# Characterization of the Tertiary Structure of Soluble CD4 Bound to Glycosylated Full-Length HIVgp120 by Chemical Modification of Arginine Residues and Mass Spectrometric Analysis

Christine Hager-Braun and Kenneth B. Tomer*

*Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709*

*Received August 7, 2001; Revised Manuscript Received November 24, 2001*

ABSTRACT: The initial step of infection of blood cells with the human immunodeficiency virus, HIV, is the formation of a complex of the viral envelope protein gp120 and its human receptor CD4. We have examined structural features of recombinant soluble CD4 (sCD4) by chemical modification of arginine residues with hydroxyphenylglyoxal and subsequent analysis by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. As R58, R59, R131, R134, R219, R240, R293, and R329 could be derivatized free in solution, these arginine residues were exposed on the surface of the protein. In the noncovalent complex of sCD4 with $HIV_{SF2}$gp120, only R58, R131, R134, R219, R240, R293, and R329 were accessible for the derivatizing agent. R59 was shielded from hydroxyphenylglyoxal and was, therefore, considered to be part of the interaction site with gp120. This indicates that the carbohydrate moieties and the flexible variable loops of the glycosylated full-length gp120 from HIV strain SF2 do not induce a reorganization of CD4 in its binding to gp120 and, therefore, do not appear to significantly affect the structural orientation of the primary receptor in complex with the HIV envelope protein as compared to the binding observed in the crystal structure of CD4 with truncated deglycosylated gp120.

Following the era of genomics, proteomics has evolved with the aim to identify all of the proteins that are expressed in a cell under certain conditions (*1*). Post-translational modifications, such as glycosylation or phosphorylation, alter the one-dimensional structure of a polypeptide and, thereby, influence the function of a protein or its regulation. The tertiary structure of a protein and the protein's interaction with other components, however, connect structure and function. Noncovalent complexes are involved in various aspects of biological reactions (e.g., multimeric protein complexes in metabolic processes, receptor−ligand complexes in signal transduction, and antigen−antibody complexes in immune response). The complexity of structural interactions and the importance of this information necessitates the development of methods that allow for the characterization of complexes of native full-length, post-translationally modified proteins with low sample consumption within a relatively short time.

For an initial characterization of proteins with unknown structure or a complementary description of a native protein with structural information available for its truncated form, chemical derivatization of the constituent amino acid residues and subsequent analysis of the sites of modification by mass spectrometry provides a relatively fast approach to analyze tertiary structural features (*2*). Under conditions that retain the native conformation of the protein, this method permits differentiation of surface exposed residues and residues buried in the core of a protein or as part of an interaction site with another protein. By this means, the surface topology of proteins can be probed (*3*−*5*), epitopes in antigen−antibody complexes mapped (*6*, *7*), and functionally essential amino acid residues in the binding site of regulatory proteins identified (*8*, *9*). We have used this approach to probe the tertiary structure of glycosylated soluble CD4 free in solution and in a complex with glycosylated full-length HIV gp120.[1] The formation of this noncovalent complex of human CD4 and the glycoprotein gp120 from the human immunodeficiency virus, HIV, represents the first step of the virus entry into its target cell.

CD4 is located on the surface of various circulating T cells, including most helper/inducer T cells, monocytes, and macrophages, and has been shown to function as the primary receptor for HIV (*10*, *11*). This glycoprotein, with a relative molecular weight of 55 kDa, consists of an N-terminal extracellular portion, one transmembrane segment, and a C-terminal intracellular tail (*12*). The extracellular residues 1−371 form four immunoglobulin-like domains (D1−D4) and are often recombinantly expressed to yield a soluble CD4 (sCD4) (*13*). Crystal structures have been reported for the human D1−D2 fragment (*14*, *15*) and the D1−D4 fragment (*16*).

Recently, the structure of the D1−D2 fragment bound to HIV gp120 in complex with the antigen-binding fragments of a neutralizing antibody was determined by X-ray crystal-

* To whom correspondence should be addressed. Phone: (919) 541-1966. Fax: (919) 541-0220. E-mail: tomer@niehs.nih.gov.

[1] Abbreviations: ESI-QTOF/MS, electrospray-ionization quadrupole time-of-flight mass spectrometry; F, phenylalanine; HIV, human immunodeficiency virus; HPG, hydroxyphenylglyoxal; K, lysine; MALDI/MS, matrix-assisted laser desorption/ionization mass spectrometry; N, asparagine; R, arginine; W, tryptophan.

