28. White-Scharf ME, Potts BJ, Smith LM, Sokolowski KA, Rusche JR, Silver S: Broadly neutralizing monoclonal antibodies to the V3 region of HIV-1 can be elicited by peptide immunization. *Virology* 1993, 192:197-206.
29. Pau CP, Lee-Thomas S, Auwanit W *et al.*: Highly specific V3 peptide enzyme immunoassay for serotyping HIV-1 specimens from Thailand.
AIDS 1993, 7:337-340.
30. Karlsson R, Michaelsson A, Mattsson L: Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system.
J. Immunol. Methods 1991, 145:229-240.
31. VanCott TC, Loomis LD, Redfield RR, Birx DL: Real-time biospecific interaction analysis of antibody reactivity to peptides from the envelope glycoprotein, gp160, of HIV-1. *J. Immunol. Methods* 1992, 146:163-176.
32. VanCott TC, Bethke FR, Polonis V *et al.*: Dissociation rate of antibody-gp120 binding interactions is predictive of V3-mediated neutralization of HIV-1.
J. Immunol. 1994, 153:449-459.
33. VanCott TC, Bethke FR, Kalyanaraman V, Burke DS, Redfield RR, Birx DL: Preferential antibody recognition of structurally distinct HIV-1 gp120 molecules. *In Press J. AIDS*
34. Gorny MK, Xu JY, Karwowska S, Buchbinder A, Zolla-Pazner S: Repertoire of neutralizing human monoclonal antibodies specific for the HIV-1 gp120.
J. Immunol. 1993, 150:635-643.

35. Parker JMR D, Guo, and R. S. Hodges: New hydrophilicity scale derived from high performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray derived accessible sites. *Biochemistry*. 1986, 25:5425-5431.
36. Chou PY, Fasman GD: β -Turns in proteins. *J. Mol. Biology* 1977, 115:135-175.
37. Wang JJG, Steel S, Wisniewolski R, Wang CW: Detection of antibodies to human T-lymphotropic virus type III by using a synthetic peptide of 21 amino acid residues corresponding to a highly antigenic segment of gp41 envelope protein. *Proc. Natl. Acad. Sci. USA* 1986, 83:6159-6163.
38. Norrby E, Biberfeld G, Chiodi Fet *al.*: Discrimination between antibodies to HIV and to related retroviruses using site directed serology. *Nature* 1987, 329:248-250.
39. Gnann JWJ, McCormick JB, Mitchell S, Nelson JA, Oldstone MBA: Synthetic peptide immunoassay distinguishes HIV-1 and HIV-2 infections. *Science* 1987, 237:1346-1349.
40. Chong YH, Ball JM, Issel CJ, Montelaro RC, Rushlow KE: Analysis of equine humoral immune responses to the transmembrane envelope glycoprotein (gp45) of equine infectious anemia virus. *J. Virol.* 1991, 65:1013-1018.
41. Fontenot JD, Hoover EA, Elder JH, Montelaro RC: Evaluation of feline immunodeficiency virus and feline leukemia virus transmembrane peptides for serological diagnostics. *J. Clin. Microbiol.* 1992, 30:1885-1890.
42. Dierks SE, Butler JE, Richerson HB: Altered recognition of surface-adsorbed compared to antigen bound antibodies in ELISA. *Mol. Immunol.* 1986, 23:403-411.
43. Ansari AA, Hattikudur NS, Joshi SR, Medeira MA: ELISA solid phase: stability and binding characteristics. *J. Immunol. Methods* 1985, 84:117-124.

44. Davies DR, Metzger H: The structural basis of antibody function. *Ann. Rev. Immunol.* 1983, 1:87-117.
45. Berberian L, Goodglick L, Kipps TJ, Braun J: Immunoglobulin V_H³ gene products: natural ligands for HIV gp120. *Science* 1993, 261:1588-1591.

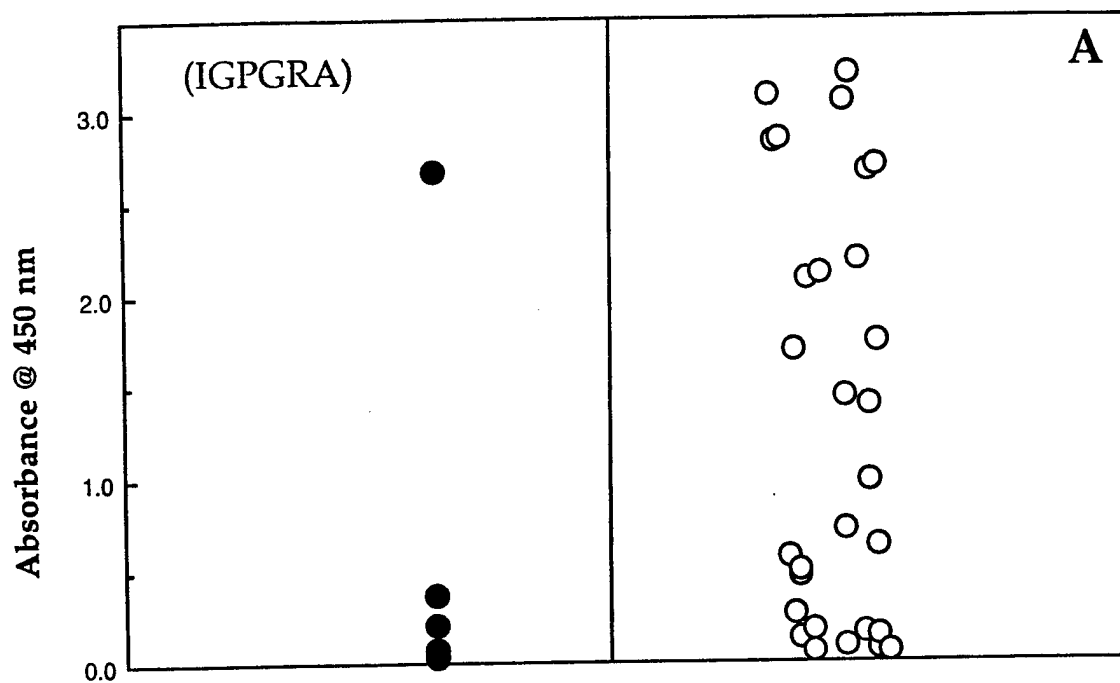
Figure Legends

Figure 1. Serological reactivity by clade B and non-clade B HIV infected patients to 120 residue mucV3 chimeras: (A) mucV3-6 which contains six (IGPGRA) residues, (B) mucV3-7 which contains seven (HIGPGRA) , and (C) mucV3-9 which contains nine (IHIGPGRAF) residues of the HIV-V3 loop. Strong, Intermediate, and Weak serological reactions correspond to the development 2.5-3.0, 0.5-2.5, or 0.0-0.5 Absorbance units after 10 minutes.

Figure 2. Serological reactivity of clade B and non-clade B HIV infected patients to peptide antigens: (A) the full 36 residue HIV-MN V3 loop, (B) a 15 residue linear peptide (DKRIHIGPGRAFYTT) of the HIV-MN V3 loop. Strong, intermediate, and weak serological reactions correspond to the development 2.5-3.0, 0.5-2.5, or 0.0-0.5 Absorbance units after 10 minutes.

Figure 3. Endpoint ELISA titers of HIV-1 sera against the cyclic and linear V3 loops, the 15 residue linear peptide, the 120 residue six knob chimeras mucV3-6, mucV3-7, and mucV3-9. (A) Serum from a patient with a clade B PND sequence of IHIGPGRAF. (B) Serum from a patient with a non-clade B PND sequence of (IHMGWGAFY).

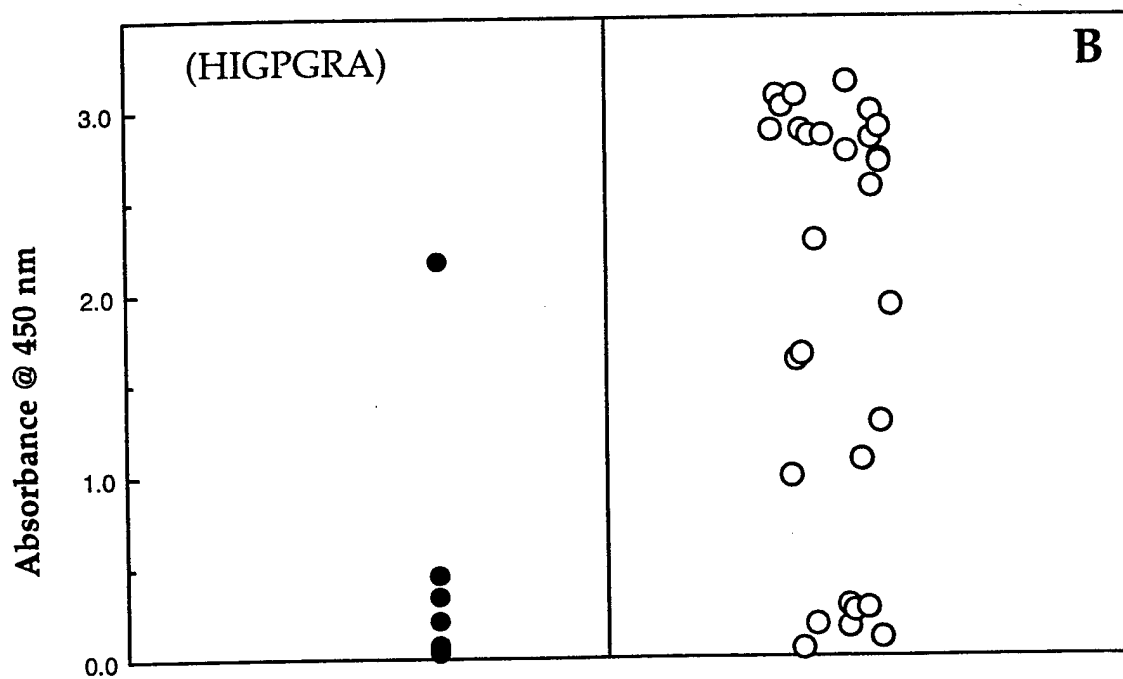
Figure 4. (A) Endpoint ELISA titers of mice immunized with 2, 10, and 20 nmoles of either the six knob 120 residue antigen, mucV3-9 (120), and the three knob 60 residue antigen, mucV3-9 (60). (B) Cross reactivity of sera from mice immunized with 10 nmoles of mucV3-9 (120) with the cyclic and linear HIV (MN) V3 loops. All points are the average of 4 mice.



Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 13

TOP ↑

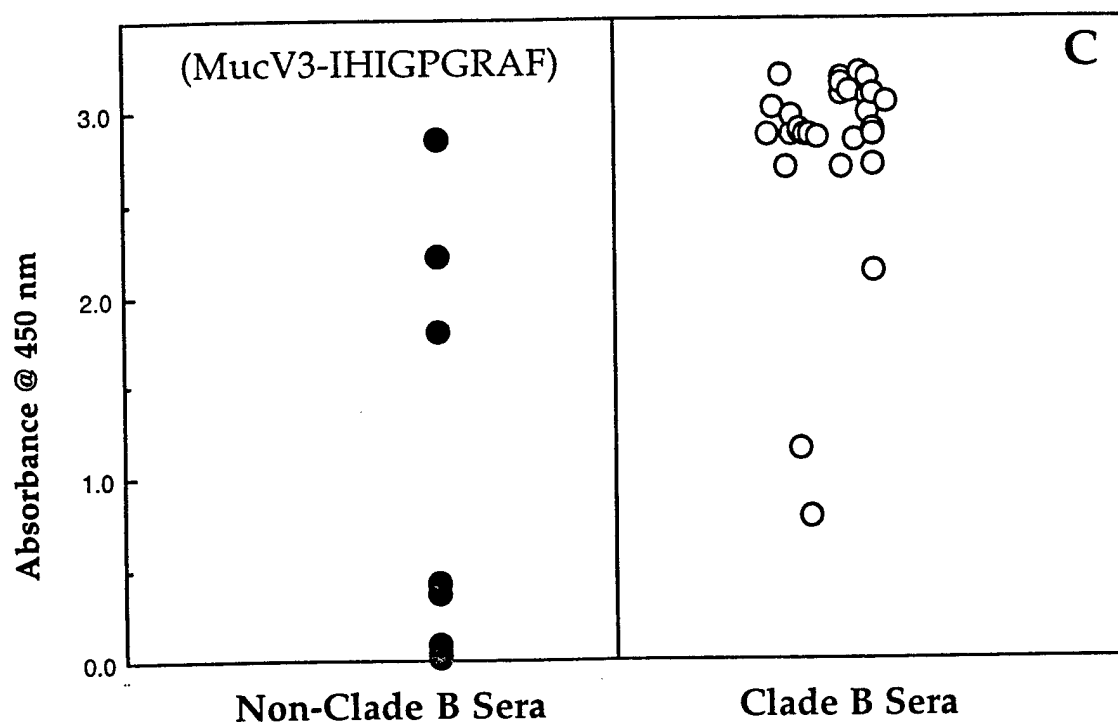


Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 13

TOP





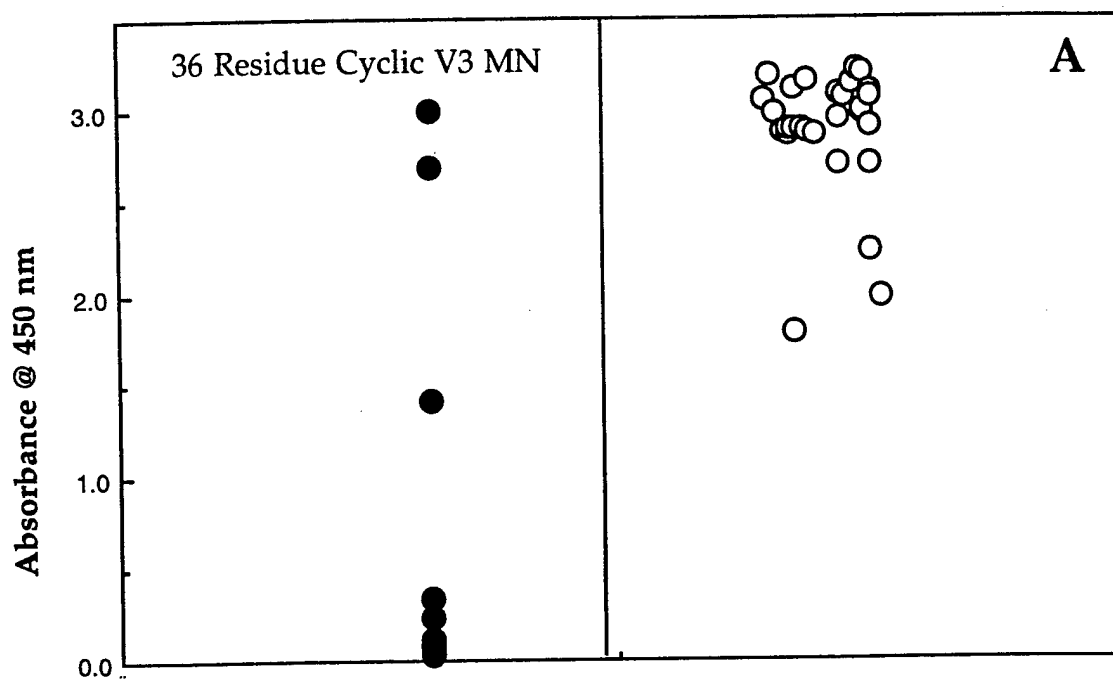
Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG #

1C

TOP

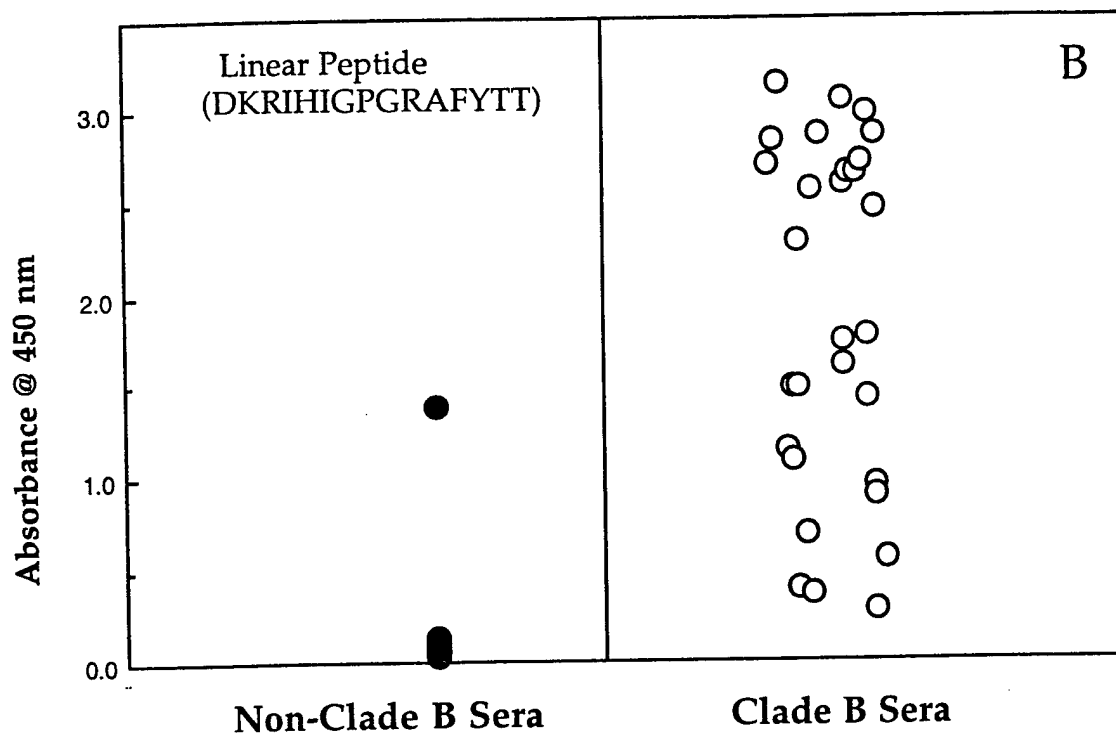
↑



Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 2A

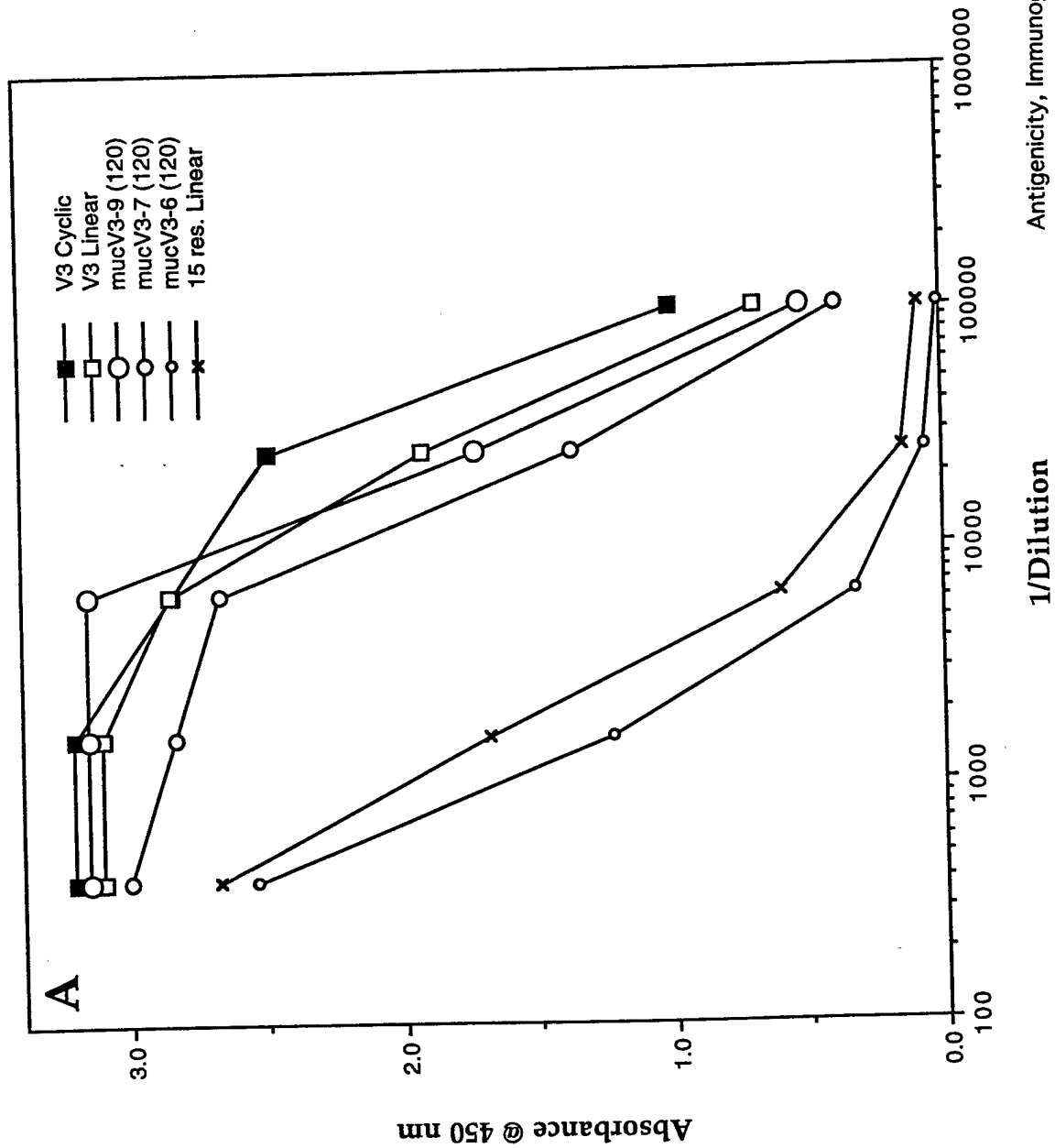
TOP ↑



Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 23

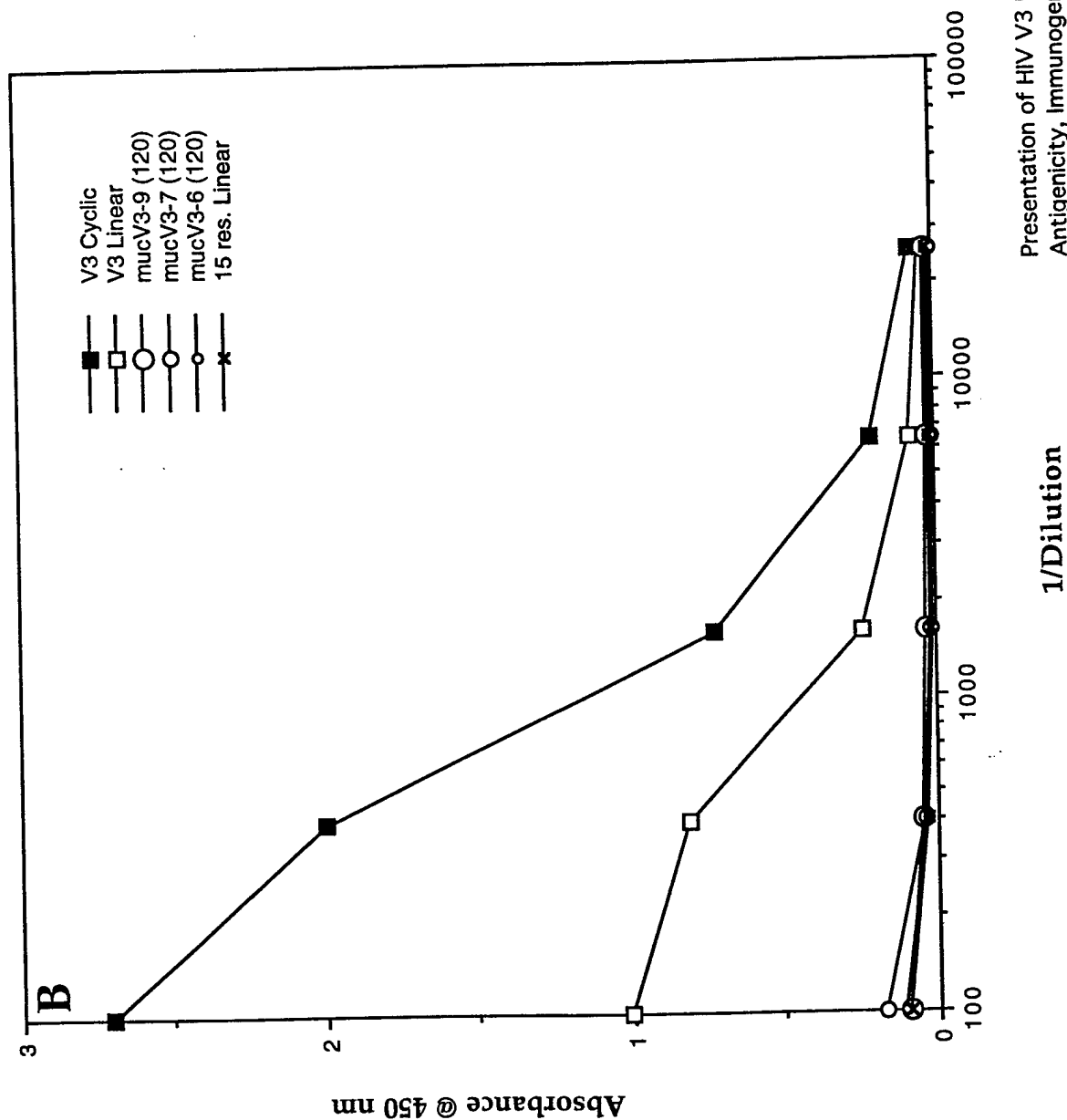
TOP ↑



Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 3A

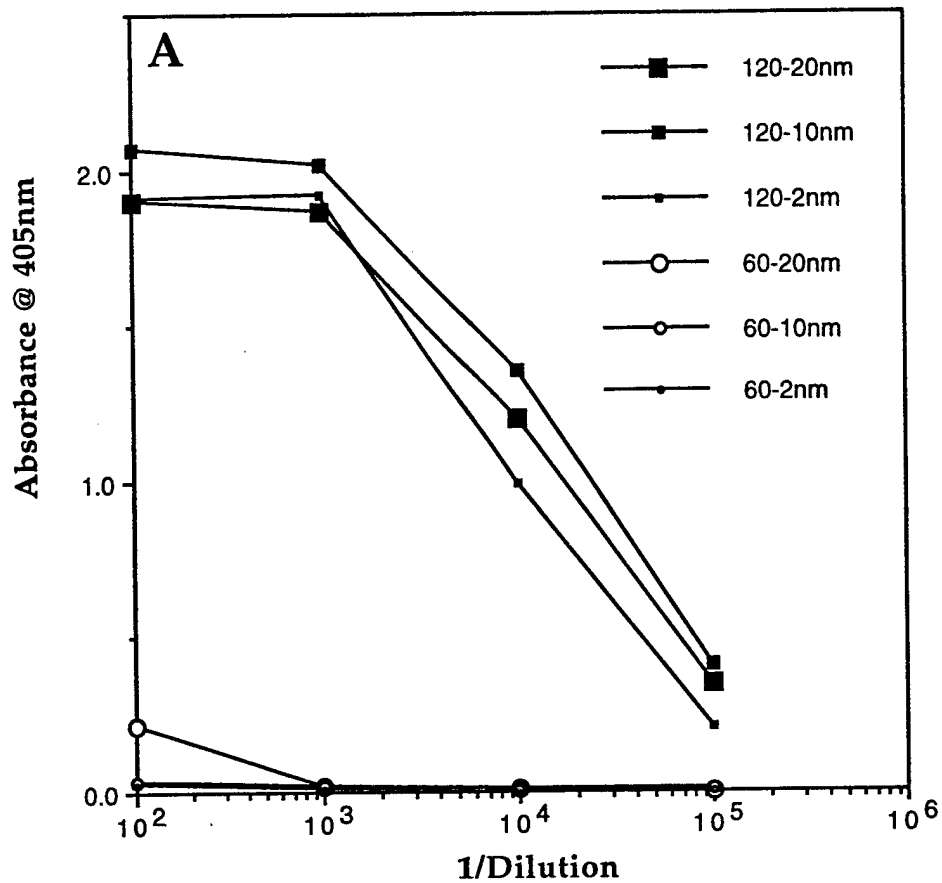
TOP ↑



Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 3B

TOP ↗

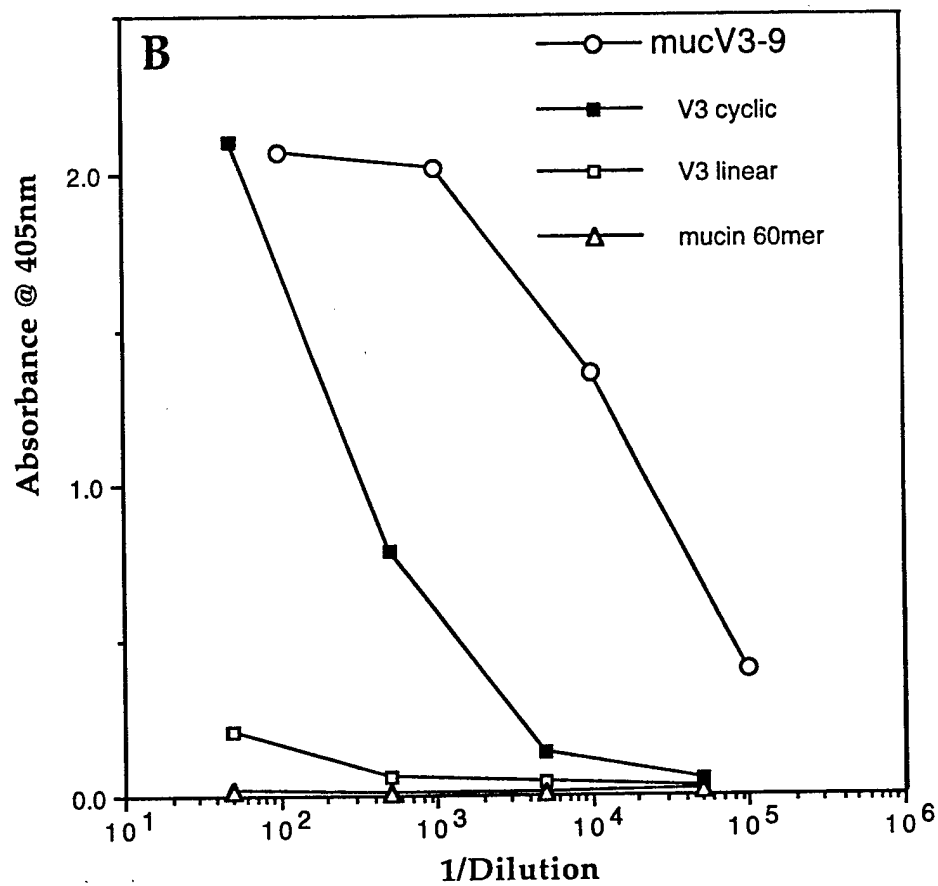


Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 4A

TOP





Presentation of HIV V3 Loop Epitopes for Enhanced
Antigenicity, Immunogenicity and Diagnostic Potential.
FONTENOT et al.

FIG # 4B

TOP ↑

Table 1

Mucin V3 Loop Chimeras

Sequence of 1 Tandem Repeat		No. Residues
	1 2 3 4 5 6 7 8 9 10 12 14 16 18 20	
mucin	(V T S A P D T R P A P G S T A P P A H G) 5.25	105
mucV3-6 (60)	(- - - G P G R A F - - - - - - - - - - -) 3	60
mucV3-6 (120)	(- - - G P G R A F - - - - - - - - - - -) 6	120
mucV3-7 (60)	(- H I G P G R A - - - - - - - - - - -) 3	60
mucV3-7 (120)	(- H I G P G R A - - - - - - - - - - -) 6	120
mucV3-9 (60)	(I H I G P G R A F - - - - - - - - - - -) 3	60
mucV3-9 (120)	(I H I G P G R A F - - - - - - - - - - -) 6	120

Table 2

Monoclonal Antibody Reactivity Preference for Cyclic V3 Loop

No	mAB	epitope	V3 cyclic	V3 linear	% differ
1	447	HIGPGRAF	4.0	3.2	25
2	412		2.4	2.1	14
3	453	KRIHIGPGR	3.1	2.8	11
4	386	RIHIGPGR	3.6	3.2	13
5	9205		4.7	4.0	17
6	268	RIHIGPGR	4.7	4.0	18
7	782		4.8	4.1	17
8	50.1	KRRIHIGPG	4.5	3.6	25
9	257	KRKRIHIGP	5.2	4.4	18
10	391		4.0	3.1	29
11	838		2.8	2.2	27
12	419		2.0	1.6	25
13	181		2.9	2.4	21
14	908		5.6	4.6	22
15	537		0.9	0.8	13
Negative Controls					
16	9284		0.01	0.01	
17	NHS		0.03	0.02	

Reactivity = Response units (RU) for antibody-antigen interaction, divided by the baseline increase in Response Units after antigen loading (measure of the amount of antigen on the matrix).

Table 3

Kinetic Constants for mABs Binding to MucV3 Chimeras

mAB	epitope	$K_1 \text{sec}^{-1} \text{M}^{-1} (\text{assoc})$	$K_{-1} \text{sec}^{-1} (\text{diss})$
447	HIGPGRAF		
	MucV3-9-60	6.9×10^4	4.6×10^{-4}
	MucV3-9-120	5.9×10^4	4.4×10^{-5}
412			
	MucV3-9-60	2.8×10^4	2.2×10^{-3}
	MucV3-9-120	2.9×10^4	2.0×10^{-4}
453	KRIHIGPGR		
	MucV3-9-60	1.1×10^5	1.2×10^{-3}
	MucV3-9-120	6.8×10^4	2.3×10^{-4}

Table 4**Average Titers of Mice Immunized with MucV3 Chimeras**

MucV3 Chimeras	IgG Titer	IgM Titer
1.1-120	5×10^4	4×10^2
1.2-120	5×10^2	5×10^1
1.3-120	3×10^6	1×10^4
mucin 105	0	3×10^3

Structural Studies Involving Different HIV-1 V3 Loops

Paolo Catasti and Goutam Gupta

MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Introduction

Studies on the feasibility of a subunit vaccine to protect against human immunodeficiency virus (HIV) infection have principally focused on the third variable (V3) loop of the envelope surface protein. One of the neutralizing determinants of HIV-1 is located inside the V3 loop. However, progress toward a vaccine based on neutralizing determinants has been impeded by the amino acid sequence variability in the V3 loop of different HIV isolates. The elusive nature of the V3 loop structure prompted us to carry out a systematic study on different isolates in an attempt to identify a common structural motif in the V3 loop regardless of the amino acid sequence variability. We have performed 2D NMR structural studies on three different V3 loop peptides: MN, Haiti, and RF (Catasti *et al.*, 1995 & 1996). The three V3 loops were all 35 residues long and S-S bridged at the terminals. The NMR studies were carried out first in water, then in a 70%/30% mixture of water/trifluoroethanol 1 (TFE). TFE is a solvent wide

????????? ?????????? ?????????? ?????????? ?????????? ??????????

Figure 1 shows that similar secondary structures are observed for the three different V3 loops: a GPG(K/R) crest in the center of the neutralizing determinant, two extended regions flanking the central crest, and a helical region in the C-terminal domain observed only in the water/TFE mixture. The RF V3 peptide did not dissolve in the water/TFE mixture, therefore we could run the experiments only in an aqueous solution. Structural prediction studies revealed that the variability in sequence and structure of the V3 loop is confined to the N and C-terminal side of the conserved GPG crest. Figure 2 is a summary of the NMR secondary structural assignments (Catasti *et al.*, 1995 & 1996), and the results of several secondary prediction algorithms. With the exception of the PSA method, most of the algorithms fail to identify the alpha helix in the C-terminal portion of the V3 loops.

References

- Catasti, P., Bradbury, E.M., and Gupta G. (1996) Structure and polymorphism of HIV-1 third variable loops. *J. Biol. Chem.*, **271**(14):8236-8242.
- Catasti, P., Fontenot, J.D., Bradbury, E.M., and Gupta G. (1995) Local and global structural properties of the HIV-MN V3 loop. *J. Biol. Chem.*, **270**(5):2224-2232.
- Deleage, G., and Roux, B. (1987) An algorithm for protein secondary structure prediction based on class prediction. *Prot. Eng.*, **1**:289-294.
- Geourjon, C., and Deleage, G. (1994) SOPM : a self optimised prediction method for protein secondary structure prediction. *Prot. Eng.*, **7**:157-164.
- Geourjon, C., and Deleage, G. (1995) SOPMA : Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *CABIOS*, **11**:681-684
- Gibrat, J.F., Garnier, J., and Robson, B. (1987) Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.*, **198**(3):425-443.
- King, R.D., and Sternberg, M.J.E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Prot. Science*, in press.
- Kneller, D.G., Cohen, F.E., and Langridge, R. (1990) Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *J. Mol. Biol.* **214**:171-182.
- Levin, J.M., Robson, B., and Garnier, J. (1986) An algorithm for secondary structure determination in proteins based on sequence similarity. *Febs Lett.*, **205**(2):303-308.

V3 Loop Structure

- McClelland, J.L., and Rumelhart, D.E. (1988) *Explorations in Parallel Distributed Processing* vol 3. pp 318–362. MIT Press, Cambridge MA.
- Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth. in Enzym.*, **266**:525–539.
- Stultz, C.M., White, J.V., and Smith, T.F. (1993) Structural Analysis Based on State-space Modeling. *Prot. Science*, **2**:305–314.
- White, J.V., Stultz, C.M., and Smith, T.F. (1994) Protein Classification by Stochastic Modeling and Optimal Filtering of Amino-Acid Sequences. *Math. Biosc.*, **119**:35–75.

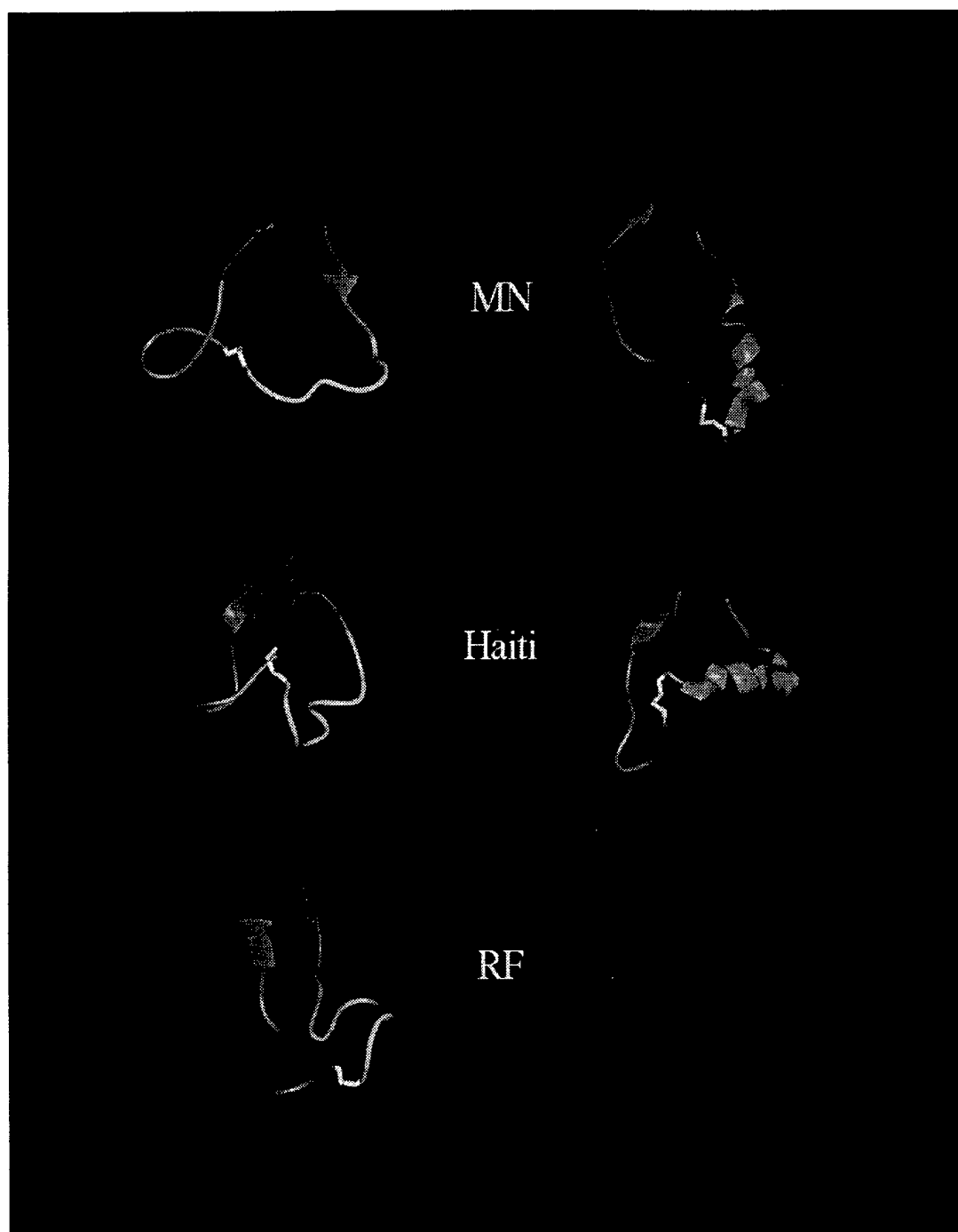


Figure 1. Ribbon diagram showing the average folding patterns of the structures of the MN, Haiti and RF V3 loops in water and in a mixture of 70%/30% water/TFE. In each case the average is done over 70 sampled low energy structures. Note that, in each case, the neutralizing epitope containing the central GPG(R/K) sequence forms a protruding loop even though the local structure and presentation of the loop in the different cases are noticeably different. Structures that satisfy the NMR constraints of the V3 loops in water show a higher degree of flexibility than those in agreement with the NMR data in the mixed water/TFE solvent. This is due to the formation of the alpha helix in the mixed solvent. Color code is as follow: GPG(R/K) crest is red, extended regions are green, disulfide bridges are yellow and the alpha helical region is cyan.

V3 Loop Structure

	1	10	20	30
V3MN	CTRPNYNKRKRIHIGPGRIFYTTKNIIGTIRQAHC			
NMR (water)	.EECTTTTCCEEEETTTTEEEBCCCTTTTTC.			
NMR (TFE)	.EECTTTTEEEECCTTTTCCCCCHHHHHHHHHHC.			
PSA method	CCCCCCCCCEEEECCTCCCCCHHHHHHHHHH			
Gibrat method	ECCCCCCEEEEECCCCCEEEEEEHEEEBCCC			
Levin method	.TSSSSCSSSCEEECTTCCEEEBCCCCCHHHSSC			
DPM method	CCCTTCTCHHHECCTCCCHHCEECCEEEHHHC			
SOPMA predict	ECCTTHCTEEEEEECCCCCTTTCCCTTCCCC			
DSC method	CCCCCCCCCEEEBCCCCCEEECCCCCCCCCCC			
PHD method	CCCCCCCCCEEE..CCC..EE..CC.....CCC			
nnpredict method	CCCCCCCCCCCCCCCCCEEEBCCCCCEECHHCCCC			
	1	10	20	30
V3Haiti	CTRPNDNTRKSIPMGPGKAFYATGDIIGNIRQAHC			
NMR (water)	.CECTTTTEEECCCTTTTEBCCCCCTTTTTC.			
NMR (TFE)	.EECTTTTCCEEEETTTTCCTTTTCHHHHHHHH.			
PSA method	CCCCCCCCCCCCCCCCCCCCCHHCHHHHHHC			
Gibrat method	ECCCCCCCCCCCCCCCCCEEEHHEEEBECCE			
Levin method	TCCCCCSSCCCCCCTCCCECTSCCSCSCCHCCC			
DPM method	CCCCTCTCCCCCCTCCCHHCEECCEHHHC			
SOPMA predict	ECCCCTTCCCCETTTCHHHHHETHHHECCTTCC			
DSC method	CCCCCCCCCCCCCCCCCEEECCCCCCCCCCC			
PHD method	CCCCCCCC....CCCC..EE..CCC.....C			
nnpredict method	CCCCCCCCCCCCCCCCCEEECCCCCEECHHCCCC			
	1	10	20	30
V3RF	CTRPNNNTRKSITKGPGRVIYATGQIIGDIRKAHC			
NMR (water)	.CTTTTEEECECCCTTTTEECCTTTTCTTTC.			
PSA method	CCCCCCCCCCCCCCCCCEEEBCHHHHHHHHH			
Gibrat method	ECCCCCCEEEBCCCCCEEEBCEEEBEEHCC			
Levin method	CCCCCCHSHS..CCCTCEEEBCCCCSCCCCC.			
DPM method	CCTTCTCCCCCCTCCCEEEBCEECCEHHHC			
SOPMA predict	ECCCCCCEEEBCCCCCEEEETTCEEEHHHHHH			
DSC method	CCCCCCCCCEEEBCCCCCEEEBCCCCCCCCCCC			
PHD method	CCCCCCCC..E..CCCC..EEE..CC..EE.....C			
nnpredict method	CCCCCCCCCCCCCCCCCEEEBCCCCCEECHHCCCC			

Figure 2. Comparison of secondary structure assignments of the NMR determined structures and secondary structure prediction for the three V3 loops, MN, Haiti and RF. The different prediction algorithms are indicated on the left. Some of these methods are discussed by Myers and Farmer in Part III of this compendium.

Meaning of Symbols

H	alpha helix	T	turn
C	random coil/loop	S	bend
E	strand	.	unassigned

Key to Prediction Algorithms

PSA	Stultz <i>et al.</i>	Gibrat	Gibrat <i>et al.</i>
Levin	Levin <i>et al.</i>	DPM	Deleage <i>et al.</i>
SOPM	Geourjon <i>et al.</i>	DSC	King <i>et al.</i>
PHD	Rost <i>et al.</i>	nnpredict	McClelland <i>et al.</i>

Human immunodeficiency virus (HIV) antigens: Structure and serology of multivalent human mucin MUC1–HIV V3 chimeric proteins

(principal neutralizing determinant/multivalent antigen)

J. DARRELL FONTENOT*, JOE M. GATEWOOD†, S. V. SANTHANA MARIAPPAN*, CHOU-PONG PAU‡, BHARAT S. PAREKH‡, J. RICHARD GEORGE‡, AND GOUTAM GUPTA*§

*Theoretical Biology and Biophysics, T-10, MS-K710, †Life Sciences Division, LS-2, MS 880, Los Alamos National Laboratory, Los Alamos, NM 87545; and ‡Division of HIV/AIDS, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30330

Communicated by Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, NM, September 23, 1994 (received for review May 26, 1994)

ABSTRACT Molecular modeling and two-dimensional NMR techniques enable us to identify structural features in the third variable region (V3) loop of the human immunodeficiency virus (HIV) surface glycoprotein gp120, in particular the principal neutralizing determinant (PND), that remain conserved despite the sequence variation. The conserved structure of the PND is a solvent-accessible protruding motif or a knob, structurally isomorphous with the immunodominant knobs in the tandem repeat protein of human mucin 1 (MUC1) (a tumor antigen for breast, pancreatic, and ovarian cancer). We have replaced the mucin antigenic knobs by the PND knobs of the HIV MN isolate in a set of chimeric human MUC1/HIV V3 antigens. This produced multivalent HIV antigens in which PNDs are located at regular intervals and separated by extended mucin spacers. In this article we show by two-dimensional NMR spectroscopy that the multivalent antigens preserve the PNDs in their native structure. We also demonstrate by ELISA that the antigens correctly present the PNDs for binding to monoclonal antibodies or polyclonal antisera from HIV-infected patients.

Fusion of a viral surface with the host-cell surface is the first step in the life cycle of the human immunodeficiency virus (HIV). The process of HIV fusion into a host cell is determined by two discrete functional sites on the 120-kDa viral surface unit glycoprotein, gp120. First, a segment near the C terminus of gp120 directly binds to the CD4 molecule on the surface of a host cell (1, 2). Next, the third variable region (V3) loop of gp120 mediates in virus fusion (3, 4). Virus fusion is abrogated when these binding events are prevented by blocking either one or both of these sites of gp120. The blocking of these sites is the primary mechanism of virus neutralization by antibodies (5–7). Two approaches have been used for viral neutralization: (i) vaccination aimed at generating antibodies specific for the principal neutralizing determinant (PND), which is located inside the V3 loop of gp120 and contains a fairly conserved Gly-Pro-Gly-(Arg or Gln) (GPGR/Q) crest at the center of the PND (6, 8); and (ii) the administration of soluble CD4 receptors (9). However, these approaches currently suffer from serious drawbacks. For example, because of sequence variability in the V3 loop (8, 10), neutralizing antibodies elicited by the V3 loop from one HIV isolate do not neutralize other HIV isolates (11, 12). Further, the short half-life of soluble CD4 molecules has limited applicability for the treatment of a “slow” virus infection.

The V3 loop of HIV surface gp120 is a potential target for protective immunity. This small 34- to 36-residue domain contains the well-characterized PND as well as a cross-reactive

cytotoxic T-lymphocyte epitope (13). The V3 loop also participates in vital functional properties of HIV-like cell tropism (14, 15) and cell fusion (3, 4, 16). Antibody binding to the V3 loop can interfere with these processes in the life cycle of the virus. To exploit the V3 loop as an antibody target, one has to understand (i) the effect of sequence variability on the structure and antigenicity and (ii) the structural features of the V3 loop (especially at the PND) that remain invariant. Our previous work using theoretical studies of 30 different V3 loops (17, 18) and three-dimensional (3-D) structure determination of two divergent cyclic V3 loops (19, 20) allows us to define the effect of sequence variability on the global structure of the entire V3 loop and the local structure of the PND centered around the conserved GPGR crest. Irrespective of the variability in the amino acid sequences on either side of the type II GPGR turn, the PNDs of different V3 loops adopt a protruding solvent-accessible motif (19–22) or knob. This work focuses on synthetic polypeptide antigens that contain PNDs that preserve and present the same conserved structural features as in the “native” V3 loop. Such a design requires a polypeptide construct that also contains a protruding motif like the PNDs in the V3 loop. The tandem repeat protein, human mucin 1 (MUC1) (a tumor antigen for breast, pancreatic, and ovarian cancer), provides us with such a structural motif (23–25). We construct a set of chimeric human MUC1–HIV V3 loop proteins in which HIV PNDs replace the immunodominant knobs of MUC1 tandem repeats. We show by two-dimensional (2-D) NMR that the PNDs in the chimeric proteins preserve the same structure as in the native V3 loops. Finally, we show that these antigens are also able to bind polyclonal antisera from HIV-infected patients and type-specific monoclonal antibodies (mAb).

MATERIALS AND METHODS

The 60- and 120-residue peptide amides were synthesized, and the products of the synthesis were deprotected, cleaved from the resin support, purified by HPLC, and subjected to molecular weight determination by electrospray mass spectroscopy as described (26). The names and sequences in single-letter amino acid code of one copy of the tandem repeat antigens described here are as follows: human MUC1, VTSAPDTR-PAPGSTAPPAHG; MUC1–V3 1.1, VTSGPGRAPAGST-APPAHG; MUC1–V3 1.2, HIGPGRAPAGSTAPPAHGV;

Abbreviations: HIV, human immunodeficiency virus; PND, principal neutralizing determinant; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; gp120, 120-kDa glycoprotein; V3, variable region third loop of HIV surface gp120; mAb, monoclonal antibody; MUC1, human mucin 1; GPGR/Q, Gly-Pro-Gly-(Arg or Gln); 2-D and 3-D, two- and three-dimensional; HIGPGR, His-Ile-Gly-Pro-Gly-Arg-Ala.

§To whom reprint requests should be addressed.

and MUC1-V3 1.3, IHIGPGRAFAPGSTAPPAHG. In each case, peptides with three and six complete tandem repeats were successfully synthesized. The cyclic and linear V3 peptides from the HIV isolate MN (HIV-MN) have the single-letter-code sequence CTRPNYNKRKRIHIGPGRAFYT^{TK}NIIGTIQAH^C, and the small 15-residue linear peptide has the sequence DKRIHIGPGRAFYT^T. The enzyme-linked immunosorbent assay (ELISA) was performed as described (27).

All NMR experiments on the MUC1-V3 60-residue antigen were carried out on the 600-MHz Bruker spectrometer at the University of Alabama at Birmingham. All NMR spectra were collected at 10°C with peptide concentration at 5 mM (pH 5.5). All 2-D data were acquired in the phase-sensitive mode with the presaturation of the ¹H₂O signal during the relaxation delay (28, 29).

As previously described (18, 19), a set of distance constraints were derived by analyzing the NMR data with the aid of full-matrix nuclear Overhauser effect (NOE) spectroscopy (NOESY) simulations, associated *R*-factor test, and energy calculations. Analyses of the 2-D NMR data of the MUC1-V3 60-residue peptide produced 220 interproton distance constraints. The energy term EDIST (14, 38) was added to the force-field as described by Scheraga and coworkers (30). Monte Carlo-simulated annealing (31) was performed to obtain a set of structures consistent with NMR data. The maximum step size of the torsion angles was set at 15 degrees, which produced an acceptance ratio of 0.20–0.50 for the 50,000-step Monte Carlo cycle at each temperature. Full-matrix NOESY calculations were repeated for the final 150 low-energy MUC1-V3 structures. These structures were analyzed in terms of their energies, end-to-end lengths (*R_e*), relative orientations of the knobs, and torsion angle parameters.