10.1021/bi011626k 
Published on Web 01/12/2002

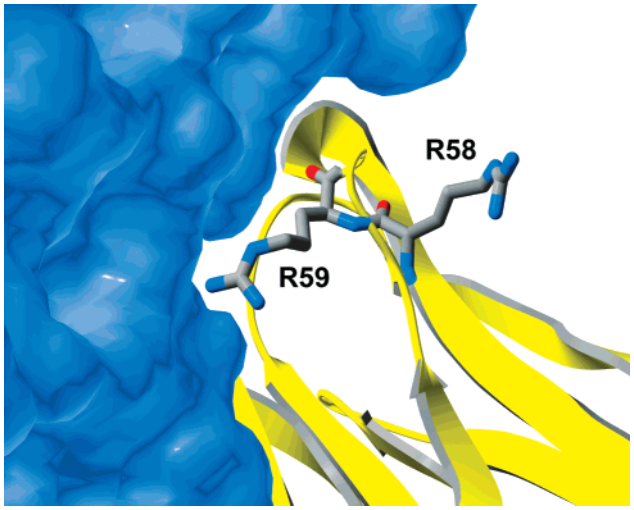


FIGURE 1: Zoomed view of the interaction site of CD4 (yellow ribbon) with truncated deglycosylated $HIV_{HXBc2}$gp120 (blue surface), as derived from the crystal structure (*17*). R59 of CD4 interacts with gp120 while R58 does not, with both arginine residues depicted as sticks.

lography (*17*). HIV gp120 is the envelope glycoprotein located on the surface of the virus. It consists of five conserved regions (C1−C5) and five variable regions (V1−V5), with V1−V4 forming loops on the surface of the protein (*17*−*19*). As both variable and conserved regions are extensively glycosylated, approximately 50% of the mass of gp120 can be attributed to carbohydrate moieties (*19*, *20*).

To crystallize the trimeric CD4/gp120/antibody−fragment complex, recombinant gp120 of the HIV-strain HXBc2 had to be significantly modified by deleting the variable loops V1/V2 and V3 and essentially deglycosylating the protein (*17*). In a more recent study, 14 peptide-proximal *N*-acetylglucosamine residues at the sites of N-glycosylation in gp120 of HIV strains HXBc2 and YU2 were found to be ordered in the crystal structure (*21*). However, Zhu et al. (*20*) determined 25 of 26 conserved glycans of recombinant gp120 HIV strain SF2 to consist of various carbohydrate constituent blocks, such as *N*-acetylglucosamine residues, mannose residues, fucose residues, and so forth, resulting in a heterogeneous highly branched modification of the gp120 protein. Therefore, the question arises as to whether these carbohydrate moieties sterically hinder the binding of CD4 to gp120 when the proteins resemble a more native state and, thereby, impose a "tilted" orientation on CD4 with respect to the orientation found in the X-ray structure of CD4 with deglycosylated gp120. According to the X-ray structure, arginine residue R59 of CD4 is one of the residues that interacts with gp120, while the adjacent R58 does not (Figure 1). Tilting the CD4 molecule would affect the accessibility of both arginine residues, with either R58 being more submerged into the binding pocket or R59 being more exposed.

Here, we present the chemical modification of arginine residues and subsequent analysis by mass spectrometry as an example to gain specific information on the tertiary structure of a noncovalent complex consisting of two fully glycosylated proteins with low sample consumption. This approach adds insight to the structure of the native proteins versus the structure of the truncated proteins elucidated by X-ray crystallography.