RESULTS

Principles of Design. We previously determined NMR structures of two disulfide-bridged V3 loops: a Thailand HIV

isolate (19) and the HIV-MN isolate (38). Amino acid sequences of the PNDs of these two V3 loops are quite different. However, as shown by superimposing two PNDs (Fig. 1 *Left*), both of them form similar protruding motifs. In the case of the HIV Thailand isolate, the central GPGQ forms a type II turn and a solvent-accessible tip; similarly in the case of HIV-MN, the central GPGR forms the type II turn and the accessible tip. In spite of the fact that the amino acid sequences flanking the central type II turn on the N- and C-terminal sides are different, the polypeptide backbone and side-chain orientations in the flanking regions are very similar in the two cases (Fig. 1 *Left*). However, when the PND is presented in the context of the V3 loop or gp120, the variability of regions flanking the central PND residues masks the conformational purity of the PND.

These protruding motifs (or knobs) are predicted from molecular modeling studies (17, 18) of a large set of different V3 loops. The “knob-like” structure is also present in the tandem repeat domain of a protein totally unrelated to HIV V3 loops. Fig. 1 *Right* illustrates the surprising result found upon solving the structure of three tandem repeats of the human breast and pancreatic tumor antigen MUC1, (PDTRPAPGSTAPPAHGVTSA)₃. In this system, we again found a knob-like motif that is created by a type II turn (isomorphous with the PND of HIV), which is immunodominant for humoral immune responses (32, 33). As shown by detailed NMR analyses (25), the elongated mucin structure contains knobs that project away from the long axis of the molecule and are connected by extended spacers (Fig. 1 *Right*). In the mucin structure, Ala-Pro-Asp-Thr from positions 0–3 in the above repeat (where Ala at position 0 is the amino acid from the previous repeat) forms the type II turn occupying the solvent-accessible tip of the knob; this tip is the immunodominant antigenic site in MUC1. In our antigen-engineering approach, we chemically synthesize a series of MUC1-V3 chimeric polypeptides in which mucin immunodominant knobs are replaced by the HIV-MN PNDs (Fig. 1 *Right*). The ability of these chimeric proteins to act as HIV antigens requires fulfillment of two criteria: (i) the PNDs in MUC1-V3 antigens

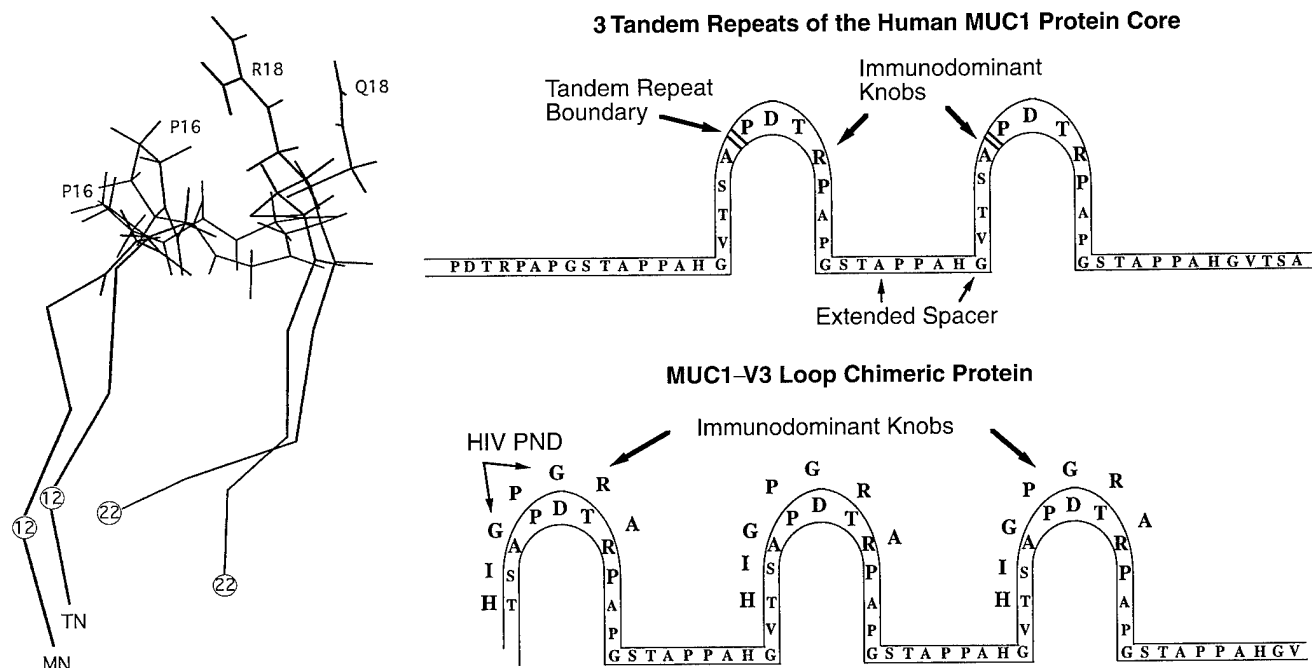


FIG. 1. (*Left*) Superimposition of the protruding motifs of two NMR structures: the V3 loop from the HIV-MN isolate (designated MN) and that from the Thailand TN243 isolate (named TN). The sequences of two motifs in single-letter amino acid code are: MN, RIHIGPGRAFYT^{TK}NIIGTIQAH^C; and TN, SITIGPGQVFYR. Note that the GPGR or GPGQ crests are oriented in the same way. (*Right*) The principle of design. The HIV PND sequences above the MUC1 sequences actually replace the MUC1 residues in the chimeras.

must be structurally equivalent with the native V3 loop, and (ii) the surface accessibility to antibodies of the PNDs in MUC1-V3 antigens must be as good as or better than in the native V3 loop. Below, we discuss how these criteria are fulfilled for the MUC1-V3 antigens that we have designed so far.

Preservation of Structure. We have chemically synthesized 60- and 120-residue peptides of three different MUC1-V3 antigens in which three different lengths of HIV-MN PND sequences are inserted. The lengths of the PND inserts are 6, 7, and 9 amino acids. The names and sequences of the chimeric antigens are listed in *Materials and Methods*. The start of the sequence is chosen in such a way that each 20-amino acid repeat has one knob (Fig. 1 *Right*). Therefore, by varying the number of repeats, we can vary the number of knobs and study the related effect on the structure and dynamics of the antigen and on the associated antibody-binding affinity. The binding data indicates that we need at least 7 PND amino acids for optimum binding.

The NOESY (400 ms of mixing) fingerprint region (containing cross-peaks due to coupling between backbone H^N and H^α) of the MUC1-V3 1.2 60-residue peptide reveals that the 20 residues in the sequence repeat also forms the structural repeat—i.e., the protons from the 20 amino acids in all three repeats sample the same chemical-shift environment and structure (Val-60 being the only exception). Comparison of the NOESY cross-section from the MUC1-V3 1.2 60-residue peptide with that of the MUC1-V3 1.2 120-residue peptide suggests that local structures of different 20-amino acid repeats in the same antigen (60 or 120 residue) are remarkably similar. However, the extent of flexibility and the nature of dynamics can be different in these two antigens. The diagnostic NOE pattern of importance is the one involving the HIV PND epitopes, His-Ile-Gly-Pro-Gly-Arg-Ala (HIGPGR), of MUC1-V3 1.2. The $H^\alpha(P_i)-H^N(G_{i+1})$, $H^\alpha(P_i)-H^N(R_{i+2})$, and $H^N(G_i)-H^N(R_{i+1})$ NOEs are indicative of a type II β -turn of GPGR centered around the proline and glycine residues (19, 34, 35). The sequential NOE pattern involving histidine, isoleucine and alanine residues of HIGPGR defines the backbone and side-chain orientations of the flanking residues in the PND with respect to the central GPGR type II turn. The NOESY (200 ms of mixing) fingerprint (H^N-H^α) region of MUC1-V3 1.2 120-residue peptide has features similar to those in the MUC1-V3 1.2 60-residue peptide except that the weak $H^\alpha(P_i)-H^N(R_{i+2})$ NOE present is not seen.

The analyses of the NMR data of MUC1-V3 1.2 60-residue peptide revealed intraresidue NOEs involving $H^\alpha-H^N$, $H^\beta-H^N$, $H^\gamma-H^N$, $H^\alpha-H^\beta$, $H^\beta-H^\gamma$, etc., and interresidue NOEs involving $H^\alpha(\text{residue } i)-H^N(\text{residue } i+1)$, $H^\beta(i)-H^N(i+1)$, $H^N(i+1)-H^N(i+1)$, and some $H^\gamma(i)-H^N(i+1)$. Full-matrix analyses and the associated R -factor test with respect to the NOESY data at 200 and 400 ms of mixing produced 220 distance constraints for the MUC1-V3 1.2 60-residue peptide (36). Simulated annealing subject to the distance constraints resulted in a set of stable (low energy) structures in agreement with the NMR data. All of the structures shared the following common features: (i) a tight GPGR type II turn; (ii) a protruding motif (or knob—colored red) consisting of residues 1–11 in each repeat, with the GPGR forming the solvent-accessible tip; (iii) seven residues from the HIV-MN PND in MUC1-V3 that are structurally isomorphous with the same residues in the cyclic HIV-MN V3 (Fig. 2); and (iv) an extended spacer (colored green), consisting of β -strand and polyproline structures, connecting the knobs. We noticed significant conformational flexibility in the MUC1-V3 1.2 60-residue peptide in terms of the end-to-end length (R_e) and in the relative distance and orientations of the knobs. This flexibility neither violates the NMR distance constraints nor disrupts structural features *i–iv* discussed above. Fig. 3 shows three different folded forms of the MUC1-V3 1.2 60-residue

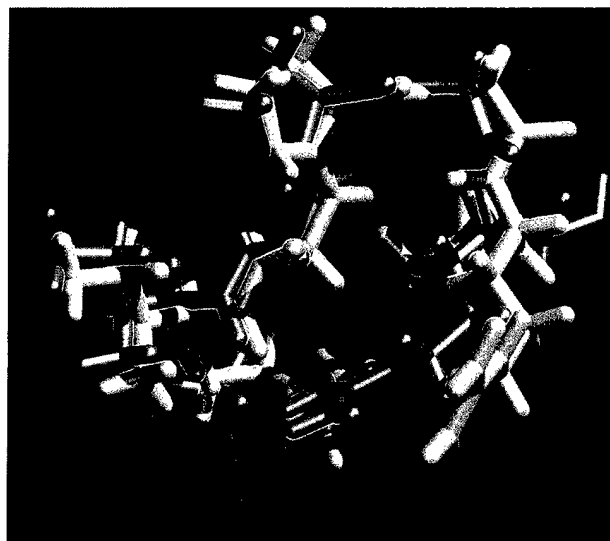


FIG. 2. Structural similarity between the HIV PND isotope HIGPGR in the HIV-MN cyclic V3 (green) and in the central knob of chimeric MUC1-V3 1.2 60-residue peptide. The residues on the N terminus are on the left, while the residues on the C terminus are on the right. Note that P occupies the left corner of the crest.

peptide in agreement with the NMR data. Fig. 3 *Left* shows a structure with $R_e = 80 \text{ \AA}$ and with knobs alternating on two sides of the long axis of the molecule, whereas Fig. 3 *Center* shows a structure with $R_e = 80 \text{ \AA}$ but with the knobs located on the same side of the long axis. A more compactly folded structure with $R_e = 30 \text{ \AA}$ and with the first and the third knobs close together is shown in Fig. 3 *Right*. The conformational variants in Fig. 3 involve only minor variations in backbone torsion angles (i.e., small correlated changes produce large cumulative deviations). Also note that some degrees of flexibility possible for the 60-residue peptide may be sterically prohibited for the 120-residue peptide. For example, the progressive extension of the 60-residue chain at the bottom in Fig. 3 to accommodate the 120-residue peptide may be sterically excluded because of the clash of the third and fourth knobs. Therefore, for steric reasons the 120-residue chain has to necessarily unfurl. This implies, even though the local structures of the 20-amino acid repeats may be identical in the

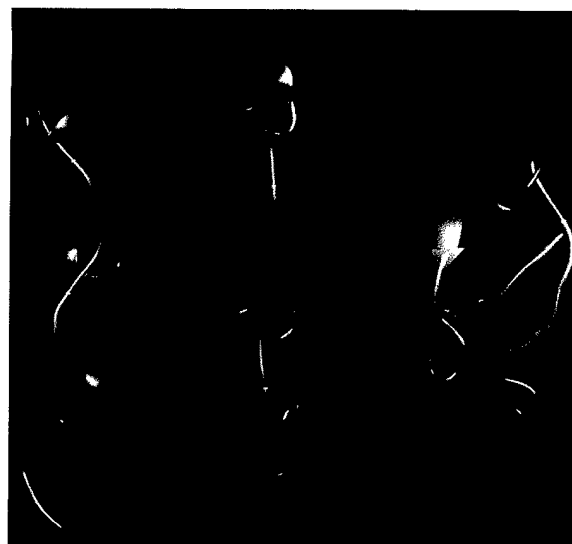


FIG. 3. Flexibility in the arrangement of the tandem repeats in MUC1-V3 1.2 60-residue peptide.

60- and 120-residue peptides, that the global folding and dynamics can be different in these two polypeptides. These differences, in addition to the difference in the number of binding sites, are likely to alter antibody binding properties for these two polypeptides. The results in Figs. 2 and 3 prove that the PND inside the chimeric construct is structurally similar to that in the native V3 loop.

Presentation of Structure. Table 1 lists the ELISA reactivity of various V3 antigens against 39 different polyclonal antisera from HIV-infected patients. Also shown in the table are the PND sequences at the tip of the predominant HIV V3 loops from the corresponding HIV-infected patients. The reactivity of the mucin 105-mer is also shown on a relative scale of 0 to 4 (0 being least reactive). Note that 37 of 39 antisera in Table 1 show no reactivity with the mucin 105-mer; the remaining

two show only weak binding. However, MUC1-V3 120-residue antigens show reactivity comparable to the full-length linear and cyclic V3 loops in the majority of the cases. MUC1-V3 1.3, which has the longest PND segment, is the most reactive among all three MUC1-V3 antigens. The conformational purity of the PND epitope in MUC1-V3 1.3 (and not the length of the sequence) decides the high reactivity. By comparison, the 14-residue linear V3 peptide (with 5 more residues than the PND in MUC1-V3 1.3) is much less reactive. This is because the small linear V3 peptide is less structured than the PND in MUC1-V3 1.3. All antisera specific for Ile-Gly-Pro-Gly-Arg-Ala-($\pm$ Phe) [IGPGRA(F); i.e., with or without phenylalanine in the epitope] show strong reactivity toward MUC1-V3 1.2. These reactivities compare well with those displayed by the HIV-MN cyclic V3 loop against the same antisera.

Table 1. Reactivity of patient polyclonal antisera and mAbs with V3 antigens

Antibody		Reactivity with V3 antigens						Reactivity with mucin 105-mer
		HIV-MN V3 loops			MUC1-V3 chimeric 120-residue			
					1.1	1.2	1.3	
Name	V3 loop PND sequence	Cyclic	Linear	Small linear				
HIV patient serum								
HOND01	IPIGPGRAF	+4	+4	+4	+4	+2	+4	+1
HOND02	IHIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND03	INIGPGRAW	+4	+4	+2	+1	+3	+4	0
HOND04	IHMGPGGAF	+4	+4	0	0	+2	+4	0
HOND05		+4	+4	+4	0	+4	+4	0
HOND06	IYIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND08	IHIGPGSAW	+4	+4	+2	+1	+4	+4	0
HOND09	IHIGPGRAF	+4	+4	+1	+1	0	+4	0
HOND10	IPIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND12	IPLGPGRAF	+4	+4	+2	+3	+2	+4	0
HOND13		+4	+4	+4	+3	+4	+4	0
HOND14	IHIGPGRAF	+4	+4	+4	+1	+4	+4	0
HOND15	IHIGPGRAF	+4	+4	+3	+2	+4	+4	0
HOND16	IHIGPGSAW	+4	+4	+4	0	0	+4	0
HOND17	IHIGPGRAF	+4	+4	+4	+1	0	+4	0
HOND18	VTIGPGKVW	+3	+4	+1	0	+3	+4	0
HOND19	INIGPGRAF	+4	+4	+3	+4	+4	+4	0
HOND20		+4	+4	+1	+4	+4	+4	0
HOND51	VHIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND52	IYIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND53	VRIGPGRAF	+4	+4	+1	0	0	+1	0
HOND54		+4	+4	+1	0	+4	+4	0
HOND55	INIGPGRAF	+4	+4	+3	0	0	+4	0
HOND58	INIGPGRAF	+4	+4	+4	+4	+4	+4	0
HOND59	VHIGPGRAF	+3	+4	+1	0	0	+2	0
HOND60	INIGPGRAF	+4	+4	+4	+2	+4	+4	0
HOND61		+4	+4	+2	0	+4	+4	0
HOND62		+4	+4	+2	+1	+1	+4	0
RW9223	IHIGPGRAF	+4	+4	+2	+4	+4	+4	0
RW9225	VRIGPGQTF	0	0	0	0	0	+3	0
BR9203	IHMGWGRAF	0	0	0	0	0	0	0
BR9221	IHMGWGRAF	+2	0	0	0	0	0	0
BR9225	IRIGPGQAF	+4	+4	0	0	+1	+3	+2
BR9228	IHMGWGRTF	0	0	0	0	0	0	0
TH9201	INIGPGQVF	0	0	0	0	0	0	0
TH9206	ITIGPGQVF	0	0	0	0	0	0	0
TH9209	ITIGPGQVF	0	0	0	0	0	0	0
UG9238	TPIGQGQVL	0	0	0	0	0	0	0
UG9365	TSIGLGQAL	0	0	0	0	0	+4	0
mAb								
268-D	HIGPGR	+2	+2	—	0	+2	+2	0
R/V3-50.1	RIHIG	+4	+2	—	0	0	+4	0

Reactivity: 0, absorbance $A = 1-5 \times$ the average of six normal controls; +1, $A = 5 \times$ the average of six normal controls up to 1.0 A_{405} unit in ELISA; +2, $A = 1-1.5 A_{405}$ units in ELISA; +3, $A = 1.5-2.0 A_{405}$ units in ELISA; +4, $A = >2.0 A_{405}$ units in ELISA.

Two neutralizing mAbs show distinct antigen specificity (Table 1). The human mAb 268-D (12) is specific for the sequence HIGPGR and reacts with MUC1-V3 1.2 and MUC1-V3 1.3 antigens, which contain this sequence, but not with MUC1-V3 1.1, which does not contain this full sequence. The mouse mAb R/V3-50.1, obtained by immunization with a cyclic V3 loop (37), is specific for the sequence Arg-Ile-His-Ile-Gly. This mAb only reacts with MUC1-V3 1.3.

DISCUSSION

MUC1-V3 antigens are unique in the following ways. (i) NMR and antibody-binding data verify that they reproduce the native structure of the PNDs even when they are presented in the context of a totally unrelated protein like MUC1. (ii) Immunogens containing identical PNDs within the MUC1 chimeras effectively allow enhanced presentation of a conserved structural feature of the virus in a fashion not possible with nonchimeric HIV antigens. The true advantage of this approach will be to induce either T-cell-dependent or T-cell-independent antibody responses to the PND, depending on the precise construction of the antigen. (iii) Multiple PNDs, present in these chimeric proteins, may be advantageous in enhancing the immune response by significantly increasing the affinity of antibody binding. The importance of multiple PNDs being present in the same antigen becomes clear by analyzing the relative binding of MUC1-V3 1.2 60- and 120-residue peptides to different antisera. The data show that the 120-residue peptide is a better ligand than the 60-residue peptide for the majority of the antisera we tested (data not shown). This is probably because the higher number of PND knobs in the 120-residue peptides are correctly disposed along the long axis of the molecule to facilitate the binding of bivalent antibodies. (iv) Alternatively, the nature of the MUC1-V3 structure (Fig. 3) suggests that if two or more different PNDs are grafted alternately along the chain, there is enough flexibility in the spacers so that two or more antibodies specific for two different PNDs can both bind bivalently, interdigitating along the molecule. Finally, there is no reason why more than two PNDs cannot be introduced in the molecule. This may be critical in designing vaccines for a highly mutating pathogen like HIV.

We acknowledge Dr. Byron Goldstein for his help and encouragement. This work was supported by National Institutes of Health Grant R01 AI32891-01A2 and Los Alamos National Laboratory Grant XL 77; and mAbs were supplied by the National Institutes of Health AIDS Research and Reference Reagent program.

- Dalgleish, A. G., Beverly, P. C. L., Clapman, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984) *Nature (London)* **312**, 763-766.
- Robey, E. & Axel, R. (1990) *Cell* **60**, 697-700.
- Rusche, J. R., Javaherian, K., McDanal, C., Petro, J., Lynn, D. L., Grimailla, R., Langlois, A., Gallo, R. C., Arthur, L. O., Fischinger, P. J., Bolognesi, D. P., Putney, S. D. & Mathews, T. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3198-3202.
- Freed, E. O., Myers, D. J. & Risser, R. (1991) *J. Virol.* **65**, 190-194.
- Robert-Guroff, M., Brown, M. & Gallo, R. C. (1985) *Nature (London)* **316**, 72-74.
- Putney, S. A., Mathews, T. J., Robey, W. G., Lynn, D. L., Robert-Guroff, M., Mueller, W. T., Langlois, A. J., Ghayeb, J., Jr., S. R. P., Weinhold, K. J., Fischinger, P. J., Wong-Staal, F., Gallo, R. C. & Bolognesi, D. P. (1986) *Science* **234**, 1392-1395.
- Chamat, S., Nara, P., Berquist, L., Whalley, A., Morrow, W. J. W., Kohler, H. & Kang, C. Y. (1992) *J. Immunol.* **149**, 649-654.
- LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Mathews, T. J., Emini, E. A. & Putney, S. D. (1990) *Science* **249**, 932-935.
- Chao, B. H., Costopoulos, D. S., Curiel, T., Bertonis, J. M., Chisholm, P., Williams, C., Schooley, R. T., Rosa, J. J., Fisher, R. A. & Marganore, J. M. (1989) *J. Biol. Chem.* **264**, 5812-5817.
- Myers, G., Korber, B., Berzofsky, J. A., Smith, R. F. & Pavlakis, G. N. (1992) *Human Retroviruses and AIDS* (Los Alamos Natl. Lab., Los Alamos, NM).
- Mathews, T. J., Langlois, A. J., Robey, W. G., Chang, N. T., Gallo, R. C., Fischinger, P. J. & Bolognesi, D. P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9709-9713.
- Gorny, M. K., Xu, J. Y., Gianakakos, V., Karwowska, S., Williams, C., Sheppard, H. W., Hanson, C. V. & Zolla-Pazner, S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3238-3242.
- Takahashi, H., Nakagawa, Y., Pendleton, C. D., Houghten, R. A., Yokomuro, K., Germain, R. N. & Berzofsky, J. A. (1992) *Science* **255**, 333-336.
- Hwang, S. S., Boyle, T. J., Lyster, H. K. & Cullen, B. R. (1991) *Science* **253**, 71-74.
- Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. (1992) *J. Virol.* **66**, 6547-6554.
- Clements, G. J., Price-Jones, M. J., Stephens, P. E., Sutton, C., Schulz, T. F., Clapman, P. R., J. A. M., McClure, M. O., Thomson, S., Marsh, M., Kay, J., Weiss, R. A. & Moore, J. P. (1991) *AIDS Res. Hum. Retroviruses* **7**, 3-15.
- Gupta, G. & Meyers, G. (1990) *Computer Analysis of HIV Epitope Sequences 1-99-105* (Pasteur Vaccins, Paris).
- Veronese, F. D. M., Jr., M. S. R., Gupta, G., Robert-Guroff, M., Boyer-Thompson, C., Gallo, R. C. & Lusso, P. (1993) *J. Biol. Chem.* **268**, 25894-25901.
- Gupta, G., Anantharamaiah, G. M., Scott, D. R., Eldridge, J. H. & Myers, G. (1993) *J. Biomol. Struct. Dyn.* **11**, 345-366.
- Gupta, G. & Myers, G. (1994) in *Peptides: Design, Synthesis, and Biological Activity*, eds. Basava, C. & Anantharamaiah, G. M. (Birkhauser, Boston), pp. 259-277.
- Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6325-6329.
- Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T. & Wilson, I. A. (1994) *Science* **264**, 82-85.
- Gendler, S., Taylor-Papadimitriou, J., Duhlig, T., Rothbard, J. & Burchell, J. A. (1988) *J. Biol. Chem.* **263**, 12820-12823.
- Jerome, K. R., Barnd, D. L., Bendt, K. M., Boyer, C. M., Taylor-Papadimitriou, J., McKenzie, I. F. C., Bast, R. C., Jr., & Finn, O. J. (1991) *Cancer Res.* **51**, 2908-2916.
- Fontenot, J. D., Tjandra, N., Bu, D., Ho, C., Montelaro, R. C. & Finn, O. J. (1993) *Cancer Res.* **53**, 5386-5394.
- Fontenot, J. D., Finn, O. J., Dales, N., Andrews, P. C. & Montelaro, R. C. (1993) *Pept. Res.* **6**, 330-336.
- Pau, C. P., Lee-Thomas, S., Auwanit, W., George, J. R., Ou, C. Y., Parekh, B. S., Granade, T. C., Holloman, D. L., Phillips, S., Schochetman, G., Young, N. L., Takebe, Y., Gayle, H. D. & Weniger, B. G. (1993) *AIDS* **7**, 337-340.
- Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **63**, 207-213.
- Jeener, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546-4553.
- Nemethy, G., Pottle, M. S. & Scheraga, H. A. (1983) *J. Phys. Chem.* **87**, 1883-1887.
- Kirkpatrick, S., Jr., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science* **220**, 671-680.
- Taylor-Papadimitriou, J. (1991) *Int. J. Cancer* **49**, 1-5.
- Devine, P. L. & McKenzie, I. F. C. (1992) *BioEssays* **14**, 619-625.
- Rose, D. R., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1-109.
- Lorimer, R. D., Moody, M. A., Haynes, B. F. & Spicer, L. D. (1994) *Biochemistry* **33**, 2055-2062.
- Gupta, G., Sarma, M. H. & Sarma, R. H. (1988) *Biochemistry* **27**, 3423-3431.
- D'Souza, M. P., Durda, P., Hanson, C. V. & Milman, G. (1991) *AIDS* **5**, 1061-1070.
- Catasti, P., Fontenot, J. D., Bradbury, E. M. & Gupta, G. (1995) *J. Biol. Chem.*, in press.

Structure and Polymorphism of HIV-1 Third Variable Loops*

(Received for publication, May 8, 1995, and in revised form, December 27, 1995)

Paolo Catasti^{‡§}, E. Morton Bradbury^{§¶}, and Goutam Gupta^{‡¶}*From the [‡]Theoretical Biology and Biophysics Group, T-10, M/S K710 and the [§]Life Science Division, M/S M881, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 and the Department of Biological Chemistry, School of Medicine, University of California at Davis, Davis, California 95616*

The third variable (V3) loop of HIV-1 surface glycoprotein, gp120, has been the target of neutralizing antibodies. However, sequence variation inside the V3 loop diminishes its effectiveness as a potential vaccine against HIV-1. The elusive nature of the V3 loop structure prompted us to carry out a systematic study on different isolates in an attempt to identify a common structural motif in the V3 loop regardless of the amino acid sequence variability. We have previously determined the structural features of two V3 loops: V3 Thailand and V3 MN. In this paper, we present the structure of two other variants: V3 Haiti and V3 RF. Our results show that similar secondary structures are observed in all the four V3 loops: a GPG(R/K/Q) crest in the center of the neutralizing domain, two extended regions flanking the central crest, and a helical region in the C-terminal domain. For the Haitian V3 loop, we also show how the conserved structural features are masked through a conformational switch encoded in the amino acid sequences on the C-terminal side of the GPGK crest.

A neutralizing determinant (1–8) located inside the V3¹ loop of the envelope glycoprotein, gp120, has been the target for protective immunity against the human immunodeficiency virus, type 1 (HIV-1). However, the amino acid sequence variation within the V3 loop has eluded the effectiveness of V3-based vaccine design (4, 5). Antibodies against the V3 loop generally exhibit type-specific neutralization profiles (6, 7), although a subset of anti-V3 antibodies specific for the less variable elements of the V3 loop show a broader range of neutralizing activity (7, 8). To better understand the effect of sequence variation on the structure and antigenicity of the HIV-V3 loop, we developed a method combining molecular modeling (9, 28) and two-dimensional NMR (10, 11) to analyze the global structure of the entire cyclic V3 loop and the local structure at the neutralizing determinant. We attempted to answer two specific questions: (i) Are there conserved secondary structural elements inside the V3 loop in spite of sequence variation? and (ii) Can the sequence variation inside the V3 loop mask this conserved secondary structure? Recently, we have shown (12) that in spite of the observed sequence variation, a conserved sec-

ondary structure is located inside the V3 loop. This structure consists of a solvent-accessible protruding motif (or a knob) spanning 8–10 residues with a central GPG(Q/R/K) type II turn at the crest of the knob (10–12). In this article, we demonstrate how amino acid sequence variation flanking the GPG crest can camouflage an otherwise conformationally pure epitope. For this purpose, we performed two-dimensional NMR and molecular modeling studies on two cyclic V3 loops from the Haitian and RF isolates (13):

	1	36	
V3-Haiti	<u>C</u> CTRPNDNTRKSIPMPGPKAFYATGDIIGNIRQAHC		net charge +3
V3-RF	CCTRPNNNTRKSITKGPGRVIYATGQIIGDIRKAHC		net charge +5

(Cysteines at position 2 and 36 are S–S bridged; the first cysteine that is underlined in the sequence has a protective group on S; site-specific differences in sequence are marked in bold).

MATERIALS AND METHODS

Synthesis and Purification—The cyclic Haitian V3 loop was synthesized and purified by Dr. Anita Hong (Anaspec, CA). The cyclic RF V3 loop was synthesized and purified by Dr. G. M. Anantharamaiah, University of Alabama at Birmingham. Both Drs. Hong and Anantharamaiah provided their services under a contract with the AIDS Division of the National Institutes of Health.

NMR Spectroscopy—All NMR experiments on the RF V3 loop were carried out on the 600 MHz Bruker spectrometer at the University of Alabama at Birmingham, whereas the data on the Haitian V3 loop were collected on a 500 MHz Bruker AMX spectrometer at Chemistry, Science, and Technology-4, Los Alamos National Laboratory. NMR spectra were collected at 10 °C with 3 mM peptide concentration (pH 5.5) for the Haitian V3 loop and at 5 mM peptide concentration (pH 5.5) for the RF V3 loop. All two-dimensional data were acquired in the phase-sensitive mode with the presaturation of the HDO signal during the relaxation delay. DQF-COSY data (14) were collected with the following acquisition parameters: $t_2 = 4096$, $t_1 = 1024$, relaxation delay = 1.5 s, number of transients = 32 for the Haitian V3 loop and number of transients = 48 for the RF V3 loop. Total correlated spectroscopy data (14) were collected with the following parameters: $t_2 = 2048$, $t_1 = 1024$, relaxation delay = 1.5 s, number of transients = 32/48, isotropic mixing = 70 ms. NOESY (14) data were collected with similar acquisition parameters and with 75 and 250 ms of mixing time (τ_m) for the Haitian V3 loop and 200 and 400 ms for the RF V3 loop. The sequential assignments² were performed by combining the total correlated spectroscopy, DQF-COSY, and NOESY data, processed on a SGI Indigo instrument using the Felix software (BioSym Inc.).

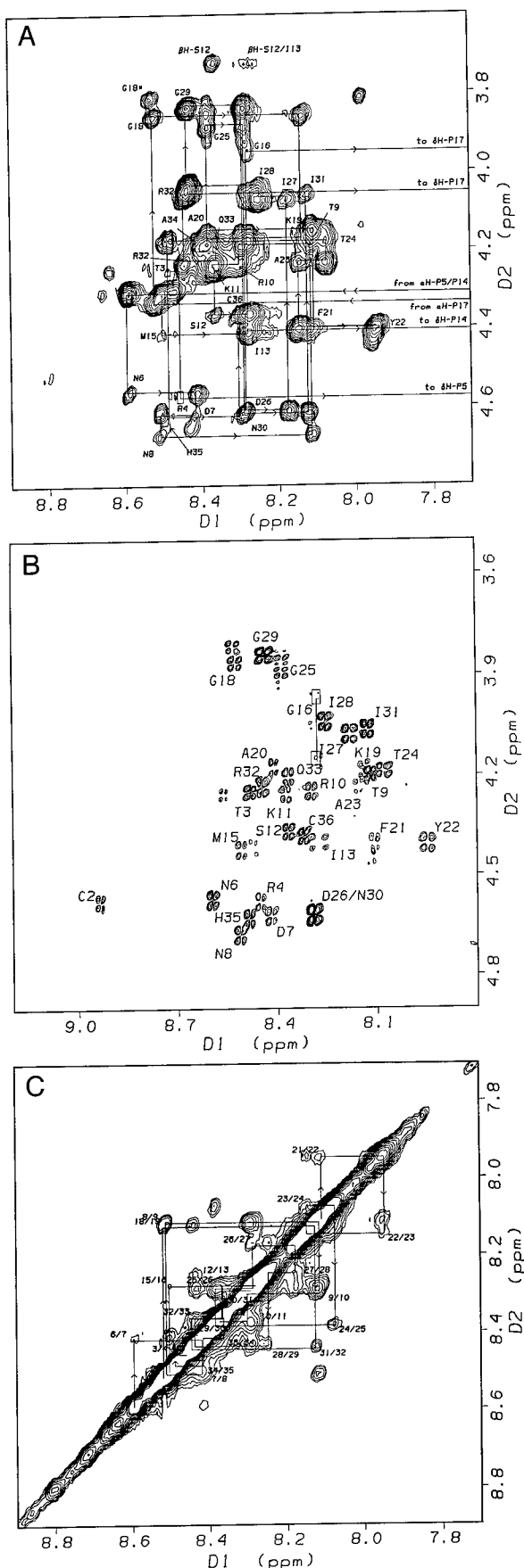
Structure Derivation—Two-dimensional NMR data of the Haitian V3 loop in water and in TFE/water mixture and for the RF V3 loop in water were analyzed with the aid of full-matrix NOESY simulations, associated R-factor test, and energy calculations (10, 11). This produced a set of distance constraints (9–12, 28). The energy term (EDIST defining the

* This work is supported by National Institutes of Health Grant R01 AI32891-01A2 and U. S. Army Grant MIPR 92MM2581. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 505-665-2587; Fax: 505-665-3493; E-mail: gxxg@temin.lanl.gov.

¹ The abbreviations used are: V3, third variable; HIV-1, human immunodeficiency virus, type 1; DQF-COSY, two-dimensional double-quantum filtered correlated spectroscopy; rms, root mean square; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TFE, trifluoroethanol.

² Authors on request will provide additional data containing proton NMR assignments, $J_{\text{HN-H}^\alpha}$ coupling constants, NOE volumes, NOE-derived distances, cartesian coordinates, torsional angle data of the three structures: V3 Haiti in water and in the mixed water/TFE solvent and V3 RF in water. The surface accessibility data of the three structures in Figs. 6–8 are also available from the authors on request.



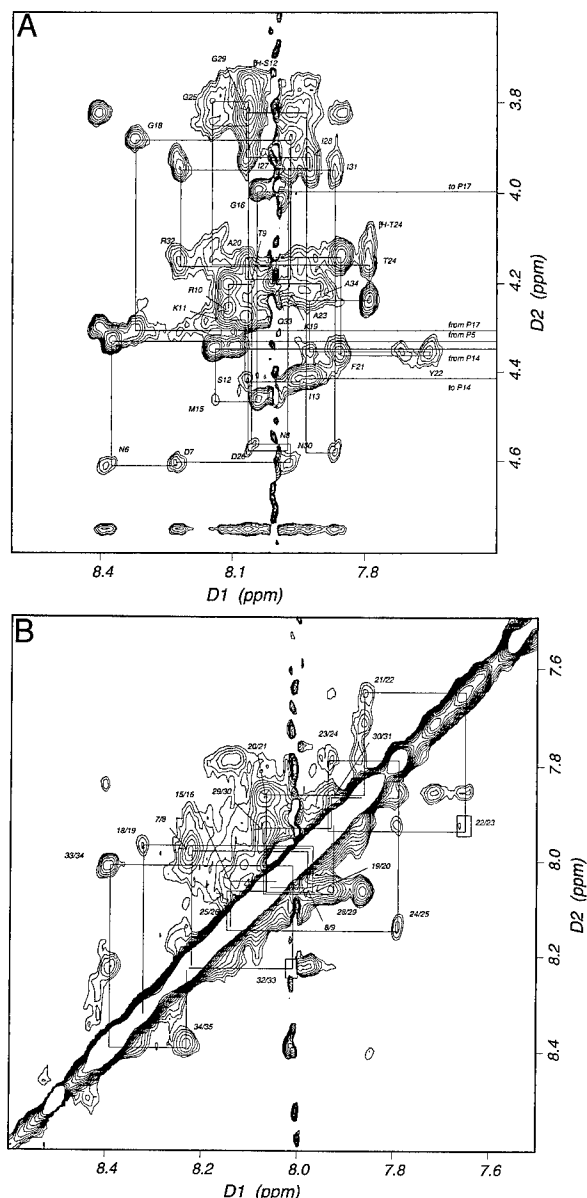


FIG. 2. NOESY cross-sections of the cyclic Haitian V3 loop in water/TFE (7:3) mixture (peptide concentration = 3.0 mM, pH 5.5, temperature = 10 °C). The HDO signal was presaturated for 1.2 s during the relaxation delay. Acquisition parameters: data matrix ($t_2 = 2048$ data points, $t_1 = 1024$ data points), relaxation delay = 1.2 s, number of transients = 32, temperature = 10 °C. **A**, the fingerprint HN-H α region. **B**, the HN-HN region. In the fingerprint region note the double population of the residue Gly¹⁸. One of the populations has an additional medium range NOE between H α -G18 and HN-F21, indicating a conformational equilibrium between two forms, one extended and one bent at the fragment Gly¹⁸-Lys¹⁹-Ala²⁰-Phe²¹. Interestingly such a double population for Gly¹⁸ is also observed in water, although the medium range NOE with F21 absent in the aqueous environment (see Fig. 1). Note that uniform upfield shift of the H α resonances in the C-terminal region of the peptide that is a strong indicator of the helical conformation.

the residues 27–34 in the C terminus of the Haitian V3 loop are high field shifted in the mixed solvent as also observed in the case of the MN V3 loop (11).

The residues in the neutralizing epitope, Ile¹³-Pro¹⁴-Met¹⁵-Gly¹⁶-Pro¹⁷-Gly¹⁸-Lys¹⁹-Ala²⁰-Phe²¹-Tyr²², are unequivocally assigned in water and in the solvent. Two conformational states of Gly¹⁸ (*i.e.* Gly¹⁸ and Gly^{18*}) are clearly evident. Gly¹⁸ not only shows two populations in terms of the chemical shift values of (H α ,HN) but also in terms of the NOESY connectivities. As shown in Fig. 2A, a second population of Gly^{18*} is

observed in mixed populations; Gly¹⁸ shows a H α (G18)-HN(F21) cross-peak (Fig. 3). The chemical shift of other residues in the second population are indistinguishable from the first. Although this conformational variant is also present in the aqueous solvent, the absence of the H α (G18)-HN(F21) cross-peak suggests that such an interaction perhaps is not stabilized in a polar environment.

NMR Experiment on the RF V3 Loop—We have chosen the RF V3 loop to examine the effect of sequence variation on the overall folding of the V3 loop and the local structure at the GPG crest. The RF V3 loop is different from the Haitian V3 loop at eight positions and is more positively charged. Fig. 4A shows the NOESY fingerprint HN-H α region of the RF V3 loop in water at 10 °C for 400 ms of mixing. Note the presence of intra-residue HN-H α and inter-residue HN(i+1)-H α (i) cross-peaks except for the residues Cys¹, Cys², Thr³, Arg⁴, Pro⁵, and Asn⁶, possibly due to inherent flexibility in the N-terminal region. Fig. 4B shows the NOESY HN-HN region of the RF V3 loop in water at 10 °C for 300 ms of mixing. Although a set of sequential HN-HN cross-peaks are observed in the C-terminal stretch (residues 29–34) of the V3 loop, no medium range NOEs indicative of helical fragments are seen. Fig. 5 summarizes the NOE data of the RF V3 loop in the aqueous environment from the analyses of the data at 200 and 400 ms of mixing. Only intra-residue and sequential NOEs are observed. The $J_{\text{HN-H}\alpha}$ coupling constants from the DQF-COSY spectrum (Fig. 4C) provide the ϕ values, which are converted into HN-H α intra-residue distances.² The lack of solubility of the highly polar RF V3 loop prevented us from carrying out NMR experiments in the mixed water/TFE (7:3) solvent.

Structures of the Haitian V3 Loop—We have previously reported structural studies on the Thailand and MN V3 loops (10, 11). The NMR data for those two sequences revealed that the structure of the V3 loop contained a few reasonably well defined secondary structural elements, *i.e.* a GPGR(Q) turn and a nascent C-terminal helix. However, the V3 loops are considerably flexible within the constraints of these secondary structural elements and the Cys²-Cys³⁶ disulfide bridge. Therefore, we developed and applied a method to explore the extent of conformational flexibility of the V3 loop that is consistent with the NMR data (10, 11). Briefly, we employ the following steps. (i) We analyze the NMR data to assign secondary structural states, *i.e.* ranges of ϕ and ψ values, to the residues of the V3 loop. The NMR data for the aqueous form include the NOEs in Fig. 3 and the DQF-COSY data (Fig. 1C); the $J_{\text{HN-H}\alpha}$ coupling constants from the DQF-COSY spectrum provide the ϕ values, which are converted into HN-H α intra-residue distances.² Line broadening in the mixed solvent prevented us from obtaining a high quality DQF-COSY spectrum in the water/TFE (7:3) mixture. (ii) We obtain a set of starting (energy minimized) structures of the V3 loop subject to the ϕ and ψ values and disulfide bridge constraints. (iii) We then use these starting structures for Monte Carlo simulated annealing and energy minimizations for sampling the conformational space. (iv) We finally select a set of low energy structures (50 for the Haitian V3 loop) and analyze the conformational parameters to examine the nature of the flexibility.

Fig. 6 shows the ribbon models of the Haitian V3 loop in water (*left*) and water/TFE (7:3) mixture (*right*). The following color coding was used in these ribbon diagrams: *gray* for the N-terminal protruding loop at position T3-R10, *green* for the N-terminal extended β -strand flanking the GPG crest, *magenta* for the central β -turn at position G16-P17-G18-K19, *yellow* for the C-terminal extended β -strand flanking the GPG crest, and *blue* for the C-terminal segment D26-H35, which can form an α -helix. In water the C-terminal segment consists of

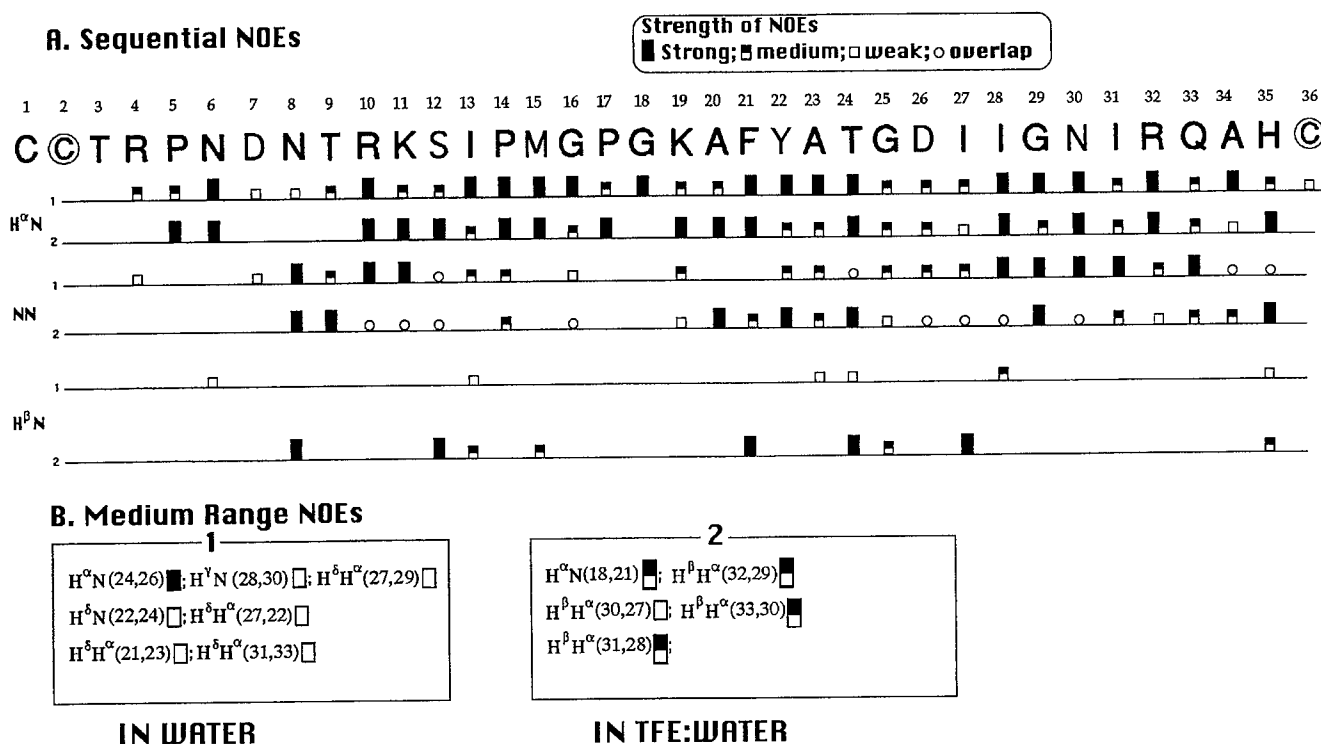


FIG. 3. Summary of the NMR data for the HIV Haitian V3 loop in water (under panel 1) and in water/TFE (7:3) mixed solvent (under panel 2). In addition to sequential H^{α} -HN, H^{β} -HN, and HN-HN NOEs, six sequential H^{γ} -HN NOEs were obtained in both the solvents. The sequential H^{α} -HN connectivity for Pro is missing, but the sequential H^{α} - H^{δ} NOEs provide the link. Note the solvent-induced change in the sequential NOE pattern in the C-terminal segment; in the mixed solvent, there is an enhancement of the sequential HN-HN NOEs relative to the corresponding H^{α} -HN NOEs, indicative of an induction of a helix.² In a previous work (24–25), the NMR assignments of the V3 loops of MN and IIB isolates in aqueous solution were obtained. The authors also performed CD studies on TFE mixed solutions but did not succeed in assigning the NMR spectra in such mixed solvents. In another NMR work (26), the authors studied two 24 linear peptides containing the neutralizing determinant of the IIB isolate of HIV-1. They reported (26) the presence of a transient turn at the GPGR crest, and the ability of mixed TFE solutions to induce helical conformation in the C-terminal domain of the peptides, whereas in water only a “nascent” helix formed by a stretch if interconnected turns was observed.

consecutive β -turns centered around Ile²⁸-Gly²⁹ and Ile³¹-Arg³², whereas in the water/TFE (7:3) mixture it further folds into a well defined α -helix as evidenced by the presence of sequential HN-HN NOEs (Fig. 2B) and medium range $H^{\beta}(i+3)$ - $H^{\alpha}(i)$ NOEs (Fig. 3B, panel 2). Due to the intrinsic conformational flexibility of the V3 loop, side chains are quite mobile, and they do not sample a single rotamer conformation. In these representations only average positions for the side chains are shown. However, in the α -helical region for the mixed solvent structure, the side chains are organized in a cylindrical array as experimentally observed by the presence of a network of $d^{\alpha\beta}(i,j+3)$ and $d^{\beta N}$ sequential connectivities (Fig. 3B). Nonetheless, in both the structures (Fig. 6) the neutralizing epitope containing the central GPGK sequence forms a protruding loop even though the local structure and presentation of the loop in the two cases are noticeably different. The aqueous structure of the Haitian V3 loop in Fig. 6 is the average of 50 sampled configurations that exhibit rms deviations below 1.5 Å with respect to the backbone atoms. Out of 50 sampled structures of the Haitian V3 loop in the TFE/water mixture, a small subset of six structures shows a large (>2.6 Å) rms deviation of the backbone atoms from the rest of the structures. The remaining 44 structures are within 1.6 Å rms deviations of the backbone atoms. The average structure in Fig. 6 is taken over these 44 structures.

Fig. 7 shows two conformations representing the aqueous environment (left) and the mixed solvent forms (right). The central region of the Haitian V3 loop containing the neutralizing determinant residues Ile¹³-Pro¹⁴-Met¹⁵-Gly¹⁶-Pro¹⁷-Gly¹⁸-Lys¹⁹-Ala²⁰-Phe²¹-Tyr²² is shown. The following color coding

was used in these skeleton models: red for the central Gly¹⁶-Pro¹⁷-Gly¹⁸-Lys¹⁹ crest, green for the N-terminal Ile¹³-Pro¹⁴-Met¹⁵ residues in extended conformation, and yellow for the C-terminal Ala²⁰-Phe²¹-Tyr²² residues, which show a solvent-induced effect. In the water structure the C-terminal fragment is in an extended conformation (open state). In the mixed solvent, two types of chain folding are observed: one folded form is similar to that of the MN-V3 loop (11), whereas in the other, the GPG crest forms the typical type II β -turn followed by a type III β -turn involving residues Gly¹⁸-Arg¹⁹-Ala²⁰-Phe²¹, as evidenced by the presence of a medium range NOE between H^{α} -G18 and HN-F21 (closed state). Such an S-shaped conformation has been previously reported for a peptide containing the V3 neutralizing determinant complexed to an antibody (18), and it will be referred as “arched” conformation for the rest of the paper. Our NMR data (Figs. 1A and 3) clearly indicated that these two states are simultaneously present in mixed solvent, whereas only the open state exists in aqueous solutions. We have not shown the open state structure of the Haitian V3 loop in the mixed solvent because it closely resembles the already published structure of the MN-V3 loop (11).

Structure of the RF V3 Loop—Fig. 8 shows the ribbon diagram for the structure of the RF V3 loop in water. Here, again the conformational analysis is done for 50 sampled low energy structures. All the sampled structures showed rms deviations of 0.22 ± 0.02 Å with respect to 79 independent distance constraints. The same color coding described in Fig. 6 was used here. The average structural features of the V3-RF in water resemble those observed for the Haitian and the MN V3 loops (11); however, the absence of any nonsequential NOE suggests

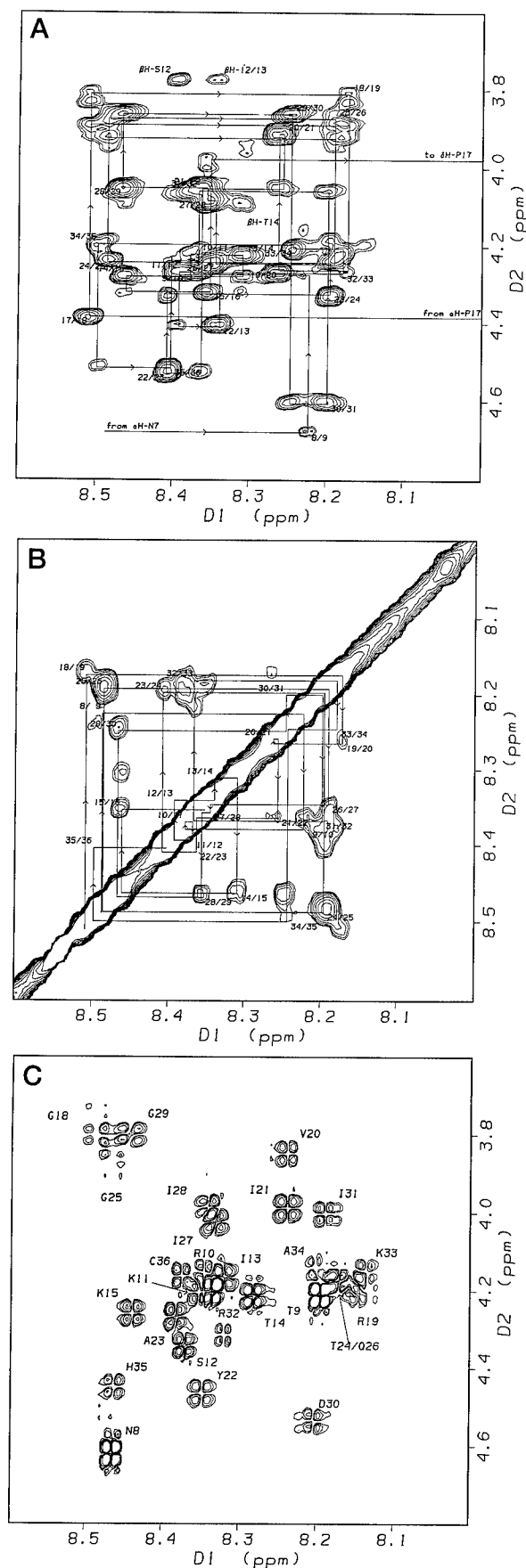


FIG. 4. NOESY (mixing time = 400 ms) and DQF-COSY cross-sections of the cyclic RF V3 loop in water (peptide concentration = 5.0 mM, pH 5.5). A, the fingerprint HN-H^a region. B, the HN-HN region. C, DQF-COSY cross-sections of the cyclic RF V3 loop. For NOESY experiments, the acquisition parameters were as follows: t2

that the RF V3 loop structure is considerably more flexible than the Haiti-V3 and MN-V3 loops. Out of 50 sampled structures, a small subset of eight structures shows a large (>2.7 Å) rms deviations of the backbone atoms from the rest of the structures. The remaining 42 structures are within 1.7 Å rms deviations of the backbone atoms. The average structure in Fig. 8 is taken over these 42 structures.