## MATERIALS AND METHODS

*Proteins and Chemical Reagents.* Recombinant soluble CD4 (sCD4) was obtained in solution through the AIDS Research and Reference Reagent Program, NIAID, NIH, from R. Sweet, SmithklineBeecham. It was produced in Chinese hamster ovarian (CHO) cells and consisted of the amino acids 1−369 of the mature protein (*13*). The buffer, which contained histidine, mannitol, and polysorbate 80, was exchanged with a sodium phosphate buffer by centrifugation of sCD4 (2 min, 1310*g*, at 25 °C) through a microcolumn (Compact Reaction Columns, USB, Cleveland, OH) filled with BioGel P-2 (45−90 $\mu$m, fine; BioRad Laboratories, Hercules, CA) which had been equilibrated in a 20 mM sodium phosphate buffer (pH 7.5). Subsequently, the actual protein concentration of the sample was determined by a BCA protein assay (Pierce, Rockford, IL). Recombinant $HIV\text{-}1_{SF2}$ glycoprotein gp120 was purchased from Austral Biologicals (San Ramon, CA). This protein was produced in CHO cells and covered amino acids 31−508 of the HIV-1 (strain SF2) envelope protein (*22*). Bovine serum albumin (initial fractionation by cold alcohol precipitation, fraction V) and α-cyano-4-hydroxycinnamic acid were obtained from Sigma (St. Louis, MO), trifluoroacetic acid (TFA) and hydroxyphenylglyoxal from Pierce (Rockford, IL), formic acid (96%) and dithioerythritol from Aldrich Chemical Co. (Milwaukee, WI), and the endoproteinases AspN and LysC from Roche Molecular Biochemicals (Indianapolis, IN). PNGaseF was purchased from Oxford GlycoSciences (Wakefield, MA) and acetonitrile (HPLC grade) from Fisher Scientific (Fair Lawn, NJ).

*Sample Preparation.* For the formation of the noncovalent complex, sCD4 and gp120 were incubated in a molar stoichiometry of 1:1 for 6 h at 37 °C under slow rotation (*23*). The complex sCD4/gp120 and the individual components sCD4 and gp120, free in solution, were incubated at a final protein concentration of 3 pmol/$\mu$L with hydroxyphenylglyoxal (HPG) ranging from 0 to 10 mM in 25 mM $NaHCO_3$ for 15 h at 25 °C in the dark under slow rotation.

*Native Polyacrylamide Gel Electrophoresis (PAGE).* Aliquots of 10 pmol/protein of the samples incubated with HPG or as a control in buffer without HPG were applied to an 8−25% gradient polyacrylamide gel for the PhastSystem (Pharmacia LKB Biotechnology, Uppsala, Sweden). Native PAGE was performed according to the instructions of the manufacturer at a temperature of 16 °C and a run time of 60 min. Subsequently, gels were stained with Coomassie following the protocol "Development Technique File No. 200" from Pharmacia. Essentially, the gels were stained in 0.1% PhastGel Blue R solution in 30% methanol and 10% acetic acid, destained with 30% methanol and 10% acetic acid, and finally incubated in preserving solution which consisted of 13% glycerol and 10% acetic acid.

*Sample Purification.* After derivatization, samples with a protein content of less than 2 $\mu$g were diluted with deionized water to 1 pmol/$\mu$L to increase the volume and separated from excessive HPG by a size exclusion chromatography. For this, microchromatography columns (Compact Reaction Columns, USB) were prepared with Sephadex G-10 (size of 100−1800 Da; Pharmacia), swollen in deionized water to give a bed volume 10 times the sample volume. The sample was loaded on the column and eluted by centrifuga-

tion (2 min, 1310*g*, 25 °C). Samples with a protein amount of more than 10 $\mu$g were purified by high-pressure liquid chromatography using a Protein-C4 column (4.6 mm × 250 mm; Vydac, Hesperia, CA), equipped with a Protein-C4 guard column (Vydac). Solvent A consisted of 0.1% TFA in deionized water, and solvent B was 0.085% TFA in acetonitrile. The sample was loaded on the column which had been previously equilibrated with 5% solvent B, washed with 5% solvent B for 5 min, and eluted with a linear gradient of 5−90% solvent B in 50 min at a flow rate of 1 mL/min. Fractions were collected for 1-min intervals, lyophilized, and resuspended in a small volume of 50% (v/v) acetonitrile and 0.09% (v/v) TFA. Fractions that contained sCD4 and gp120, respectively, were combined to give a total volume of 50 $\mu$L.

*Digest and Deglycosylation.* Combined fractions of sCD4 modified free in solution ($sCD4_{free}$), and sCD4 modified as part of the noncovalent complex sCD4/gp120 ($sCD4_{bound}$), were subjected to cleavage by endoproteinases. For this, the concentration of acetonitrile was reduced to 10% by a 5-fold dilution of the samples with buffer to give a final concentration of 10 mM $NH_4HCO_3$ and a protein concentration of approximately 3.5 pmol/$\mu$L. Endoproteinase AspN was added at a ratio of proteinase/substrate of 1:50, and the samples were incubated for 17 h at 25 °C. Subsequently, the endoproteinase was inhibited by 2 mM EDTA (pH 7.5). The sCD4-derived peptides were deglycosylated by incubation with PNGase F for 8.5 h at 25 °C under slow rotation. Further cleavage of the peptides was achieved by incubation of a 5-$\mu$g aliquot of the AspN/PNGaseF-treated samples with endoproteinase LysC under slow rotation for 17 h at 25 °C at a stoichiometry of proteinase/substrate of 1:25. Aliquots of the peptide samples after AspN/PNGaseF treatment and AspN/PNGaseF/LysC treatment, respectively, were incubated with 2.5 mM dithioerythritol for 30−60 min at 25 °C directly before analysis by mass spectrometry. The remaining samples were stored at −70 °C.

*Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI/MS).* Aliquots of the protein samples purified by gel filtration or by HPLC were cocrystallized on the MALDI target with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (in acetonitrile/deionized water/formic acid, 60/30/10, v/v) according to the dried-droplet method (*24*). Peptides from the digests were diluted 1:1 with deionized water to reduce the concentration of buffer and then cocrystallized 1:1 with matrix solution.

Proteins and peptides were analyzed by MALDI/MS under positive ion conditions on a Voyager DE STR Super equipped with a nitrogen laser (Applied Biosystems, Framingham, MA). The masses of the proteins were determined in the linear mode using a 2 m flight path, an accelerating voltage of 25 kV, a grid voltage of 90%, and an extraction delay time of 400 ns. For each spectrum, 200 shots were combined. An external calibration was performed using the singly and doubly charged ions of bovine serum albumin (mass accuracy of 0.02%). Peptides were analyzed in both the linear and the reflector mode. In the linear mode, the accelerating voltage was 25 kV, the grid voltage 94%, and the extraction delay time 300 ns. In the reflector mode (3 m flight path), the accelerating voltage was set to 20 kV, the grid voltage to 65%, the mirror voltage ratio to 1.12, and the extraction delay time to 400 ns. In both modes, default calibration settings were used, and 100 shots were summed for each spectrum.

The relative abundances of peptides were calculated from their peak height (mean value of 3 measurements ± standard error), assuming similar characteristics for the desorption/ionization process of a specific ion in the sample $sCD4_{free}$ and $sCD4_{bound}$, respectively. The sum of the signals for unmodified and modified peptides was defined as 100% abundance.

*Liquid Chromatography with Mass Spectrometric (LC/MS) and Tandem Mass Spectrometric (LC/MS/MS) Detection.* The peptide sample was first diluted 1:1 with deionized water to reduce the concentration of acetonitrile to 5% and then loaded on a reversed-phase microcolumn for liquid chromatography (PepMap C18, 5 $\mu$m, 75 $\mu$m × 150 mm; LC Packings, San Francisco, CA), which had been equilibrated with 95/5 solvent A/B. Solvent A consisted of 0.1% formic acid in deionized water, and solvent B was 0.1% formic acid in acetonitrile. The amount of peptides injected on the column was approximately 4 pmol. Peptides were washed with 5% solvent B for 5 min and eluted with a gradient from 5% to 60% solvent B in 45 min and from 60% to 90% in 10 min at a flow rate of approximately 200 nL/min. The microcolumn was connected to an electrospray ionization source on a quadrupole/time-of-flight (ESI-QTOF) hybrid mass spectrometer (Micromass, Altrincham, U.K.), which allowed online analysis of the eluting peptides. For the analysis of positively charged ions in the MS mode, the capillary voltage was set to 2832 V, the cone voltage to 20 V, and the source temperature to 80 °C. Total ion chromatograms were recorded for the mass-to-charge range of 200−3000. The relative abundances of peptides were calculated from the integrated signals of the various charge states in the selected ion chromatograms, with the sum of integrated signals of unmodified and modified peptides defining 100% (mean value of two measurements ± standard error). For the detection of ions in the tandem MS mode (MS/MS), the capillary voltage was 2832 V, the cone voltage 20 V, and the source temperature 80 °C. The MS/MS spectra were acquired over a mass range of 50−3000 amu. Fragmentation was induced by collision with argon gas in the collision cell at a collision energy ranging from 4 to 30 V. Data were acquired using MassLynx 3.4 (Micromass, Altricham, U.K.) and deconvoluted via MaxEnt3. (For the nomenclature of fragment ions see refs *25* and *26*.)