DISCUSSION

Previous NMR studies on the Thailand and MN V3 loops (10, 11) and the current work on the Haiti and RF V3 loops (Figs. 6–8) can be summarized as follows: (i) A GPG-turn at the crest of the V3 loop is present in all the four sequences. (ii) Stretches of β -strand adjacent to the GPG-turn on the N- and C-terminal sides are common to all the four sequences. (iii) The residues in the C-terminal segment form a few turns in water and a helix in the less polar mixed solvent. (iv) In spite of the constraints of secondary structures (i)–(iii) and the disulfide bridge, the V3 loop exhibits conformational flexibility as evidenced by the absence of long range NOESY interactions commonly observed in well folded globular proteins (14).

However, a “protruding knob” formed by the central GPG turn and the β -strands on either side emerges as the secondary structural feature conserved among diverse V3 loop sequences. The single crystal structure of the HIV-1 neutralizing antibody (monoclonal antibody 50.1) complexed to 16-residue-long linear MN-V3 fragment shows the hint of such a protruding knob, although the segment on the C-terminal side of the GPGR type II turn remains disordered (18). The crystallographic observation suggests that the protruding knob of the V3 loop that includes the neutralizing epitope might well be specifically recognized by the antibody. However, the conserved protruding knob of the V3 loop need not always be presented in its conformationally pure form because HIV may find a way to mask this conserved secondary structural element. In this work we report one such mechanism of masking as revealed by the closed state in Fig. 7. In this form of the Haitian V3 loop, the NMR data indicate an arching of the residues on the C-terminal side of the GPGK turn. This is a departure from the protruding knob motif that contains the central GPG turn and two β -strands on either side. Such an arched conformation of the neutralizing epitope has also been observed in an antibody (monoclonal antibody 59.1) complexed with a linear V3 fragment (18).

When combined with the single crystal data (17, 18), our NMR data (Refs. 10 and 11 and this work) indicate that the closed or arched conformation of the neutralizing epitope of the V3 loop is possible and can be recognized by the antibody. In addition, our data also indicate that an equilibrium between the closed and open state (Fig. 7) is possible for the same V3 loop sequence. The arching around Ala²⁰-Phe²¹ tends to mask Lys¹⁹ and Ala²⁰ (Fig. 7). The closed form of the V3 loop may camouflage some essential elements of the neutralizing epitope from the immune system. For instance, this masking will interfere with the binding of antibodies (8, 19) that recognize the PGRAF epitope.

Most importantly such a local masking of Ala²⁰ and Phe²¹ should affect the proteolysis of the (Arg/Gln/Lys)¹⁹-Ala²⁰ peptide bond by thrombin and tryptase (20, 21); the second enzyme

= 2048 data points, t1 = 1024 data points, relaxation delay = 1.5 s, number of transients = 32. Same acquisition parameters were used for the DQF-COSY experiment except for t2, which was increased to 4 K. Sequence-specific assignments were obtained starting from Val²⁰ (only Val in the sequence) and moving backward and forward along the connectivity route until completion of the assignments. Assignments in the fragment Cys¹-Asn⁷ were not possible, presumably due to the intrinsic flexibility in the region Asn⁶-Asn⁷-Asn⁸.

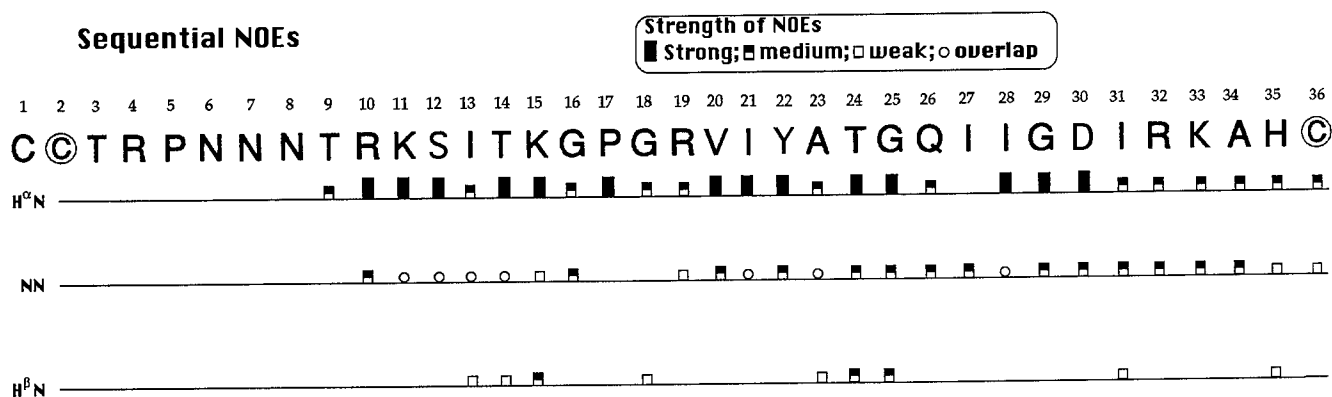


FIG. 5. Summary of the NMR data for the HIV RF V3 loop in water. Sequential H^α-HN, H^β-HN, and HN-HN NOEs. The sequential H^α-HN connectivity for Pro is missing, but the sequential H^α-H^β NOEs provide the NOE link.² Lorimier *et al.* (26) studied a 40-residue-long peptide containing a T-helper epitope 16 residue long in the N-terminal region and a 24-residue-long segment derived from the V3 loop of the HIV-RF isolate studied in this work. In the V3 loop region the authors observed the GPGR turn, whether the rest of the C-terminal fragment was mostly disordered. However, we would like to stress that in the study of the linear peptides (24, 25), the authors did not consider the importance of the disulfide bridge locking the V3 fragment in a closed loop. We have previously shown that the cyclic V3 loops are better ligands for V3-specific antibodies (11).

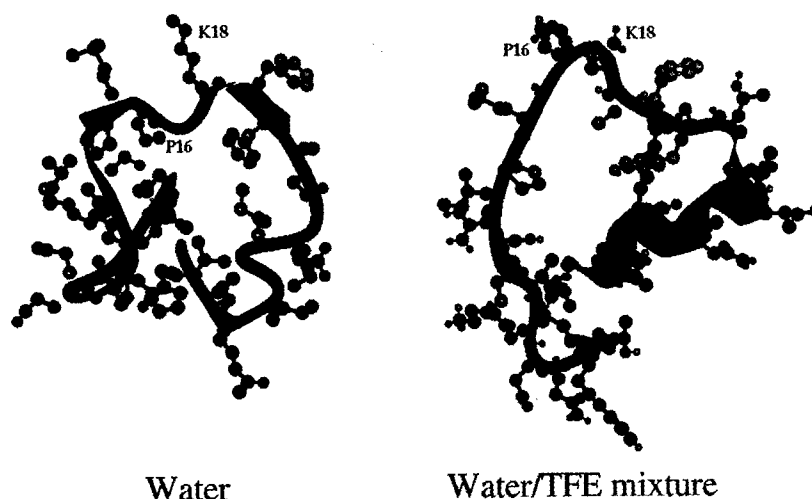
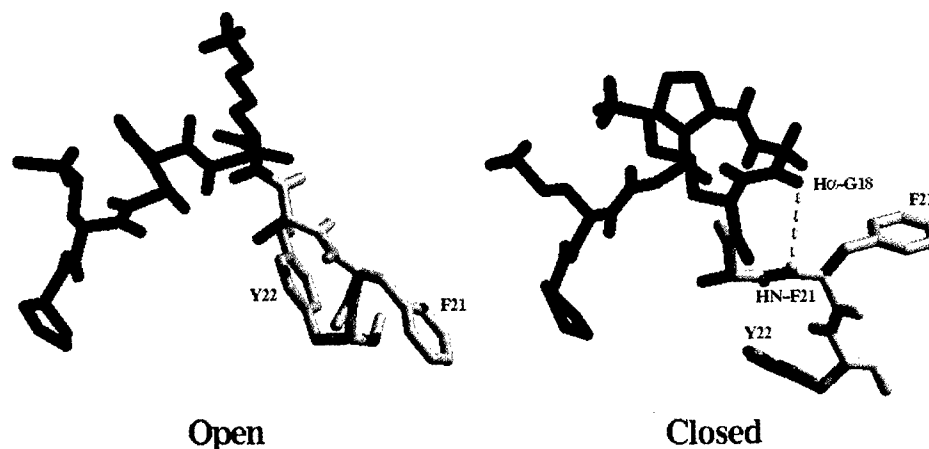


FIG. 6. The ribbon diagrams describe representative folding patterns for the structures of the Haitian V3 loop in water (*left*) and in mixed water/TFE solvent (*right*). The following color coding was used in these ribbon diagrams: gray for the N-terminal protruding loop, green for the N-terminal extended β -strand flanking the GPG crest, magenta for the central β -turn at the GPG crest, yellow for the C-terminal extended β -strand flanking the GPG crest, and blue for the C-terminal segment, which can form an α -helix. In each case, the average is done over 50 sampled low energy structures. Ribbon models in the two cases correspond to the average structure. All the sampled structures of the Haitian V3 loop in water showed rms deviations of 0.24 ± 0.02 Å with respect to 95 independent distance constraints. All the sampled structures of the Haitian V3 loop in mixed water/TFE solvent showed rms deviations of 0.27 ± 0.02 Å with respect to 123 independent distance constraints. The structures of the Haitian V3 loop in water show a greater degree of flexibility than those in the mixed water/TFE solvent; this is due to the formation of the C-terminal helix in the mixed solvent.

FIG. 7. Two conformations representing the aqueous environment (*left*) and mixed solvent forms (*right*). The central region of the Haitian V3 loop containing the neutralizing determinant residues Ile¹³-Pro¹⁴-Met¹⁵-Gly¹⁶-Pro¹⁷-Gly¹⁸-Lys¹⁹-Ala²⁰-Phe²¹-Tyr²² are shown. The following color coding was used in these skeleton models: green for the N-terminal Ile¹³-Pro¹⁴-Met¹⁵ residues in extended conformation, red for the central Gly¹⁶-Pro¹⁷-Gly¹⁸-Lys¹⁹ crest, and yellow for the C-terminal Ala²⁰-Phe²¹-Tyr²² residues, which show solvent induced arching effect. Solvent-accessible areas were calculated using the Molecular Surface Package due to Connolly (27) with a probe radius of 1.5 Å. In the fragment Lys¹⁹-Phe²², the aqueous structure has a lower surface accessibility than the structure in the mixed TFE/water solvent.



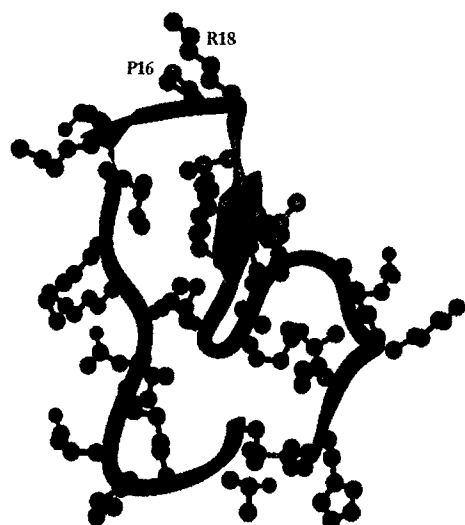


FIG. 8. The ribbon diagrams describe representative folding pattern of the RF V3 loop in water. The same color coding described in the legend to Fig. 6 was used here. The structural features of the V3-RF in water resemble those observed for the Haitian and the MN V3 loops in water (11).

lies on the T-cell surface. When gp120 is used as a substrate these two enzymes show exceptional specificity for cleavage of the (Arg/Gln/Lys)¹⁹-Ala²⁰ peptide bond inside the V3 loop. Most striking is the observation that the V3 loops of T-cell tropic virus strains are 1,000 times more susceptible to cleavage by these two enzymes than the V3 loops of macrophage tropic strains (21). The T-cell tropic V3 loops are more positively charged than the macrophage tropic V3 loops (21–23). Our studies reveal that the open state of the neutralizing epitope of the V3 loop is exclusively preferred for MN and RF V3 loops with net charges of more than +5, whereas the closed state of the neutralizing epitope begins to appear for the Haitian V3 loop with a net charge of +3. Therefore, the proteolysis data (21) are consistent with our structural conclusions.

Acknowledgments—We thank Dr. James Bradac of the National Institutes of Health vaccine branch for help and suggestions, Dr. Cliff Unkefer of Chemistry, Science, and Technology-4 for giving us access to the 500 MHz Bruker-AMX instrument, and the NMR Facility at the University of Alabama, Birmingham for giving us access to the 600 MHz Bruker-AMX instrument.

REFERENCES

1. Gyer, H., Holschbach, C., Hunsmann, G., and Scheider, J. (1988) *J. Biol. Chem.* **263**, 11760–11767

2. Putney, S. D., Mathews, T. J., Robey, W. G., Lynn, D. L., Robert-Guroff, M., Mueller, W. T., Langlois, A. J., Ghayeb, J., Petteway, S. R., Weinhold, K. J., Fischinger, P. J., Wong-Staal, F., Gallo, R. C., and Bolognesi, D. P. (1986) *Science* **234**, 1392–1395
3. Goudsmit, J., Debouck, C., Meleon, R. H., Smit, L., Bakker, M., Asher, D. M., Wolff, A. V., Gibbs, C. J., and Gajdusck, D. C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4478–4482
4. Javaherian, K., Langlois, A. J., McDanal, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Herlihy, W. C., Putney, S. D., and Mathews, T. J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6768–6772
5. LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shaddock, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Mathews, T. J., Emini, E. A., and Putney, S. D. (1990) *Science* **249**, 932–935
6. Mathews, T. J., Langlois, A. J., Robey, W. G., Chang, N. T., Gallo, R. C., Fischinger, P. J., and Bolognesi, D. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 9709–9713
7. Gorny, M. K., Xu, J., Gianakakos, V., Karwowska, S., Williams, C., Sheppard, H. Y., Hanson, C. V., and Zolla-Pazner, S. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3238–3242
8. Steimer, K. S., Scandella, C. J., Skiles, P. V., and Haigwood, N. L. (1991) *Science* **254**, 105–108
9. Gupta, G., and Myers, G. (1990) *Cinquieme Colloque des Cent Gardes*, pp. 99–105, Pasteur, Paris
10. Gupta, G., Anantharamaiah, G. M., Scott, D. R., Eldridge, J. H., and Myers, G. (1993) *J. Biomol. Struct. & Dyn.* **11**, 345–366
11. Catasti, P., Fontenot, J. D., Bradbury, E. M., and Gupta, G. (1995) *J. Biol. Chem.* **270**, 2224–2232
12. Fontenot, J. D., Gatewood, J. M., Mariappan, S. V. S., Pau, C.-P., Parekh, B. S., George, J. R., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 315–319
13. Milich, L., Margolin, B., and Swanstrom, R. (1993) *J. Virol.* **67**, 5623–5634
14. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York
15. Roterman, I. K., Gibson, K. D., and Scheraga, H. A. (1989) *J. Biomol. Struct. & Dyn.* **7**, 1–25
16. Kirkpatrick, S., Gelatt, C. D. Jr., and Vecchi, M. P. (1983) *Science* **220**, 671–680
17. Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T., and Wilson, I. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6325–6329
18. Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T., and Wilson, I. A. (1993) *Science* **264**, 82–85
19. Bou-Habib, D. C., Roderiquez, G., Oravec, T., Berman, P., Lusso, P., and Norcross, M. A. (1994) *J. Virol.* **68**, 6006–6013
20. Clements, G. J., Price-Jones, M. J., Stephens, P. E., Sutton, C., Schulz, T. F., Clapham, P. R., McKeating, J. A., McClure, M. O., Thomson, S., Marsh, M., Kay, J., Weiss, R. A., and Moore, J. P. (1991) *AIDS Res. Hum. Retroviruses* **7**, 3–16
21. Ebenbichler, C., Westervelt, P., Carrillo, A., Henkel, T., Johnson, D., and Ratner, L. (1993) *AIDS* **7**, 639–646
22. Hwang, S. S., Boyle, T. J., Lyster, H. K., and Cullen, B. R. (1991) *Science* **253**, 71–74
23. Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1992) *J. Virol.* **66**, 6547–6554
24. Chandrasekhar, K., Profy, A. T., and Dyson, H. J. (1991) *Biochemistry* **30**, 9187–9194
25. Zvi, A., Hiller, R., and Anglist, J. (1992) *Biochemistry* **31**, 6972–6979
26. de Lorimier, R., Moody, M. A., Haynes, B. F., and Spicer, L. D. (1994) *Biochemistry* **33**, 2055–2062
27. Molecular Surface Package, version 2.6.2. (1994) M. L. Connolly, Menlo Park, CA
28. Veronese, F. D., Reitz, M. S., Gupta, G., Robert-Guroff, M., Boyer-Thompson, C., Louie, A., Gallo, R., and Lusso, P. (1993) *J. Biol. Chem.* **268**, 25894–25901

Local and Global Structural Properties of the HIV-MN V3 Loop*

(Received for publication, September 21, 1994, and in revised form, November 10, 1994)

Paolo Catasti^{‡§}, J. Darrell Fontenot[‡], E. Morton Bradbury^{§¶}, and Goutam Gupta^{‡||}

From the [‡]Theoretical Biology and Biophysics Group and [§]Division LS, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 and the [¶]Department of Biological Chemistry, School of Medicine, University of California, Davis, California 95616

Studies of the feasibility of a subunit vaccine to protect against human immunodeficiency virus (HIV) infection have principally focused on the third variable (V3) loop. The principal neutralizing determinant (PND) of HIV-1 is located inside the V3 loop of the surface envelope glycoprotein, gp120. However, progress toward a PND-based vaccine has been impeded by the amino acid sequence variability in the V3 loops of different HIV isolates. Theoretical studies revealed that the variability in sequence and structure of the V3 loop is confined to the N- and C-terminal sides of the conserved GPG crest. This leaves three regions of the V3 loop conserved both in sequence and secondary structure. We present the results of NMR studies that test the validity of our theoretical predictions. Structural studies are reported for the HIV-V3 loop (HIV-MN) in the linear and cyclic (S-S-bridged) forms. For the V3 loop sequence of the HIV-MN isolate, the three conserved secondary structural elements are as underlined below:

turns turn helix
CTRPNYNKRKRIHIGPGRIFYTTKNIIGTIROAHC

Finally, the conformational requirement of the PND in the V3 loop-antibody interaction is tested by monitoring the monoclonal antibody binding to the HIV-MN V3 loop in the linear and cyclic forms by enzyme-linked immunosorbent assay. The binding data reveal that the cyclic V3 loop is a better ligand for the monoclonal antibodies than the linear form although the latter has the same sequence. This means that the monoclonal antibodies recognize the PNDs as conformational epitopes.

The surface of human immunodeficiency virus (HIV)¹ is studded with several copies of the glycoprotein gp120 (1). A segment of gp120 is being considered as a potential antigenic target for protective humoral immunity. This segment is located inside the third variable loop, called the V3 loop (Fig. 1). Monoclonal antibodies raised against the V3 loop can neutralize the viral infection by specifically binding to the principal

neutralizing determinant (PND) located inside the V3 loop (2–4). However, the HIV V3 loop undergoes sequence mutations at a rapid rate in order to escape the immune surveillance of the host cell. But certain segments of the V3 loop remain fairly conserved among different HIV isolates; these conserved segments of the V3 loop are probably essential in the life-cycle of the virus. Therefore, the virus mutates the V3 loop only enough to escape the immune pressure without risking its own life inside the host cell. The elusive nature of the V3 loop calls attention toward two important structural aspects in relation to V3 loop-antibody interaction: (i) the global tertiary folding of the V3 loop and (ii) the structure and presentation of the PND. This article describes the results of a study aimed at exploring these structural aspects by combining two-dimensional NMR spectroscopy, molecular modeling, and antibody binding measurements of the V3 loop from the HIV isolate from Minnesota (HIV-MN).

EXPERIMENTAL PROCEDURES

Materials

The linear and cyclic HIV-MN V3 loops were obtained from Peptide and Protein Research Consultants, Washington Singer Laboratories, UK, using an NIH reagent contract. The purity of the linear V3 loop (TRPNYNKRKRIHIGPGRIFYTTKNIIGTIROAHC) was ~99% by high pressure liquid chromatography; a major peak at 3878.8 was obtained by fast atom bombardment mass spectroscopy which agrees with the average molecular weight (M_r) of 3878.5 as required by the desired structure. The cyclic form (CTRPNYNKRKRIHIGPGRIFYTTKNIIGTIROAHC) showed purity of [SIM]95%, i.e. a single major peak with a few minor impurities. A major fast atom bombardment mass spectroscopy peak at 4083.1 corresponded well with the M_r of 4082.7 as required by the desired structure.

Structure Determination: NMR and Modeling

The methodology for structure determination consisted of three steps: steps 1 and 2 for theoretical analyses of the structure and flexibility of the V3 loop and step 3 for experimental verification of theoretical predictions by NMR spectroscopy.

Step 1: Prediction of Secondary Structures—The secondary structural elements were predicted for a V3 loop sequence by computing the probability S of a given residue i in the V3 loop to adopt a k -type of conformation (k = helix, h , β sheet, b , coil, c , or turn, t), where:

$$S(k,i) = \sum_{l=-\gamma}^{\gamma} \frac{P(k,i+l)}{|l|+1}$$

(The summation is over $1 - \gamma$ to γ , where γ = size of the window chosen to account for the effect of the neighboring amino acid residues: $\gamma = 5$ for h ; $\gamma = 3$ for b ; and $\gamma = 4$ for c or t). $P(k,i)$ = potential for the k -type of conformation of individual residue i derived from the analysis of the single crystal structures of about 65 proteins. The highest $S(k,i)$ determines the conformation k for the i residue (5). Use of any existing algorithm for secondary structure prediction is only 60% accurate. In order to improve accuracy, we tested our predictions by requiring an S-S bridge formation that achieves local energy minima for the cyclic V3 loop; this led to step 2 in our method.

Step 2: Generation of Energy-minimized S-S-bridged V3 Loop—This step involved obtaining an energetically stable S-S-bridged structure

* This work was supported by National Institutes of Health Grant R01 AI32891-01A2. The NMR work was done at the NMR facility at the University of California, Davis, using the GE 500 MHz spectrometer (funded by National Science Foundation Grant DIR-88-04739 and United States Public Health Service Grant RR04795). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

|| To whom correspondence should be addressed: Theoretical Biology and Biophysics Group, T-10, M/S K710, Los Alamos National Laboratory, Los Alamos, NM 87547. Tel.: 505-665-2587; Fax: 505-665-3493; E-mail: gxxg@temin.lanl.gov.

¹ The abbreviations used are: HIV, human immunodeficiency virus; PND, principal neutralizing determinant; V3, third variable; NOE, nuclear Overhauser effect; TFE, 2,2,2-trifluoroethanol; NOESY, nuclear Overhauser and exchange spectroscopy.



Step 3: Use of Two-dimensional NMR Experiments—This step involved (i) sequential assignment of the protons belonging to constituent amino acid residues, (ii) extraction of sequential and medium-range inter-residue interactions by employing full-matrix NOESY simulations with respect to observed NOESY data (10), and (iii) conformational sampling by Monte Carlo simulated annealing subject to the distance constraints derived from NOESY data (11). Two-dimensional NMR experiments were conducted in 90% H₂O, 10% D₂O and in 100% D₂O under the following solution conditions: peptide concentration = 1–3.5 mM, pH = 5.5 in phosphate buffer, temperature = 6–25 °C. NMR experiments were done over a wide range of peptide concentrations to examine whether there were complications due to inter-molecular associations. Analyses of the results of total correlation spectroscopy, double quantum filtered correlation spectroscopy, and NOESY (at two mixing times) experiments in 90% H₂O, 10% D₂O led to the sequential assignment of the spin system (HN, H^α, H^β) and also to the identification of secondary structural states of various residues in the V3 loop (12). Additional NMR data in D₂O further confirmed sequential assignment of all non-exchangeable protons. Prominent secondary structural elements emerged from the characteristic NOE pattern present in the two-dimensional NMR data. In addition to the characteristic structural features (*i.e.* the presence of β -strand, a turn or a helix) a complete set of structural constraints were derived from two-dimensional NMR data: φ from $J_{\text{HN-H}^{\alpha}}$ coupling and inter-residue HN-HN; H^α-HN, H^β-HN distances from two-dimensional NMR experiments in 90% H₂O, 10% D₂O; χ^1 , χ^2 from $J_{\text{H}^{\alpha}\text{-H}^{\beta}}$, $J_{\text{H}^{\beta}\text{-H}^{\gamma}}$, $J_{\text{H}^{\beta}\text{-H}^{\text{ring}}}$ coupling and intra-residue H^αH^β, HNH^β; and inter-residue H^αH^α and H^αH^β distances from two-

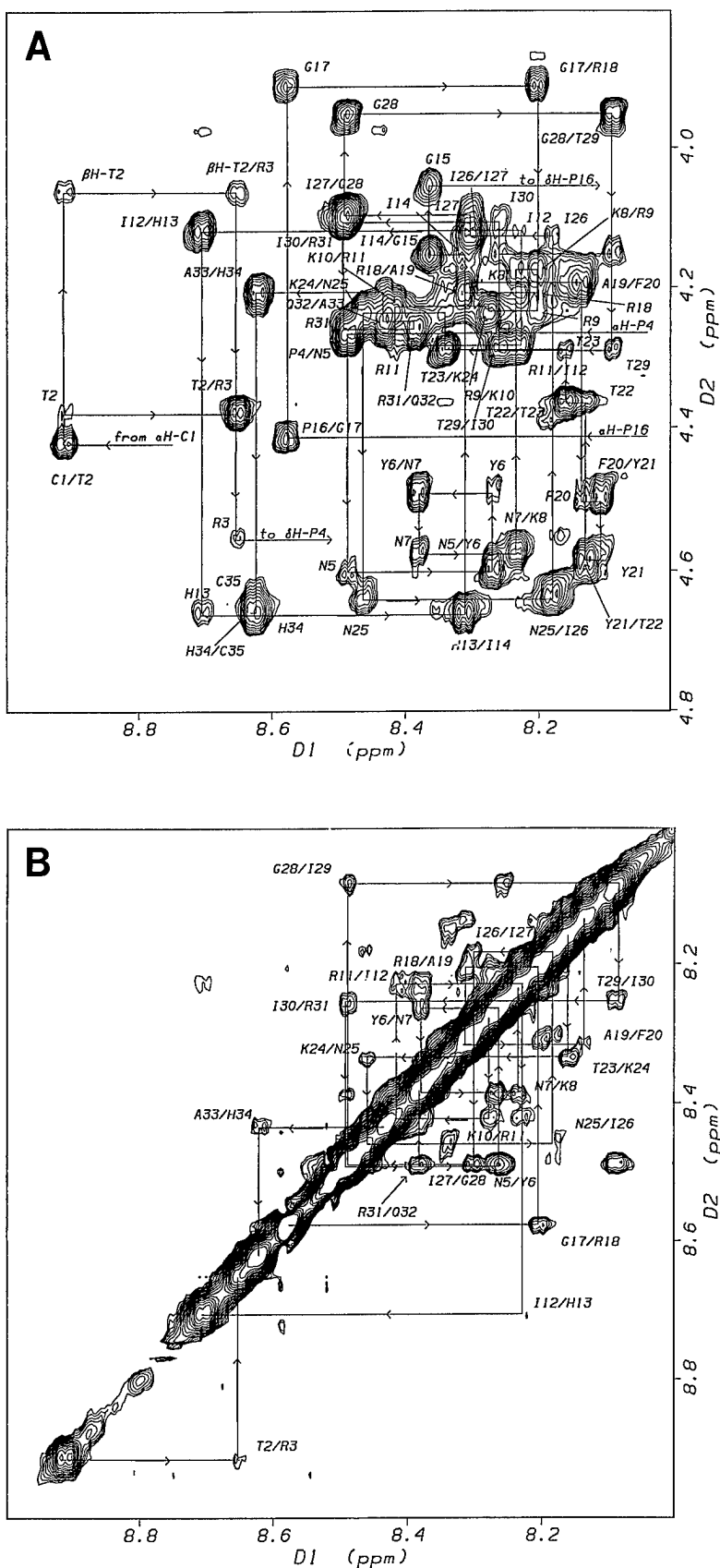


FIG. 2. NOESY cross-sections of the cyclic MN V3 loop in water (peptide concentration = 3.5 mM; pH 4.5). The pulse-sequence due to Sklenar-Bax (24) was used for solvent suppression. Acquisition parameters: data matrix ($t_2 = 2\text{K}$, $t_1 = 1\text{K}$), relaxation delay = 1.5 s, number of transients = 32, temperature = 10 °C. A, the fingerprint HN-H α region; B, the HN-HN region. Sequence specific assignments (25) were obtained starting from F20, which represents a unique residue in the sequence, and moving backward and forward along the connectivity route until completion of the assignments.

dimensional NMR experiments of the V3 loop in D₂O. All these structural constraints were used for structure determination. The structure determination consisted of two steps: (i) extraction of inter-proton distances and (ii) incorporation of these distance constraints for obtaining a cluster of structures in agreement with the NMR data. Full-Matrix NOESY simulations with respect to experimental data at two mixing

times (150 and 300 msec) enabled us to include both primary and higher orders of NOEs. Thus, the complications in the distance estimate using a two-spin model often encountered at a high mixing time due to spin-diffusion (*i.e.* higher order NOEs) are avoided in the full-matrix NOESY simulations where all spins are considered in the relaxation (10). Such a simulation at two mixing times improves the rigor in

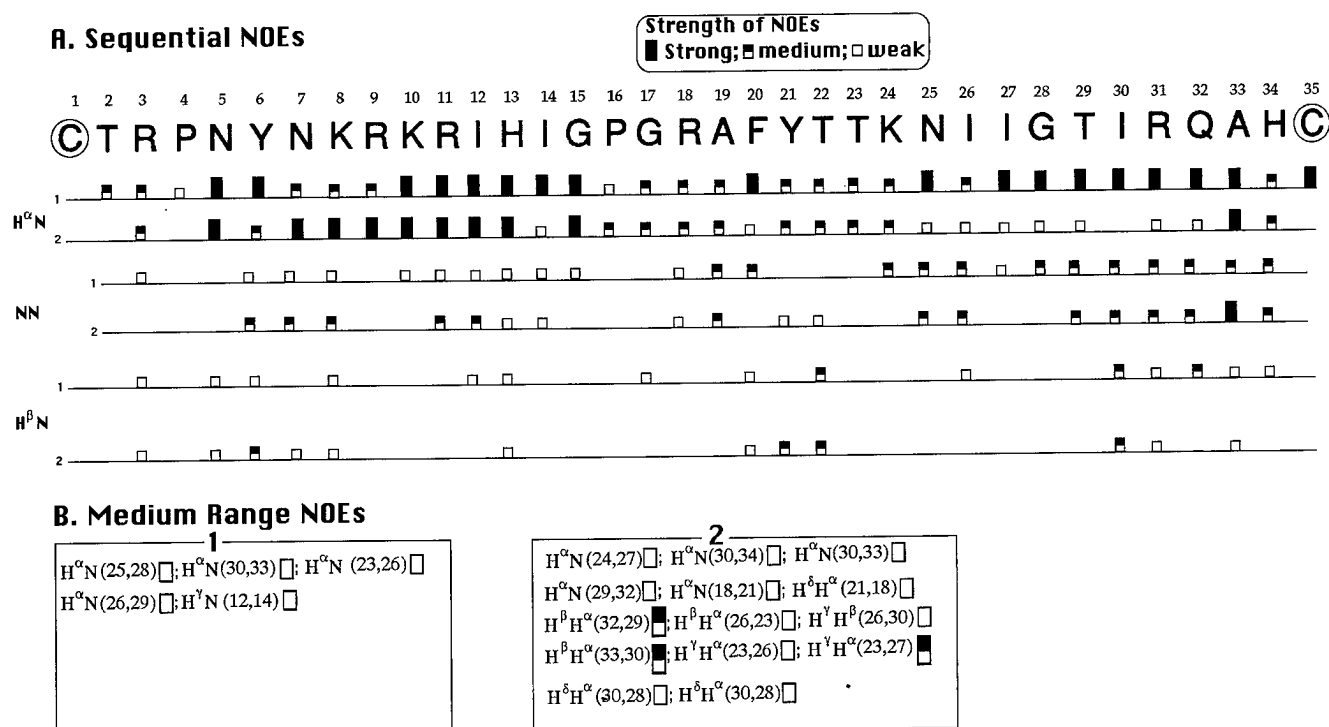


FIG. 3. Summary of the NMR data for the HIV-MN V3 loop in water and in water/TFE (7:3) mixed solvent. Nomenclature of various sequential and medium range NOEs are taken from Wuthrich *et al.* (12). In addition to sequential H^α-HN, H^β-HN, and HN-HN NOEs about 10 sequential H^γ-HN NOEs were obtained in both the solvents. The sequential H^α-HN connectivity for Pro is missing but the sequential H^α-H^β NOEs provide the NOE link. Note the solvent-induced change in the sequential NOE pattern in the C-terminal segment; in the mixed solvent, there is an enhancement of the sequential HN-HN NOEs relative to the corresponding H^α-HN NOEs, indicative of an induction of a helix.

estimating structural constraints for pairwise inter-proton interactions, *i.e.* for each constraint, an upper and a lower limit of the distance. Two types of constraints are identified (11).

Type I is given as

EDIST = 0

if the distance r is within a specified range (between r_1 and r_2)

$$= k(r - r_1)^2 \text{ if } r < r_1$$

$$= k(r - r_2)^2 \text{ if } r > r_2. k: \text{force constant.}$$

Type II is given as

$$\text{EDIST} = 0 \text{ if } r \geq r_1$$

$$= k(r - r_1)^2 \text{ if } r < r_1$$

This type is particularly useful for an unobserved NOE where we can set a lowest allowable distance limit for the corresponding proton pair. The φ -angle constraints are also included as 1–4 distance constraints.

The energy term, EDIST, is added to the force-field QCEP 454 due to Scheraga and co-workers (7). The simulated annealing is performed in the following manner. First, a starting energy-minimized structure is chosen and Monte Carlo simulations are performed for 50,000 steps at 1000K in the $(\varphi, \psi, \omega, \chi)$ -space; the last accepted configuration is stored to be subsequently used as a starting configuration in the next lower temperature-cycle. Second, 50,000 Monte Carlo steps are repeated in several cycles of gradually decreasing temperature until a temperature of 100K is reached. Third, the lowest energy configuration at 100K is further energy-minimized to a low energy gradient. This is the "temperature quenching" step in which thermally excited single bond rotations around the equilibrium positions are quenched. Finally, first through third steps are repeated for 20 different starting configurations.

All sampled low energy structures are analyzed to define the extent of conformational variability. Although 20 starting structures chosen for Monte Carlo simulation are conformationally different, they are not included in the analyses for conformational flexibility. Because these structures obtained by NMR pattern recognition followed by energy minimization do not adequately define the population density of the energy basins to which they belong. However, after simulated anneal-

ing energy barriers are crossed and different energy basins are visited sufficient number of times. Therefore, after such a sampling analyses of conformational variants become physically meaningful.

Solid Phase Peptide Enzyme-linked Immunosorbent Assay with Monoclonal Antibodies

Peptides (0.5 $\mu\text{g}/\text{ml}$) were bound to Dynatech Immulon IV 96 well plates (Chantilly, VA) by overnight incubation in 0.05 M Bicarbonate buffer. The remaining protein binding sites were blocked after one hour of room temperature incubation in 5% Carnation nonfat dry milk in phosphate-buffered saline at pH 7.4. The plates were then incubated with 50 μl of the appropriately diluted monoclonal antibodies for 1 h at room temperature. The plates were then washed three times with phosphate-buffered saline. This was followed by a 1-h incubation with 50 μl of the secondary antibody consisting of Sigma goat anti-mouse IgG conjugated to alkaline phosphatase and diluted 1/3000 in 5% carnation nonfat dry milk in phosphate-buffered saline at pH 7.4. The plates were then washed three times with phosphate-buffered saline. Detection was accomplished with 4 mg/ml phosphatase substrate in 0.25 M diethanolamine with 68 μM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ at pH 9.8. The reaction was terminated after 1 h by adding 50 μl of 3 N NaOH and the absorbance was read at 405 nm.

RESULTS

Molecular Modeling—Amino acid sequence analyses of V3 loops from various HIV-1 strains show that variability in amino acid sequence occurs mainly within specified regions of the V3 loop, leaving three regions that are fairly conserved. Fig. 1A shows the North American consensus V3 loop sequence and the variability in amino acid sequence observed at different sites (13). The relatively conserved regions are: (i) the N-terminal segment which generally includes a site of glycosylation, (ii) the GPG crest, and (iii) the C-terminal segment. Amino acid sequence variability among different V3 loop sequences is confined mainly to the two regions flanking the GPG crest. Fig. 1B shows the HIV-MN V3 loop which, although lacking the glycosylation site at the N terminus, shows a close sequence resemblance with the North American consensus V3 loop. Previously, we have reported a method that defines secondary structural

states of relatively conserved and highly variable regions of V3 loop sequences and predicts the energetically stable tertiary fold(s) of the S-S-bridged V3 loop (8, 9). Using the same method we analyzed the secondary structural elements and the global structure of twenty different V3 loops. The set of V3 loop sequences included the MN isolate, V3 loops from different geographic locations, and V3 loops from isolates showing different tropisms. The analyses predicted the following secondary structural states for the three conserved regions: (i) an 8-residue long loop at the N terminus, (ii) a type II β -turn at the GPG crest, and (iii) a C-terminal helix. Two-dimensional NMR spectroscopy revealed that these conserved structural features were also present in the HIV-MN V3 loop.

NMR Experiments—NMR studies were performed on the control linear peptide only in the aqueous environment. A complete sequential assignment was achieved by combining the total correlation spectroscopy and NOESY data. NMR data showed that only the central principal neutralizing determinant sequence adopted a protruding loop with a flexible GPGR turn and disordered N- and C-terminal segments. The structural studies were performed on the cyclic HIV-MN V3 loop in aqueous and in mixed (water/TFE) solvents. Combination of total correlation spectroscopy and NOESY data in 90% H_2O , 10% D_2O was used to obtain the sequential assignment. Fig. 2, A and B, show the NOESY HN-H^α (fingerprint) and HN-HN regions for 300 ms of mixing time. Note the presence of continuous $\text{HN}(i)\text{-H}^\alpha(i-1)$ sequential connectivity and a number of sequential HN-HN cross-peaks. However, for structure determination one requires relative strengths (not the mere presence) of these cross-peaks. The relative strengths of the sequential and medium range NOEs were obtained by performing NOESY experiments in 90% H_2O , 10% D_2O at two different mixing times (150 and 300 ms). In addition, 27 φ -angle constraints were obtained from the double quantum filtered correlation spectroscopy data of the MN V3 loop in water. Various sequential and medium range NOEs of the cyclic MN V3 loop in aqueous solution are summarized in Fig. 3. The full-matrix NOESY analyses result in 200 pairwise inter-proton distances corresponding to sequential and medium range interactions. The 200 distance and 27 dihedral angle constraints indicate the following effects of the cyclization: (i) induction of an N-terminal loop containing residues 1–9, (ii) stabilization of the GPGR turn, and (iii) formation of two turns at residues 23–26 and 30–33 in the C terminus. The presence of these two turns in the C terminus reveals an incipient helix even in a polar solvent like water.

NMR experiments are, therefore, conducted in a less polar water/TFE (7:3) mixed solvent to promote the formation of a C-terminal helix. Fig. 4, A and B, show the corresponding NOESY HN-H^α (fingerprint) and HN-HN regions for 300 ms of mixing. Spectra in Fig. 4, A and B, indicate that the cross-peaks are better resolved in the mixed solvent although the one-dimensional signals in the mixed solvent are slightly broader than in water. NMR data of the HIV-MN V3 loop in the mixed solvent allow identification of 220 NOESY cross-peaks including the intra-residue HN-H^α NOEs containing the φ -angle information. The induction of the C-terminal helix in the mixed water/TFE solvent is supported by the following NMR data characteristic of a helix (12): (i) the chemical shift of the H^α protons belonging to the residues in the 23–33 segment shows a solvent-induced high field shift, (ii) the sequential NH-NH cross-peaks for the residues in segment showed an appreciable solvent-induced increase in the intensity relative to the corresponding sequential $\text{H}^\alpha\text{-HN}$ intensities, (iii) emergence of the $\text{H}^\alpha(i)\text{-HN}(i+3/4)$ and $\text{H}^\beta\text{-H}^\alpha(i+3/4)$ NOESY cross-peaks for the C-terminal residues. Various sequential and medium range

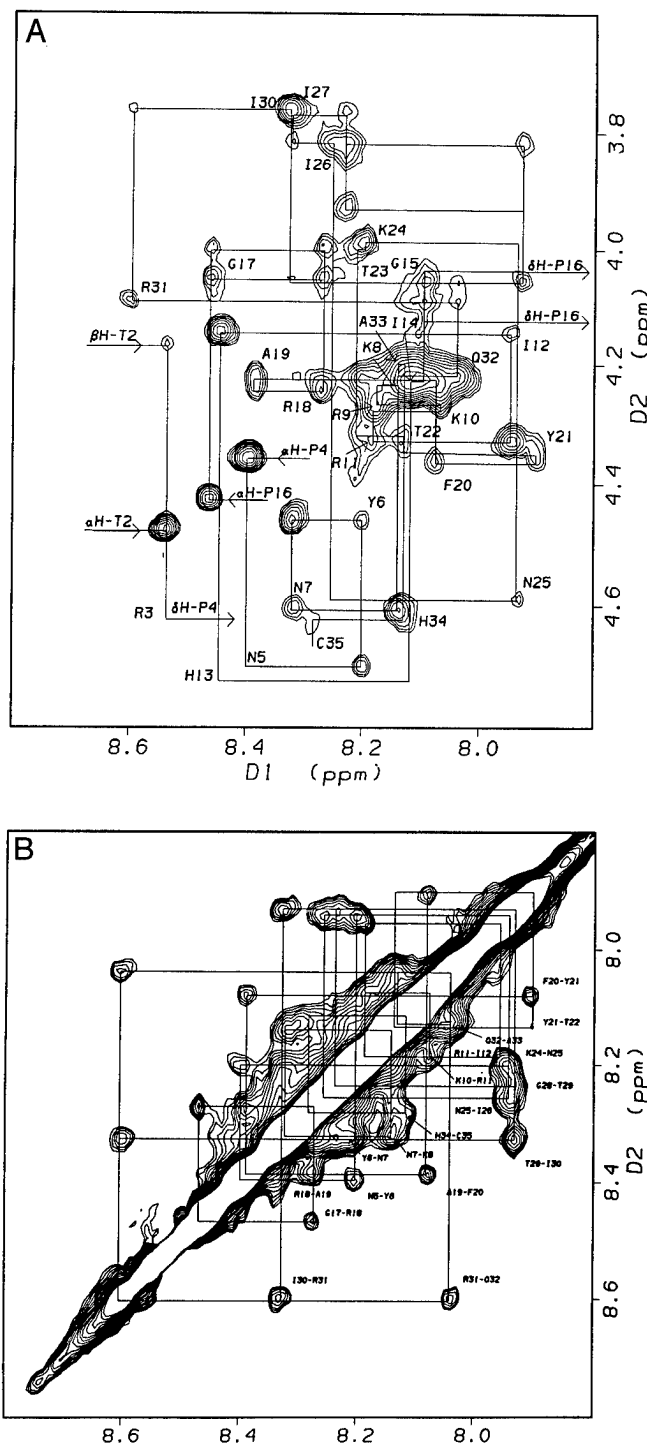


FIG. 4. NOESY cross-sections of the cyclic MN V3 loop in water/TFE (7:3) mixture (peptide concentration = 2.5 mM; pH 4.5; temperature = 10 °C). The H_2O signal was pre-saturated for 1 s during the relaxation delay. Acquisition parameters: data matrix ($t_2 = 2\text{K}$, $t_1 = 1\text{K}$), relaxation delay = 1.5 s, number of transients = 32, temperature = 10 °C. A, the fingerprint HN-H^α region; B, the HN-HN region.

NOEs of the cyclic MN V3 loop in the mixed solvent are summarized in Fig. 3.

Solvent-induced Structural Changes—A Monte Carlo simulated annealing procedure (11) subject to the distance and the torsion angle constraints derived from the NMR data leads to a cluster of structures for the MN V3 loop in water and in a mixed water/TFE solvent. Fig. 5, A and B, show ribbon diagrams of two different folding patterns of the MN V3 loop in water and in a mixed water/TFE solvent, respectively. In each case, the

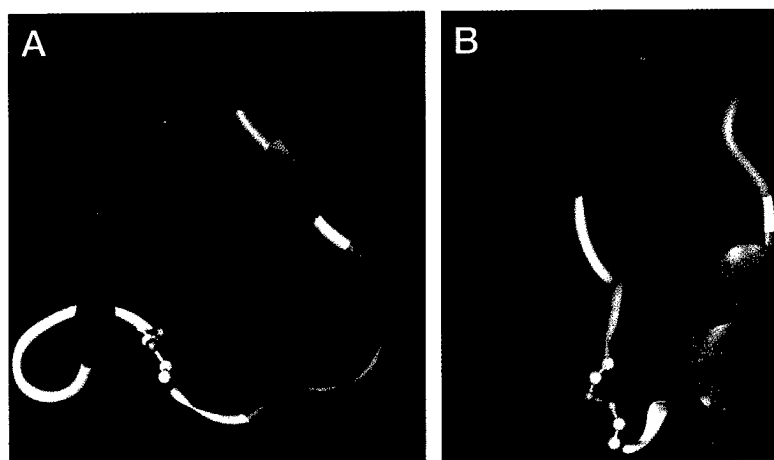


FIG. 5. A and B, the ribbon diagram showing the average folding patterns of the structures the MN V3 loop in water and in mixed water/TFE solvent. In each case, the average is done over 70 sampled low energy structures. Note that, in each case, the neutralizing epitope containing the central GPGR sequence forms a protruding loop even though the local structure and presentation of the loop in two cases are noticeably different. All the sampled structures in A and B showed rms deviations of 0.25 ± 0.01 with respect to 140 specified distance constraints. The structures that satisfy the NMR constraints of the V3-MN loop in water show greater degree of flexibility than those in agreement with NMR data in the mixed water:TFE solvent; this is due to the formation of the C-terminal helix in the mixed solvent.

average folding patterns are shown and averages computed over 70 low energy structures in agreement with the NMR data. Three conserved segments of the V3 loop are color coded: yellow for the N-terminal segment, red for the GPGR crest, and cyan for the C-terminal segment. Note that in both folding motifs the local secondary structures of the N-terminal segment and the GPGR crest remain the same; however, the induction of the C-terminal helix in the mixed solvent changes the spatial interrelations of the three secondary structural elements in the two structures. In addition, a short but well defined β -strand conformation present in aqueous solution disappears in the mixed water/TFE solvent. Therefore, solvent induced changes are also detected in the structure and presentation of the neutralizing epitope (comprising of the central GPGR and 3–4 flanking amino acids on either side) of the MN V3 loop.

The flexibility of the V3 loop in two solvents is markedly different. The nature of flexibility of the two structures is identified by examining the standard deviations (S.D.) of the backbone and side chain torsion angles in these two structures around their average values. These deviations (observed by analyzing energy minimized structures) reflect the lowest possible values because the thermal motions (particularly for the side chains) are filtered off by energy minimization. Tables I and II show the S.D. values in the torsion angles for the V3 loop structures in the aqueous and mixed solvents, respectively. Note that the structure in the aqueous solvent is more flexible than the structure in the mixed solvent. The flexibility of the residues (Lys¹⁰, Arg¹¹, His¹³, Ile¹⁴) in the aqueous V3 loop structure is greatly reduced in the mixed solvent structure because of the induction of a C-terminal helix involving residues 23–33. The helix for residues 23–33 of the V3 loop in the mixed solvent is a distorted one *i.e.* the continuous stretch of $C=O(i) \rightarrow (i+4)HN$ H-bonds is weakened where (ϕ, ψ) values deviate from the ideal helix values (for examples, residues Ile²⁶ and Gln³² in Tables I and II). Also note that the residues 20–23 and 28–34 on the C-terminal side of the V3 loop in the aqueous solvent are more flexible than the corresponding residues of the V3 loop in the mixed solvent.

The biological relevance of TFE-induced structural change is often questioned. However, it may be pointed out that water molecules are largely excluded from the surface of the V3 loop in its active form when it is interacting with antibodies or the host-cell receptor or with other domains of gp120 (14). Therefore, TFE-induced structural changes may shed some light on

the process that accompanies the activation of the V3 loop. In addition, the physico-chemical observation that the C-terminal residues of the V3 loop adopt a helical structure in the mixed solvent is a testimony that the same residues have intrinsic helix forming propensity which is masked in water due to competing water-peptide H-bonds.

Monoclonal Antibody Binding Data for the Linear and Cyclic V3 Loops—The induction of structure due to the S–S bridge between C1 and C35 has a strong bearing on the antibody binding properties of the HIV V3 loop. The lack of a well defined structure in the linear peptide is also evident from antibody binding studies. Binding of linear and cyclic V3-MN loops to three different monoclonal antibodies is compared in Fig. 6. Antibodies 1510, 1511, and 1289 bind to the V3 epitopes KRIHI, HIGPGR, and GPGRF, respectively (15). Note that the cyclic V3-MN loop is a better ligand than the linear analog in all three cases. This is consistent with the NMR evidence that the cyclic V3-MN loop is more structured than the linear analog. As expected, the most pronounced difference in binding occurs for the mAb 1510 which recognizes the sequence KRIHI on the N-terminal side of the GPG crest; this sequence also shows more ordered structure upon cyclization. For the other two antibodies, the difference in binding is smaller, because both of them include the GPGR which even in the linear analog shows a residual turn. Therefore, the NMR and antibody binding studies imply that vaccine attempts using the cyclic V3 loop would be more effective than the linear analog in inducing protective humoral immunity to the conserved structural features. The binding profile of human monoclonal antibodies 1510 and 1511 (both derived from HIV infected patients) reinforces the notion that the cyclic V3 loop presents the epitope structures similar to that found in native gp120. Interestingly, the Lys¹⁰-Arg¹¹-Ile¹²-His¹³-Ile¹⁴-Gly¹⁵-Pro¹⁶-Gly¹⁷ fragment which is a part of the neutralizing epitope of the cyclic MN V3 loop shows the same structure in water and in the mixed solvent as in the complex co-crystal of the neutralizing antibody and the MN V3 loop peptide antigen complex (16).

DISCUSSION

The following conclusions can be drawn based on the data of the HIV-MN V3 loop presented in this article. (i) The S–S bridge between C1 and C35 introduces structure in the V3 loop. (ii) The overall tertiary folding of the V3 loop as well as the local structure at the PND are critical in deciding the affinity of the

TABLE I

The average values (in degrees) of the backbone and side chain torsion angles and their standard deviations of the sampled structures in agreement with the NMR data of the MN V3 loop in water

The backbone torsion angles (marked *) show bimodal distributions.

Residue	No.	φ	Ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
Cys	1	38	62	-178	58						
		26	13	4	2						
Thr	2	-134	121	-179	44	-71	57				
		3	8	3	2	3	2				
Arg	3	-167	90	179	-173	-176	170	-167	-7	177	0
		3	4	3	5	5	3	3	1	0	0
Pro	4	-75	65	-172							
		0	5	4							
Asn	5	-179	105	-179	169	-105	0				
		2	3	4	2	1	0				
Tyr	6	52	48	-179	-163	-114	0				
		8	3	2	7	3	0				
Asn	7	-174	161	178	-147	-17	0				
		5	16	5	18	92	1				
Lys	8	-66	-176	-169	-115	77	176	179	59		
		2	10	13	25	8	6	0	0		
Arg	9	-151	131	-172	-132	174	165	-83	6	-177	0
		10	15	8	39	6	13	2	3	0	0
Lys	10	-159	-60/180*	-177	-155	-179	174	133	-63		
		2	12/13	4	28	4	3	58	6		
Arg	11	-155	90/180*	173	-143	149	-175	121	-1	0	0
		20	11/12	8	22	38	5	44	5	2	0
Ile	12	-131	132	173	-65	173	-176	74			
		5	9	7	7	6	8	3			
His	13	-115	94	-178	-174	41					
		29	8	2	12	42					
Ile	14	-125	171	175	-157	148	-49	62			
		3	15	3	6	16	2	4			
Gly	15	162	94	-169							
		3	2	6							
Pro	16	-75	75	-176							
		0	8	3							
Gly	17	114	-31	175							
		3	3	6							
Arg	18	-73	-6	173	-111	-150	178	-179	0	0	0
		3	10	7	52	35	2	1	0	0	0
Ala	19	-142	174	-177	-51						
		4	22	4	9						
Phe	20	-85	-157	-175	0	98					
		17	10	5	48	18					
Tyr	21	-140	148	179	42	83	-179				
		23	17	3	14	7	0				
Thr	22	-55	169	173	-176	167	-175				
		9	13	7	1	4	1				
Thr	23	-135	61	-177	46	-70	-58				
		13	2	3	3	10	10				
Lys	24	77	-63	-176	-154	140	137	-179	-59		
		3	3	4	1	48	47	0	0		
Asn	25	-139	59	168	-61	-77	178				
		7	5	8	1	8	0				
Ile	26	-37	116	-173	-172	78	-111	121			
		7	3	3	1	4	62	56			
Ile	27	-55	94	-179	-170	172	-149	-53			
		9	3	6	2	0	1	0			
Gly	28	149	65/140*	-166							
		21	13/12	15							
Thr	29	-144	50	-169	42	75	141				
		21	10	8	8	4	67				
Ile	30	-87	-40/160*	179	-110	169	-9	127			
		15	12/11	3	59	2	47	64			
Arg	31	-85	-11	-178	-22	-137	-133	-130	176	179	179
		19	3	3	72	65	59	51	2	0	0
Gln	32	-85	-33	168	128	135	10	179			
		2	12	10	46	41	23	0			
Ala	33	-71	90/160*	-177	0						
		3	10/12	2	60						
His	34	-59	137	179	175	77					
		4	18	3	1	14					
Cys	35	-80	94	179	-179						
		6	16	1	3						

V3 loop for antibody binding. (iii) The structure of the HIV V3 loop is intrinsically flexible and structural transitions of the loop are possible due to a subtle change in the environment (for

example, the effect of TFE) (17). In the total correlation spectroscopy and NOESY spectra of the MN V3 loop in water/TFE, we observed a broadening of the NMR line due to the mixed

TABLE II

The average values (in degrees) of the backbone and side chain torsion angles and their standard deviations of the sampled structures in agreement with the NMR data of the MN V3 loop in a mixed water:TFE (7:3) solvent

The backbone torsion angles (marked *) show bimodal distributions.

Residue	No.	ϕ	Ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
Cys	1	-60/60*	76	-175	-174						
		10/10	9	5	4						
Thr	2	-154	163	-172	-176	164	63				
		4	3	3	1	0	1				
Arg	3	-162	78	175	-168	-143	70	-177	-1	179	1
		3	3	4	2	8	1	4	3	0	1
Pro	4	-75	96	-171							
		0	4	3							
Asn	5	-170	166	176	-155	-99	0				
		3	2	4	5	3	0				
Tyr	6	49	45	-173	-165	-111	0				
		3	4	4	2	1	0				
Asn	7	-172	147	172	-174	29	-1				
		3	3	3	2	2	0				
Lys	8	-68	157	178	-86	68	169	178	59		
		4	3	3	1	0	0	1	0		
Arg	9	-163	138	175	-165	175	-179	-85	7	-177	0
		4	4	3	2	4	2	4	4	1	0
Lys	10	-145	169	178	-94	-175	174	68	-71		
		5	3	3	27	4	1	0	0		
Arg	11	178	161	178	-117	174	-155	82	-6	-3	0
		3	3	4	2	3	3	0	1	0	0
Ile	12	-94	118	178	-66	155	173	62			
		2	2	2	1	0	0	0			
His	13	-160	-177	-172	-148	57					
		4	2	2	3	1					
Ile	14	-136	115	179	-168	172	-38	67			
		3	5	3	1	0	1	0			
Gly	15	146	93	-176							
		3	3	3							
Pro	16	-75	71	-176							
		0	3	3							
Gly	17	101	-34	170							
		3	4	3							
Arg	18	-73	79	178	-169	178	179	-179	0	0	0
		3	2	4	0	0	2	0	0	0	0
Ala	19	-157	-88	160	164						
		4	3	4	1						
Phe	20	-82	136	166	-58	105					
		3	3	4	2	3					
Tyr	21	-72	-177	-171	-57	113	-179				
		3	4	3	6	4	0				
Thr	22	-59	-103	-168	156	156	135				
		3	3	3	1	0	1				
Thr	23	-77	-15	-170	-118	-61	-61				
		4	2	3	20	5	6				
Lys	24	-60	-20	176	-84	-88	177	-179			
		3	3	4	17	6	0	1	-59		
Asn	25	-77	-40	-179	177	42	178				
		3	3	2	3	11	0				
Ile	26	-103	47	165	-40	-176	-161	-167			
		3	3	2	1	1	4	1			
Ile	27	-97	-26	-163	-131	-72	179	-67			
		3	3	2	3	3	0	2			
Gly	28	-89	-24	-171							
		2	6	3							
Thr	29	-79	-22	-178	-167	77	-177				
		3	3	1	5	22	2				
Ile	30	-96	-20	165	79	-101	-176	-60			
		3	2	2	7	7	2	3			
Arg	31	-68	-45	-179	-173	176	179	-179	179	179	179
		4	2	4	2	1	0	0	0	0	0
Gln	32	-114	49	167	-60	-179	-102	179			
		4	4	2	1	2	2	0			
Ala	33	-71	-53	176	-68						
		3	4	3	4						
His	34	-69	93	176	-179	78					
		2	4	2	0	2					
Cys	35	-154	156	179	60						
		3	1	0	12						

solvent. However, unlike the suggestion made in a previous work (18), the broadening was not big enough to hinder complete sequential assignments and NOE determination. We as-

sume this to be due to the fact that we used a low peptide concentration of 2.5 mM in the mixed solvent experiments which prevented the formation of aggregates. (iv) The amino

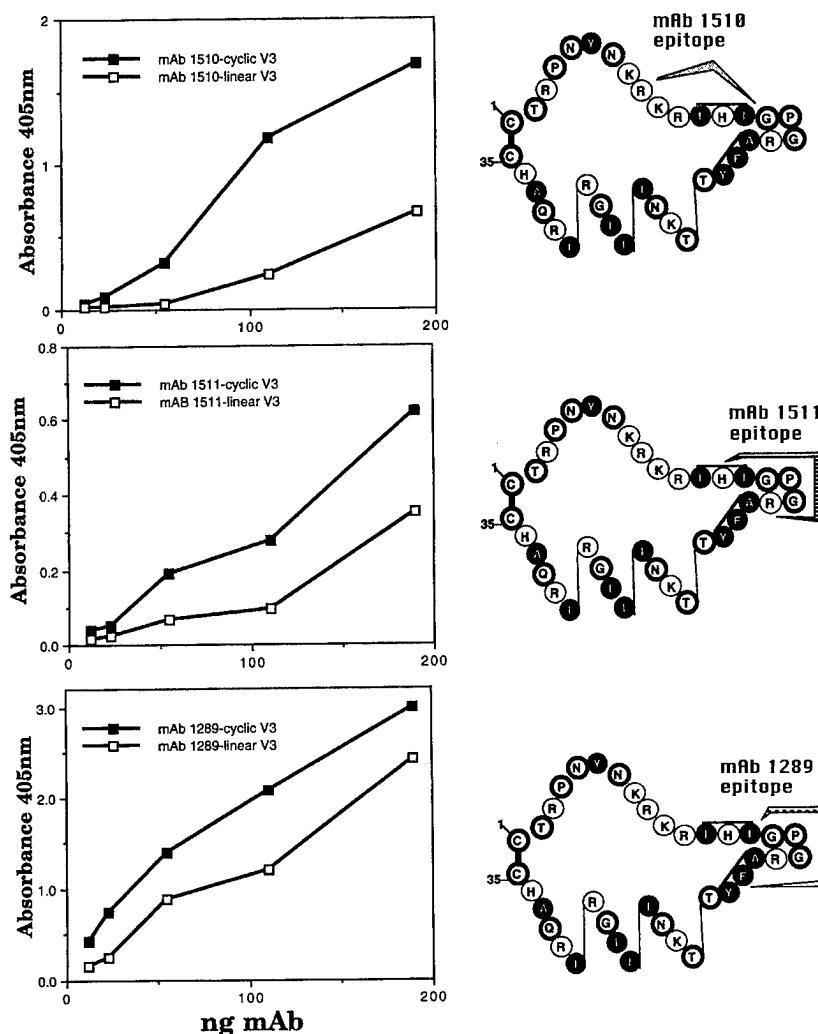


FIG. 6. An enzyme-linked immunosorbent assay showing the preference of monoclonal antibodies for the cyclic over the linear form of the HIV-MN V3 loop. Human monoclonal antibodies 1510 (top) and 1511 (center), and mouse antibody 1289 (bottom) all bind to a greater extent to the cyclic V3 loop peptide. The recognized epitopes 1510 (top), 1511 (center), and 1289 (bottom) are shown on the right. In the schematic representations of the HIV-MN V3 loop shown on the right, solid circles depict hydrophobic residues, open circles charged residues, and outlined circles polar uncharged residues.

acid sequence variability of the V3 loop is restricted on the two sides of the GPG crest (19). Amino acid sequence variability in the regions flanking the conserved GPGR turn can alter the stability of the turn and/or alter (camouflage) the surface accessibility of this conserved secondary structural element. Therefore, structural studies such as ours on this type of sequence variability will be useful in the PND-based vaccine design and also in deciphering the role of the V3 loop in cell fusion (20, 21) and cell tropism (22, 23) where structure and presentation of the V3 loop might be crucially important.

Acknowledgments—We thank Dr. James Bradac of NIH Vaccine branch for providing us with the linear and cyclic MN V3 loop peptides. We also thank Dr. Angel E. Garcia for making technical comments and important suggestions. The antibodies were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: monoclonal antibodies 1510, 1511, and 1289 were all provided by Dr. Susan Zolla-Pazner, Veterans Administration Medical Center, New York, NY.

REFERENCES

- Geyer, H., Holsbach, C., Hunsmann, G. & Schneider, J. (1988) *J. Biol. Chem.* **263**, 11760–11767
- Putney, S. D., Mathews, T. J., Robey, W. G., Lynn, D. L., Robert-Guroff, M., Mueller, W. T., Langlois, A. J., Gharyb, J., Petteway, S. R., Weinhold, K. J., Fischinger, P. J., Wong-Staal, F., Gallo, R. C. & Bolognesi, D. P. (1986) *Science* **234**, 1392–1395
- Goudsmit, J., Debouck, C., Meleone, R. H., Smit, L., Bakker, M., Asher, D. M., Wolff, A. V., Gibbs, C. J. & Gajdusik, D. C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4478–4482
- Javaherian, K., Langlois, A. J., McDaniel, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Herlihy, W. C., Putney, S. D. & Mathews, T. J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6768–6772
- Deleage, G. & Roux, B. (1989) *Prediction of Protein Structure and Principles of Protein Conformation*, pp. 587–594, Plenum Press, New York
- Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–438
- Roterman, I. K., Gibson, K. D. & Scheraga, H. A. (1989) *J. Biomol. Struct. & Dyn.* **7**, 1–25
- Gupta, G. & Myers, G. (1990) *Cinquieme colloque des cent gardes*, pp. 99–105, Pasteur, Paris
- Veronese, F. D., Reitz, M. S., Jr., Gupta, G., Robert-Guroff, M., Boyer-Thompson, C., Louie, A., Gallo, R. & Lusso, P. (1993) *J. Biol. Chem.* **268**, 25894–25901
- Gupta, G., Sarma, M. H. & Sarma, R. H. (1988) *Biochemistry* **27**, 3423–3431
- Gupta, G., Anantharamaiah, G. M., Scott, D. R., Eldridge, J. H. & Myers, G. (1993) *J. Biomol. Struct. & Dyn.* **11**, 345–366
- Wuthrich, K., Billeter, M. & Braun, W. (1984) *J. Mol. Biol.* **180**, 715–740
- LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Mathews, T. J., Emin, E. A. & Putney, S. D. (1990) *Science* **249**, 932–935
- Wyatt, R., Thali, M., Tilley, S., Pinter, A., Posner, M., Ho, D., Robinson, J. & Sodroski, J. (1992) *J. Virol.* **66**, 6997–7004
- Gorny, M. K., Xu, J., Gianakakos, V., Karwowska, S., Williams, C., Sheppard, H. Y., Hanson, C. V. & Zolla-Pazner, S. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3238–3242
- Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6325–6329
- Zvi, A., Hiller, R. & Anglist, J. (1992) *Biochemistry* **31**, 6972–6279
- Chandrasekhar, K., Profy, A. T. & Dyson, H. J. (1991) *Biochemistry* **30**, 9187–9194
- Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–106
- Rusche, J. R., Javaherian, K., McDaniel, C., Petro, J., Lynn, D. L., Grimaldi, R., Langlois, A., Gallo, R. C., Arthur, L. O., Fischinger, P. J., Bolognesi, D. P., Putney, S. D. & Mathews, T. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3198–3202
- Freed, E. O., Myers, D. J. & Risser, R. (1991) *J. Virol.* **65**, 190–194, 333–336
- Hwang, S. S., Boyle, T. J., Lyster, H. K. & Cullen, B. R. (1991) *Science* **253**, 71–74
- Chesebrough, B., Wehrly, K., Nishio, J. & Perryman, S. (1992) *J. Virol.* **66**, 6547–6554
- Sklenar, V. & Bax, A. (1987) *J. Magnet. Reson.* **74**, 469–479
- Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York

ANALYSES OF VARIOUS FOLDING PATTERNS OF THE HIV-1 LOOP

Goutam Gupta and Gerald Myers

Introduction

The surface of the human immunodeficiency virus (HIV-1), which causes the immunodeficiency syndrome (AIDS), is studded with several copies of the surface glycoprotein, gp120 (Allan et al., 1985; Robey et al., 1985; Veronese et al., 1985). As shown in Figure 17-1, the glycoprotein gp120 consists of multiple (S-S)-bridged loops (Leonard et al., 1990). The amino acid sequences of the loops vary across different HIV-1 isolates. Of particular importance is the third variable or the V3 loop (see Figure 17-1). Antibodies elicited by the V3 loop or the parent gp120 block virus infectivity, thus neutralizing the virus (Goudsmit et al., 1988; Javaherian et al., 1988; Putney et al., 1986). Neutralizing antibodies can also block viral infection by inhibiting fusion of HIV-infected cells with CD4-positive uninfected cells (Rusche et al., 1988). It has also been shown that the changes in sequence inside the V3 loop affect the pattern of cell fusion (Freed and Risser, 1991; Robey and Axel, 1990). The V3 loop has been implicated as a determinant in HIV-1 tropism (Wertervelt et al., 1989). Neutralizing determinants are also able to serve as epitopes for both cytotoxic T cells (Takahashi et al., 1989) and helper T cells (Webster et al., 1982). The role of the V3 loop in

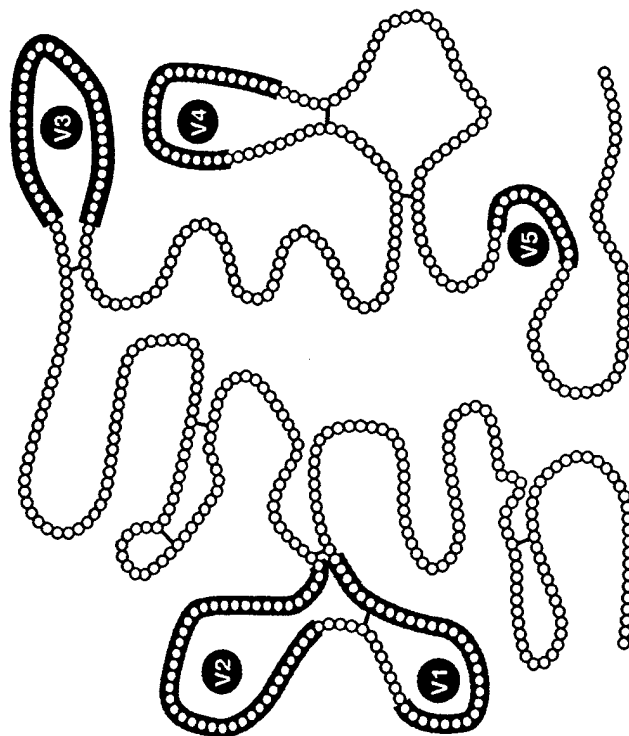


Figure 17-1. Schematic representation of the glycoprotein, gp120, that studs the surface of the AIDS-causing HIV-1. This surface glycoprotein contains several variable domains (shaded and labeled V1 through V5). Each open circle in this diagram represents an amino acid. Disulfide (S-S) bridges are shown as solid bars connecting two cysteines. Primary attachment point of the gp120 molecule on the T4 receptor, CD4 (also a glycoprotein), lies between V4 and V5. However, because of its involvement in neutralization, cell tropism, and cell fusion, the (S-S)-bridged V3 loop has been the subject of the most intense research investigation. The V3 loop is also the most variable region of gp120.

virus neutralization, in cell fusion, and in determining cell tropism has made it a primary focus of vaccine development against HIV-1 infection.

Successful vaccine development, however, has been impeded by the amino acid sequence variability found among different isolates of HIV-1, particularly within the V3 loop sequence. It is probable that HIV-1, like the influenza virus, undergoes changes in the V3 loop to escape selective pressure imposed by neutralizing antibodies *in vivo* (Hart et al., 1990): neutralizing antibodies elicited by the V3 loop from one HIV-1 isolate may not

neutralize HIV-1 isolates with V3 loops with different amino acid sequences (LaRosa et al., 1990). It is therefore important to analyze three-dimensional structures of V3 loop sequences from a large number of HIV-1 isolates to develop a broadly reactive vaccine.

Antibody-V3 Loop Interactions

There is little information about the specific interactions that determine the structure and stability of the antibody-V3 loop complex. However, single crystal studies reveal that protein and polypeptide antigens undergo little conformational change upon binding to specific antibodies (a "lock-and-key" mechanism) (Davies et al., 1990). Assuming such a mechanism, the degree of contact of complementary accessible surfaces of the V3 loop and its antibody (approaching each other as rigid bodies) should determine the specificity of the complex; that is to say, it is important to study the V3 loop structure in the free state because the latter will have a direct bearing on the several thermodynamic factors that determine the contact specificity of the V3 loop-antibody complex. Some of these factors are (i) the loss of entropy in the complex, resulting from the freezing of the mobile backbone segments and the side chains in the V3 loop, which in turn is directly related to the equilibrium tertiary folding and the fluctuations of the secondary structural elements in the fold; (ii) the gain in entropy from the loss of water layers, which is related to the pattern of water relaxation at or near atoms or groups in the V3 loop structure in the free state; and (iii) enthalpic contributions from van der Waals' interactions, H-bonding, and salt bridges, which are related to the surface properties (that is, to orientations of the aliphatic, aromatic, and charge groups with respect to the surface of the V3 loop in the free state).

Surface complementarity between the V3 loop and the antibody might also be achieved, however, via an "induced fit" mechanism in which the V3 loop, the antibody, or both undergo conformational changes to obtain greater surface complementarity. A particularly important situation emerges when two conformationally distinct V3 loop sequences bind the same antibody with similar affinity by exploiting the flexibility of the single binding site in the antibody. Because such a flexibility can potentially expand the antibody repertoire by allowing a single combining site to recognize a range of conformationally distinct V3 loop epitopes, this situation would therefore have a special significance in vaccine development. Whether the "lock-and-key" or the "induced fit" mechanism is operative, knowledge of the structural properties of V3 loops in the free state is essential for explaining the roles of the tertiary fold and the amino acid sequence at the contact region and for determining the structure and specificity of V3 loop-antibody complexes.

Two Assumptions

As just explained, the first step in analyzing the V3 loop-antibody interaction is the structural study of the V3 loop. These structural analyses are relevant because of two important facts: (i) an antibody recognizes a specific three-dimensional structure of the antigen (Davies et al., 1990), and (ii) the (S-S)-bridged V3 loop is the smallest part of gp120 that is likely to present the antigenic-determinant site to the antibody in a manner similar to the entire envelope protein. Therefore, it is essential to analyze the various V3 loop sequences in terms of the *conformational flexure* of their three-dimensional structures. This knowledge will, in turn, help identify the critical residues that determine the stereospecificity of antibody-V3 loop interactions.

Conformational Flexure of the V3 Loop

Analyses of amino acid sequences of the V3 loops from various HIV-1 isolates show that the variability in amino acid sequence occurs only in specified regions of the V3 loop, leaving three regions that remain fairly conserved in the amino acid sequence (Gupta and Myers, 1990). The relatively conserved regions are (1) a site of glycosylation [NN*NT, where N* is the site of sugar attachment and N*XT or N*XS (X being any amino acid) is the sequence requirement for glycosylation], (2) the GPG crest, and (3) the C-terminal helix GDIRQAHC. Our modeling studies on 20 different V3 loop sequences suggest that these invariant amino acid segments have very interesting structural properties (Gupta and Myers, 1990).

NN*NT: A Type I β -Turn

The presence of NN*NT leads to a type I β -turn (Rose et al., 1985; Wilmot and Thornton, 1988) where N* is exposed to the environment, thereby making it accessible for glycosylation. From the analysis of various glycoproteins, accessibility of N*, usually as a member of a β -turn, is viewed as an obligatory stereochemical requirement for glycosylation (Bush, 1982).

GPG Crest: A Type II β -Turn

The residue occurring most frequently in the sequence after GPG is R/Q, followed in frequency by K. Based on simulated annealing studies (Kirkpatrick et al., 1983), a type II β -turn appears to be energetically the most favored structure for the GPCR/Q/K crest. The presence of the GPCR/Q-like sequence (henceforth referred to as the GPG crest) induces a chain folding, which in turn facilitates the S-S bridge between two invariant cysteines in the V3 loop.

GDIRQAHC: A Stretch of Helix

A helix is predicted for the (DIRQAHC) segment. D and Q are the most variable sites in this stretch: D is most often replaced by N (D and N show a preference for similar secondary structures), and Q is replaced by R or K (Q, R, and K are all helix formers). Computer modeling studies, by fully accounting for the variability of D and Q (Gupta and Myers, 1990) show that the DIRQAHC segment in the V3 loop tends to adopt a helical conformation. The presence of the fairly conserved G before the helical segment perhaps helps to retain the integrity of the helix and yet allows an appropriate orientation of the helix by utilizing the flexibility at the preceding G such that an (S-S) bridge can be formed between the two invariant cysteines.

Thus it is clear that variability in the V3 loop sequence of HIV-1 is somewhat limited if the virus is to remain functionally active but escape the pressure imposed by neutralizing antibodies *in vivo*. Our computer modeling studies reveal that the amino acid sequence variability of the V3 loop is largely confined to the two sides of the highly conserved GPCR/Q crest (Gupta and Myers, 1990).

The most effective neutralizing antibodies map at or near the GPCR/Q crest (Gorny et al., 1991; Javaherian et al., 1989; Ohno et al., 1991), making direct contact to five to seven amino acids. The two regions flanking the GPG crest, being hypervariable, are the sites for potential antigenic drift (gradual accumulation of point mutations), so that changes in the amino acid sequence in these regions can either alter the structure and stability of the GPG-crest or induce a structural change in the flanking regions, or both. Such changes in the amino acid sequence in the regions flanking the GPG crest drastically reduce the binding affinity of the antibody produced by the host in response to the virus, and the virus thus escapes a neutralizing effect.

Assuming that the "conserved" (less variable) regions of the V3 loop are also conserved in their secondary structures (i.e., the site of glycosylation, the GPG crest, and the C-terminal helix in Figure 17-2), analyses of sequence variability and structures of the V3 loops can focus on the effect of the two hypervariable regions flanking the GPG crest on the structure and stability of the V3 loop, to which we now turn.

Classification of Tertiary Folding Patterns

Figure 17-2 schematically describes the flowchart of the method of analyzing the three-dimensional structures of the V3 loops belonging to different folding motifs. The technical details of the methodology are described in the Appendix. Following this methodology, we can construct a

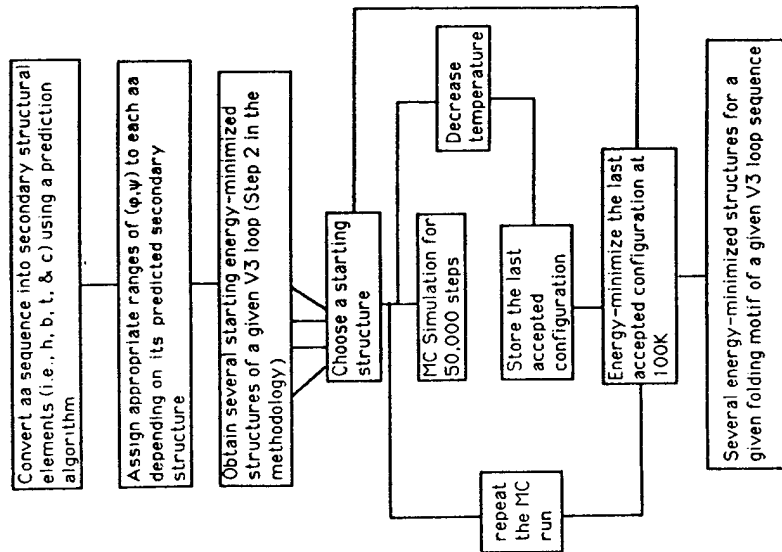


Figure 17-2. Flowchart of the methodology that combines secondary structure prediction, Monte Carlo simulated annealing, and energy minimization. Several starting structures are obtained after step 2 of our methodology. From this set, about 20 such structures with lower energies are chosen as starting structures for simulated annealing. The first MC run (with 50,000 steps) is performed at 1000 K. Temperatures of subsequent MC runs are gradually lowered in 50 temperature cycles to final temperature of 100 K. The 500,000-step MC run for 50 temperature cycles is repeated 20 times for the same starting structure using different random number generators such that the variables are chosen and varied in different ways in different MC runs. This is likely to guarantee conformationally different energy-minimized structures at the end of the calculation. Structures obtained at the end of simulated annealing and energy-minimization are again used as starting structures. Finally, we select 100 low-energy structures for each folding motif for a given V3 loop.

17 • Analyses of Various Folding Patterns of the HIV-1 Loop

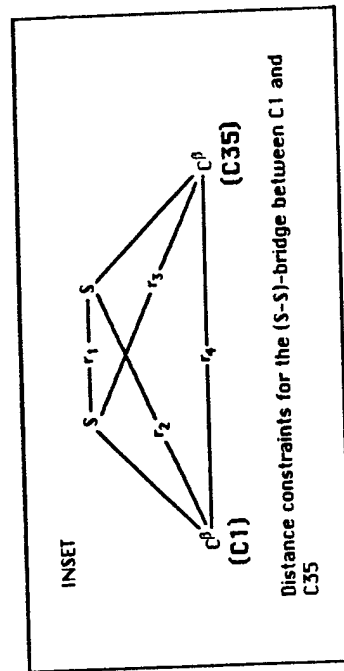
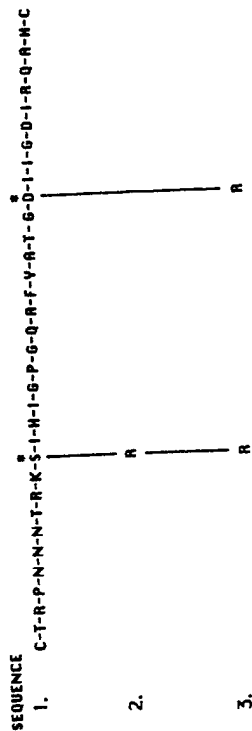


Figure 17-3. Three sequences used in our calculations: sequence 1, consensus (Fletcher, 1984); sequence 2, single site mutation, S11R; sequence 3, from double mutations, S11R and D25R. The invariant cysteines, C1 and C35, are involved in the (S-S) bridge formation. Inset: Four distance constraints for (S-S) bridge between C1 and C35.

Table 17-1. A List of Representative Folding Patterns Expected for V3 Loop Sequences

Pattern	H ₃ ⁺ N ⁻ C	NNNT	GPGR/Q	DIRQAHC
1. hh	turn	turn	turn	helix
2. hb	turn	turn	turn	helix
3. bh	turn	turn	turn	helix
4. bb	turn	turn	turn	helix
5. cb	turn	turn	turn	helix
6. ch	turn	turn	turn	helix
7. bc	turn	turn	turn	helix
8. hc	turn	turn	turn	helix
9. cc	turn	turn	turn	helix

set of V3 loop sequences that are expected to be drastically different from one another in terms of their secondary structures. Each member in this set should represent a given type of tertiary folding pattern. Basing our classification purely on phenomenological grounds (i.e., step 1 of our methodology), assuming a type I turn for the site of glycosylation, a type II turn for the GPGR crest, and a helical stretch for DIRQAHC, and by fully accounting for the variability in the two regions flanking the GPG crest by assigning all possible secondary structural states, we can classify the folding patterns into the categories shown in Table 17-1.

Variable secondary structural elements flanking the conserved GPGR/Q turn span four to five amino acids immediately preceding or following the GPGR turn. Classification of the V3 loops in terms of their tertiary folding patterns immediately provides us with an opportunity to use fewer trial V3 loops in vaccine development. Before vaccine trials can begin, however, we need to obtain complete and accurate knowledge about each of these V3 loop folding patterns.

We carried out modeling calculations on the consensus V3 loop sequence shown in Figure 17-3. Low-energy structures belonging to different folding categories were obtained. The relative stabilities of the folded structures are calibrated with respect to the reference state in which all residues except the type II turn at GPGQ are in the extended state. Figure 17-4 shows different folded states and the reference state for the consensus V3

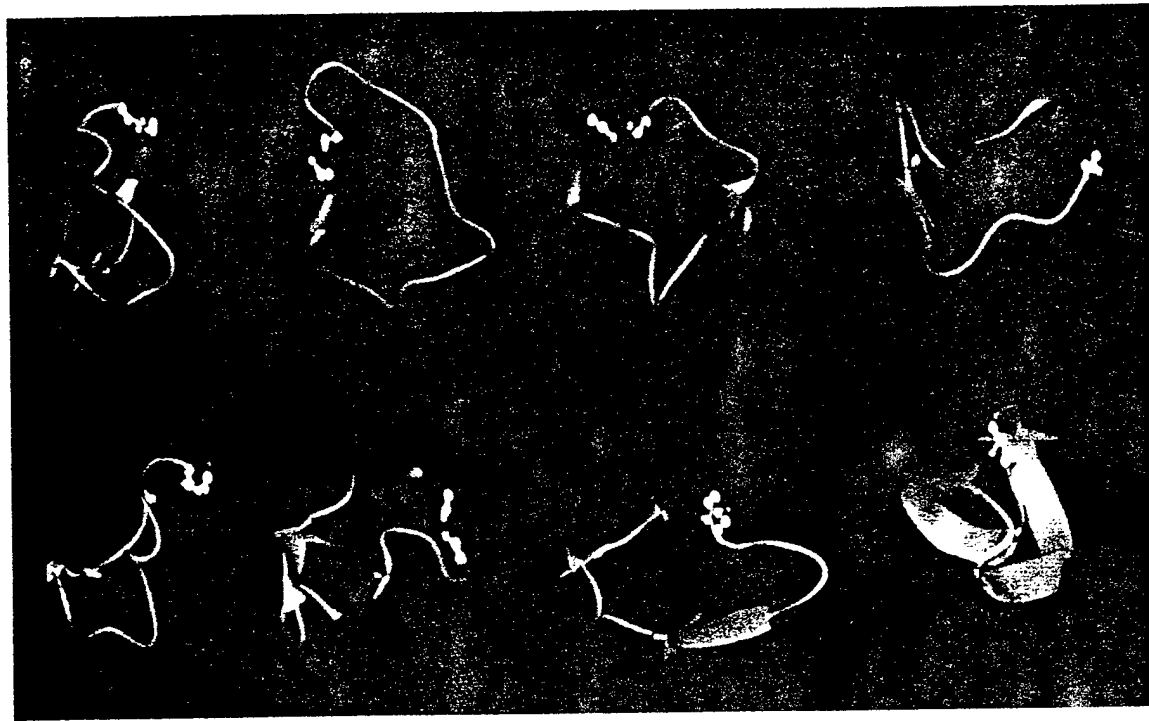


Figure 17-4. Seven folding motifs of consensus V3 loop (sequence 1 in Figure 17-3) and reference extended form. In the reference form, all residues except putative GPGQ-turn are in extended form. As discussed for Figure 17-2, seven folded forms shown here are the average of several structures in same family. Top, left to right: reference state; 'bb'; 'cb1'. Bottom, left to right: 'hc'; 'cc'; 'cb2'. Secondary structural elements are color coded: helix, red; β -extended, green; turn/coil, blue. C-terminal helix is present in all folded forms. C α atoms of C1 and C35, white; S atoms of C1 and C35, magenta. Although folding motifs 'cb1' and 'cb2' belong to same family, the degeneracies of the coil state allow obtaining two variants with similar energies belonging to the same 'cb' motif. The two forms, 'cb1' and 'cb2', are shown to illustrate that there could be significant conformational variation even within the same folding family. Relative energy of stabilization of the folded states is computed as $(ETOT)_{folded} - (ETOT)_{ref}$. ETOTs for each folded form and for reference form are averages of several sampled minima (structures) in same family. Relative energies (kcal/mole) are -32 ('bb'), -45 ('hc'), -35 ('cb1' and 'cb2'); ± 10 ('cc'). In starting structures, for a stretch of residues in the coil state no two neighboring residues belong to the same (ϕ, ψ) region. However, after simulated annealing and energy minimization of starting structures belonging to the 'cb' family, two or more neighboring residues adopted the β or extended conformation. This may be why they show similar stability. However, the 'bb' motif is richer than the 'cb' in secondary structure content.

loop sequence. The secondary structural elements are color coded in Figure 17-3: green for a β -strand (b), red for an α -helix (h), and blue for a coil (c) or a turn (t). The sulfur atoms of the terminal carbons are shown in magenta; the N-terminal carbon is always on the left while the C-terminal carbon is always on the right. Note that in all the folded forms and in the reference state, the GPGQ segment forms the exposed tip of the loop. In all the folded forms, N6* (the attachment point for glycosylation) is also exposed to the environment because this residue is a part of a type I turn. The folded pattern, designated as 'hb', is the most stable of all the folded forms; 'bb' follows as the next most stable form. We therefore discuss in greater detail the structural analyses of these two folded forms.

Conformational Variation Inside the Family of 'hb' and 'bb' Folded Motifs

Table 17-2 shows the average values and the corresponding standard deviations of the torsion angles of the consensus V3 loop belonging to the 'hb' motif. Following the procedure described in the Appendix and in Figure 17-2, 100 low-energy structures were obtained in the 'hb' folding family. The deviations of the total energy, ETOT, for these structures were within 20 kcal/mole. The averages and the standard deviations are computed over these 100 structures. Note that the standard deviations of the backbone torsion angles (ϕ, ψ) of the *trans* peptide units ($\omega \sim 180^\circ$) are within 10° . However, small correlated changes in the (ϕ, ψ) values lead to average root mean square (rms) deviations of about 1.5 Å among all the structures by considering only the positions of the C α atoms. The structures within the 'hb' folding motif are primarily different with respect to the relative position and orientation of the 'helix' and the ' β '- or 'extended' strand on either side of the putative GPGQ turn. It may be pointed out that the standard deviations of the (ϕ, ψ) values for sampled energy-minimized structures only reflect the lower limit of the flexibility because the thermal motions are filtered out after temperature quenching (energy minimization). The largest deviations are observed for the side-chain torsion angles, especially for the residues that are exposed on the surface.

Table 17-3 shows the average values and the standard deviations of the torsion angles for the 100 sampled structures of the consensus V3 loop sequence belonging to the 'bb' folding motif. Here again small changes ($<10^\circ$) in the (ϕ, ψ) values are correlated to produce an average rms deviation of 1.5 Å among all sampled structures with respect to the positions of the C α atoms. As in the case of the 'hb' motif, this variation corresponds to the changes in the relative position and orientation of the β - or extended strand on either side of the putative GPGQ turn.

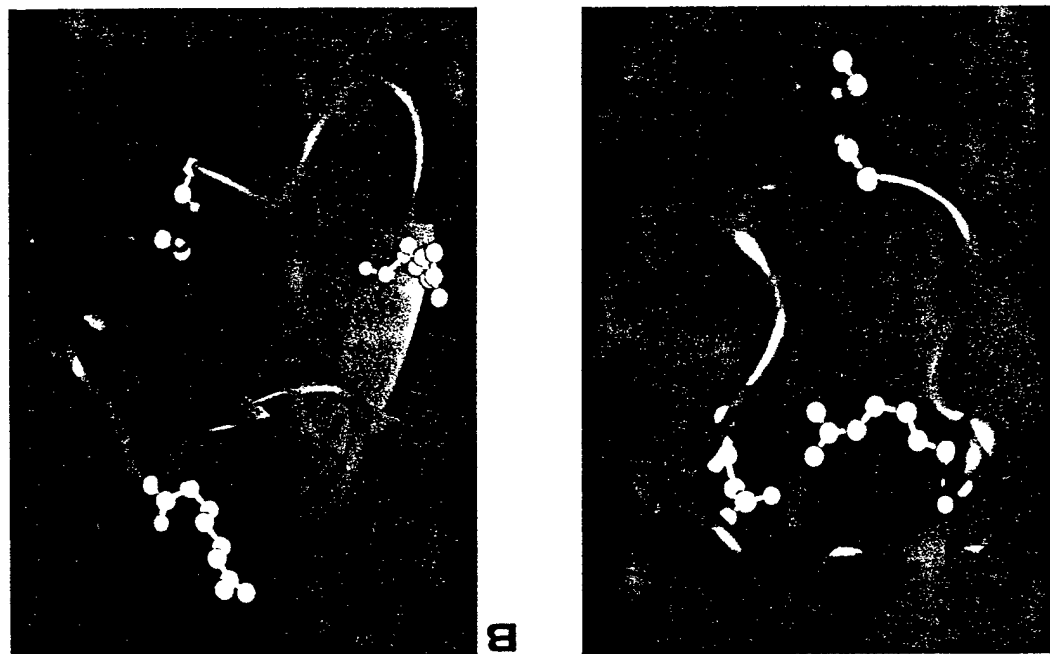


Figure 17-5. Preferred folded forms for sequence 2 and sequence 3 in Figure 17-3. A. 'hb' motif for sequence 2. Note proximity of R11 (cyan) and D25 (white). B. 'bb' motif for sequence 3. Note that R11 and R25 are moved far apart. Color codes of C β and S of C1 and C35 are as in Figure 17-4.

Table 17-2. Averages and Standard Deviations (in Degrees) of the Torsion Angles for the "hb" Conformation

Residue #	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
C 1	-122	34	-178	-67						
T 2	-108	-55	-170	-60	73	-57				
R 3	-56	-53	-171	-131	-176	179	179	0	179	0
P 4	-75	114	179							
N 5	-97	-55	-178	-170	40	0				
N 6	-138	165	-175	-120	-29	0				
N 7	-106	7	-177	-165	-27	0				
T 8	-123	12	-177	55	164	-161				
R 9	-128	-95	-163	-75	-98	-176	-82	4	-177	0
K 10	-66	-45	-167	-66	-179	-172	-68	-169		
S 11	-56	-40	-176	-34	139					
I 12	-63	-47	-174	-68	168	58	-50			
H 13	-58	-50	-174	-58	-58					
I 14	-60	-72	-177	-66	154	50	61			
G 15	153	100	-178							
P 16	-75	89	176							
G 17	101	25	178							

Effect of Amino Acid Substitutions on the 'hb' and 'bb' Folding Motifs

One interesting feature emerges on comparison of the 'hb' and 'bb' folding motifs. The average conformation of the 'hb' folding shows the possible proximity of the side chains of S11 and D25 while in the average conformation of the 'bb' folding motif the same two side chains are moved far apart. As shown in Figure 17-3, a single substitution of the type S11R (sequence 2 in Figure 17-3) and double substitutions of S11R and D25R (sequence 3) are commonly observed. While the virus particles with the V3

Table 17-2. (Continued)

Residue #	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
Q 18	-64	-36	-174	-166	179	-90	0			
A 19	6	7	8	18	8	19	0			
F 20	-144	-57	-178	-56	94					
Y 21	-165	29	-174	40	70	-178				
A 22	-140	130	-177							
T 23	-89	77	179	-54	158	179				
G 24	-169	130	-178							
D 25	-158	81	-179	46	108					
I 26	-146	83	176	42	161	79	-55			
I 27	47	-141	167	-24	-175	-150	-47			
G 28	-141	57	175							
D 29	-61	-38	-175	-59	87					
I 30	-66	-46	-173	-65	153	53	61			
R 31	-64	-38	-175	-69	-178	177	-179	179	179	179
Q 32	-67	-40	-177	-78	176	93	0			
A 33	-58	-45	-176							
H 34	-65	-59	-168	-61	7					
C 35	-80	129	-179	-48						

loops of sequences 1 and 2 are syncytium inactive, the virus particles with the V3 loop of sequence 3 are syncytium active. Therefore, it is of interest whether these mutations that alter surface charge distribution on the V3 loop have any structural significance. Our calculations show that the 'hb' folding is favored for the sequence 2 in Figure 17-3 with a single S11R mutation. In this folded form the side chains of R11 and D25 are in close proximity because of electrostatic stabilization. Such a conformation is unfavorable for sequence 3 (Figure 17-3) with double mutations of S11R and D25R because of electrostatic repulsion, and a 'bb' folding is preferred for this sequence (Figure 17-4). Figure 17-5A shows the average 'hb' fold-

Table 17-3. Average Values and Standard Deviations (in Degrees)

Residue #	ψ	ϕ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
C 1	-148	-80	175	27						
T 2	13	5	5	3						
R 3	-134	-54	-169	-62	88	-63				
P 4	4	4	4	2	17	1				
N 5	-52	-53	178	-112	175	-176	178	0	-179	-17
N 6	3	4	6	4	5	5	8	4	2	18
N 7	-75	117	179							
N 8	0	5	5							
N 9	-115	-54	175	-66	103	0				
N 10	7	5	5	22	4	1				
N 11	-127	160	-177	-133	-95	0				
N 12	6	6	4	33	5	0				
N 13	-95	-16	-178	-171	-76	1				
N 14	6	5	3	1	11	0				
N 15	-92	-5	-178	-13	121	-178				
N 16	4	3	5	39	34	2				
N 17	-117	164	179	-59	-84	179	-82	3	-177	0
N 18	4	4	6	4	21	5	2	2	1	0
N 19	-150	136	-178	-176	177	-173	-66	-167		
N 20	6	4	5	0	0	0	1	0		
N 21	-157	144	178	-77	174					
N 22	7	4	6	38	58					
N 23	-53	142	-178	-172	165	-50	68			
N 24	4	5	7	2	2	1	1			
N 25	-150	175	-178	67	97					
N 26	4	7	5	6	10					
N 27	-154	146	-179	-171	165	70	68			
N 28	4	6	5	2	2	1	1			
N 29	161	85	-176							
N 30	6	6	6							
N 31	-75	93	179							
N 32	0	5	5							

ing for sequence 2 in Figure 17-3. Note that the tip of the positively charged side chain of R11 (shown in cyan) is close (<4 Å) to that of the negatively charged tip of D25 (shown in white). Figure 17-5B shows the average 'bb' folding sequence 3 in Figure 17-3. In this sequence, the side chains R11 (cyan) and R25 (white) are sticking out from two different sides of the V3 loop and are beyond 20 Å.

Such a sequence-specific change in the folding pattern of the V3 loop may have a special bearing on the virus-cell fusion and syncytium. The first step in fusion is the attachment of gp120 to the cell surface CD4 receptor. Because both are glycoproteins, the CD4 and the gp120 would have to

Table 17-3. (Continued)

Residue #	ψ	ϕ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
G 17	99	11	-178							
Q 18	5	6	5							
A 19	-55	-40	177	-172	156	113	0			
F 20	74	103	178	83	13	34	0			
Y 21	4	4	5	4						
A 22	-108	-58	179	-59	95					
T 23	6	9	5	11	9					
G 24	-165	30	177	40	72	-178				
D 25	6	6	4	3	7	0				
I 26	-142	131	178							
I 27	5	6	6							
G 28	-85	77	-179	-54	162	179				
D 29	5	5	4	1	0	0				
I 30	5	5	4							
R 31	-159	127	-178							
Q 32	-159	82	-179	47	108					
A 33	-158	79	179	38	164	86	-56			
H 34	6	6	5	3	1	9	1			
C 35	49	-159	179	-16	-172	-153	-47			
	3	5	5	3	1	2	0			
	-176	63	-177							
	5	6	4							
	-61	-49	-176	-61	70					
	5	3	5	1	8					
	-61	-38	-175	-66	152	51	59			
	3	4	4	0	1	0	1			
	-67	-41	-178	-162	178	-176	179	-178	-179	179
	4	4	2	22	3	5	3	2	0	0
	-66	-43	-177	-75	174	96	0			
	5	5	5	4	4	5	0			
	-60	-61	-176							
	6	6	4							
	-55	-65	-171	-172	155					
	4	5	5	7	69					
	-83	108	179	-46						
	5	39	1	6						

cross a repulsion barrier to approach each other. The increase in the surface distribution of positive charges and the net reduction of negative charges are likely to mediate an efficient surface attachment of gp120 on the host T4 cells. Sequences of the type 3 in Figure 17-3 and the corresponding structure shown in Figure 17-5B fulfill the criteria of efficient cell-surface attachment as mentioned, and therefore provide a potential example of an efficient mode of fusion.

Concluding Remarks

The modeling studies reported here describe the intrinsic conformational flexibility of the HIV-1 V3 loop. These studies also show that the relative stability of one folded form or the other depends on the sequence of the V3 loop. Experimental characterization of different folded forms and the corresponding sequence dependence requires the following: (i) accurate determination of the structure and flexibility of a V3 loop in solution by two- or three-dimensional NMR techniques and (ii) structure-function correlation of antibody-V3 loop binding by combining structural biology and immunological methods. Such experiments are already in progress (Gupta et al., in press; Jong et al., in press; Veronese et al., in press). Results obtained so far demonstrate that, although quite flexible in the free state, the V3 loops when presented to specific antibodies as a part of the whole gp120 can and do adopt well-defined structures.

Acknowledgments

This work is supported by U.S. Army Grant MIPR 92MM2581 and NIH Grant R01 AI32891-01A2 (G.G.). G.G. wishes to thank Drs. Darrell Fontenot and Patricia Reitemeier for proofreading the manuscript.

Appendix

Methodology

The methodology involves the following steps.

Step 1. Prediction of Secondary Structures

The secondary structural elements are predicted for a V3 loop sequence by computing the probability S of a given residue i in the V3 loop to adopt a k -type of conformation (k = helix, β -sheet, coil, or turn), where

$$S(k, i) = \sum_{l=-\gamma}^{\gamma} \frac{P(k, i+l)}{|l|+1}.$$

(The summation is over $l = -\gamma$ to γ , where γ = size of the window chosen to account for the effect of the neighboring amino acid residues: $\gamma = 5$ for a helix; $\gamma = 3$ for a β -sheet; and $\gamma = 4$ for a coil or a turn.) $P(k, i)$ = potential

for the k type of conformation of an individual residue i derived from the analysis of the single crystal structures of about 65 proteins. The highest $S(k, i)$ determines the conformation k for the i residue. (This methodology is adopted from Deleage and Roux, 1989.) Use of any algorithm for secondary structure prediction is only 60% accurate. To improve accuracy, we test our predictions by requiring S-S bridge formation that achieves local energy minima for the cyclic V3 loop; this leads to step 2 in our method.

Step 2. Generation of Energy-Minimized S-S-Bridged V3 Loop

This step involves obtaining an energetically stable S-S-bridged structure for a V3 loop sequence given the secondary structural states of the constituent amino acid residues as obtained after step 1. Appropriate ranges of (ϕ, ψ) values are assigned to all amino acids. For example,

$$\begin{aligned} \phi &= -55^\circ + 25^\circ, & \psi &= -55^\circ + 25^\circ \text{ for residues in a helix} \\ \phi &= -140^\circ + 30^\circ, & \psi &= 140^\circ + 30^\circ \text{ for residues in a } \beta\text{-strand} \\ \phi_{i+1} &= -65^\circ + 20^\circ, & \psi_{i+1} &= -50^\circ + 20^\circ \\ \phi_{i+2} &= -90^\circ + 20^\circ, & \psi_{i+2} &= 0^\circ + 20^\circ \text{ for residues in a type I turn} \\ \phi_{i+1} &= -65^\circ + 20^\circ, & \psi_{i+1} &= 120^\circ + 20^\circ \\ \phi_{i+2} &= 90^\circ + 20^\circ, & \psi_{i+1} &= 0^\circ + 20^\circ \text{ for residues in a type II turn} \end{aligned}$$

(ϕ, ψ) of residues in the coil state are set free to choose any point in the allowed space (for definitions of different secondary structures and corresponding (ϕ, ψ) values, see Ramachandran et al., 1963). We simplify the sequence by assuming A for residues with side chains extending beyond the C^β atom, except for the Ps and the terminal Cs. Our rationale for doing this is that the allowed (ϕ, ψ) space of residues with a side chain longer than A is only a subspace of that allowed for A (Ramachandran et al., 1963).

We obtain an S-S-bridged structure of a V3 loop by using a linked-atom-least-square refinement equation (Sippl et al., 1984) that minimizes the function F in the space (ϕ, ψ) :

$$F = \sum_i \lambda_i G_i + \sum_{ij} (d_{ij}^{mn} - D^{mn})^2,$$

where $G_i (= |r_i - r_i'| = 0)$ indicates distance constraints for an S-S bridge as shown in Figure 17-2. Distances in the S-S-bridged V3 loop configuration are defined as $r_1 = S(C1) - S(C35)$, $r_2 = C\beta(C1) - S(C35)$, $r_3 = C\beta(C35) - S(C1)$ and $r_4 = C\beta(C1) - C\beta(C35)$; corresponding equilibrium distances are $r_1' = 2.04$ Å, $r_2' = r_3' = 3.05$ Å, $r_4' = 3.85$ Å (Korber et al., 1992). λ_i indicates Lagrangian multipliers; d_{ij}^{mn} indicates distance between atom i (type m) and atom j (type n); and D^{mn} indicates the contact limit between atom (type

m) and atom (type n) (Ramachandran et al., 1992). In this refinement, the (ϕ, ψ) values of various residues are treated as elastic variables (i.e., variables with weights) such that by appropriate choice of weights the predicted secondary structural states of residues (after step 1) are minimally altered. This method guarantees a stereochemically orthodox structure for the S-S-bridged ($CA_{33}C$)-like sequence. Finally, appropriate side chains are attached to generate an actual V3 loop sequence, and the potential energy of the system is minimized in the $(\phi, \psi, \omega, \chi)$ space using the force-field of Scheraga and co-workers (Fletcher, 1984). The total conformational energy, ETOT (kcal/mole), has the following components:

- ETOT = EES** (coulomb interactions between pairs of partial charges, dielectric constant = 80)
- + **ETOR** (torsional energy from barriers around single and partially double C-N bonds)
 - + **ENB** (van der Waals attraction and repulsion terms between nonbonded atom pairs)
 - + **ESS** (constraint energy from S-S bonds)
 - + **EDIS** (energy from distance constraints as present in different secondary structures) structural elements).

Several initial structures are chosen within the specified ranges of (ϕ, ψ) , and each structure is subjected to step 2, which finally produces several low-energy structures for the (S-S)-bridged V3 loop. These energy-minimized structures are used as starting configurations in a Monte Carlo (MC) simulated annealing procedure, as described in step 3.

Step 3. Exploration of the Conformational Flexibility of the V3 Loop by MC Simulated Annealing

The simulated annealing is performed (Kirkpatrick et al., 1983; Korber et al., 1992) in the following manner. First, a starting energy-minimized structure is chosen and Monte Carlo (MC) simulations are performed for 50,000 steps at 1000 K in the $(\phi, \psi, \omega, \chi)$ space and the last accepted configuration is stored to be subsequently used as a starting configuration in the next lower temperature cycle. Second, 50,000 MC steps are repeated in several cycles of gradually decreasing temperature until a temperature of 100 K is reached. Third, the lowest energy configuration at 100 K is further energy minimized to a low-energy gradient (Sippl et al., 1984). Finally, the first through third steps are repeated for several different starting configurations. The EDIS term is removed from ETOT in this step.

As discussed in the flowchart of the methodology (Figure 17-2), steps 1-3 produce several low-energy structures for each folding motif of a given V3 loop.

REFERENCES

- Allan et al. (1985): *Science* 228:1091.
 Bush G (1982): *Biopolymers* 21:535.
 Davies et al. (1990): *Annu Rev Biochem* 59:439.
 Deleage G, Roux B (1989): *Prediction of Protein Structure and Principles of Protein Conformation*, p. 587. New York: Plenum.
 Fletcher R (1984): *Practical Methods of Optimization I*. New York: Wiley.
 Freed EO, Risser A (1991): *AIDS Res Hum Retroviruses* 7:807.
 Gorny et al. (1991): *Proc Natl Acad Sci USA* 88:3238.
 Goudsmit et al. (1988): *Proc Natl Acad Sci USA* 85:4478.
 Gupta G, Myers G (1990): *Cinquieme Colloque des Cent Gardes*, p. 99.
 Gupta et al. (1993): *J Biol Struct Dyn* 11:345.
 Hart et al. (1990): *J Immunol* 145:2677.
 Javaherian et al. (1989a): *Proc Natl Acad Sci USA* 86:6768.
 Javaherian et al. (1989b): *Proc Natl Acad Sci USA* 86:8768.
 Jong et al.: *J Virol* (in press).
 Kirkpatrick et al. (1983): *Science* 220:671.
 Korber et al. (1992): In: *Vaccines 92: Modern Approaches to New Vaccines Including Prevention of AIDS*, pp. 75-79. Cold Spring Harbor, New York: Cold Spring Harbor Press.
 LaRosa et al. (1990): *Science* 249:932.
 Leonard et al. (1990): *J Biol Chem* 265:10373.
 Ohno et al. (1991): *Proc Natl Acad Sci USA* 88:10726.
 Putney et al. (1986): *Science* 234:1392.
 Ramachandran et al. (1963): *J Mol Biol* 7:95.
 Robey E, Axel R (1990): *Cell* 60:697.
 Robey et al. (1985): *Science* 228:593.
 Rose et al. (1985): *Adv Protein Chem* 37:1.
 Rusche et al. (1988): *Proc Natl Acad Sci USA* 85:3198.
 Sippl et al. (1984): *J Phys Chem* 88:6231.
 Takahashi et al. (1989): *Science* 224:118.
 Veronese et al. (1985): *Science* 229:1402.
 Veronese et al. (1993): *J Biol Chem* 268:25894.
 Webster et al. (1982): *Nature (London)* 296:115.
 Wertevelt et al. (1991): *Proc Natl Acad Sci USA* 88:3097.
 Wilmot CM, Thornton JM (1988): *J Mol Biol* 203:221.