*Surface Accessibility of Arginine Residues.* The exposed areas of arginine residues were calculated from the crystal structure of CD4 with gp120 and the fragments of a neutralizing antibody (*17*, *21*) with CD4 docked and undocked to gp120, respectively, using AREAIMOL version 4.1 (*27*).

## RESULTS

*Modification of sCD4 and sCD4/gp120.* Recombinant expressed soluble CD4 (sCD4) contains nine arginine residues according to Deen et al. (*13*) (Figure 2). Surface accessible arginine residues were derivatized on sCD4 free in solution ($sCD4_{free}$), on sCD4 noncovalently complexed with gp120 ($sCD4_{bound}$) as well as on $gp120_{free}$, and on $gp120_{bound}$ using 5 and 10 mM hydroxyphenylglyoxal (HPG), and the proteins were subsequently purified by gel filtration.

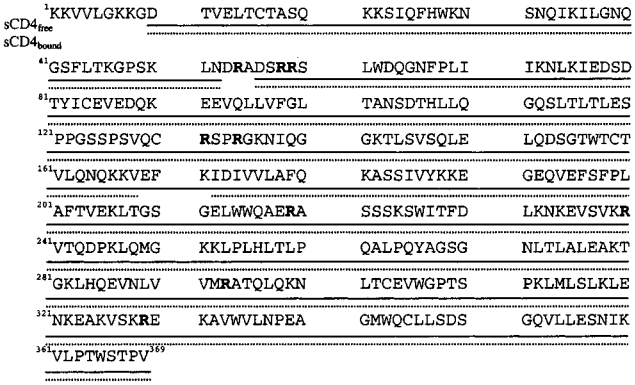


FIGURE 2: Amino acid sequence of soluble CD4 (amino acids 1−369), as derived from the cDNA-sequence according to Deen et al. (*13*) with arginine residues highlighted with bold characters. Sequence coverage as found by MALDI/MS of the peptides derived from modified $sCD4_{free}$ (solid line) and modified $sCD4_{bound}$ (dotted line) is shown by the underlined amino acids.

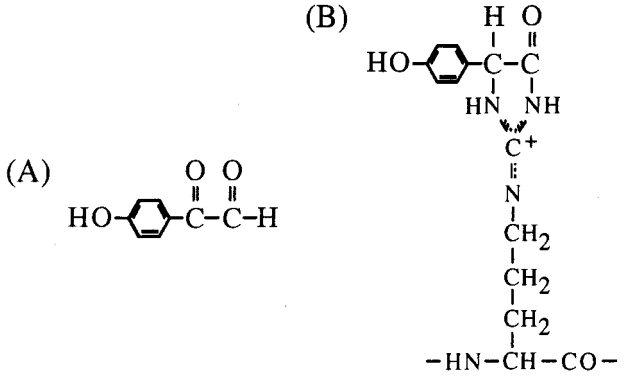


FIGURE 3: (A) Chemical structure of hydroxyphenylglyoxal, HPG, and (B) the product of an equimolar reaction of HPG with an arginine residue, as proposed by Wood et al. (*9*).

Analysis by MALDI/MS showed that the extent of modification was similar for $gp120_{free}$ and $gp120_{bound}$ for both concentrations. This indicated that there was a sufficient molar excess of the reagent present, and, therefore, saturating conditions for the derivatization reaction. At both concentrations of HPG, the sCD4/gp120 noncovalent complex was stable after derivatization of the arginine residues, as demonstrated by native PAGE (data not shown).

The MALDI mass spectra of $sCD4_{free}$ and $sCD4_{bound}$ derivatized with 5 mM HPG and subsequently purified by HPLC showed a mass difference of 352 Da between these two proteins. Although phenylglyoxal and its derivatives can react with the guanidino group of arginine via different mechanisms to yield different final products (*9*, *28*), mass spectrometric analyses of peptides derived from modified sCD4 indicated that a homogeneous reaction mechanism was involved which resulted in a mass increase of 132 Da per derivatized arginine residue (Figure 3). Consequently, on average, 2.7 of the 9 arginine residues were shielded from the derivatizing agent HPG when sCD4 was complexed with gp120.

*Localization of Derivatized Arginine Residues by MALDI/MS.* Soluble CD4, modified with 5 mM HPG free in solution as well as bound to gp120 and subsequently purified by HPLC, was digested with endoproteinases and analyzed by MALDI/MS to identify modified peptides. In the digest with endoproteinase AspN, peptide ions of $sCD4_{free}$ were observed

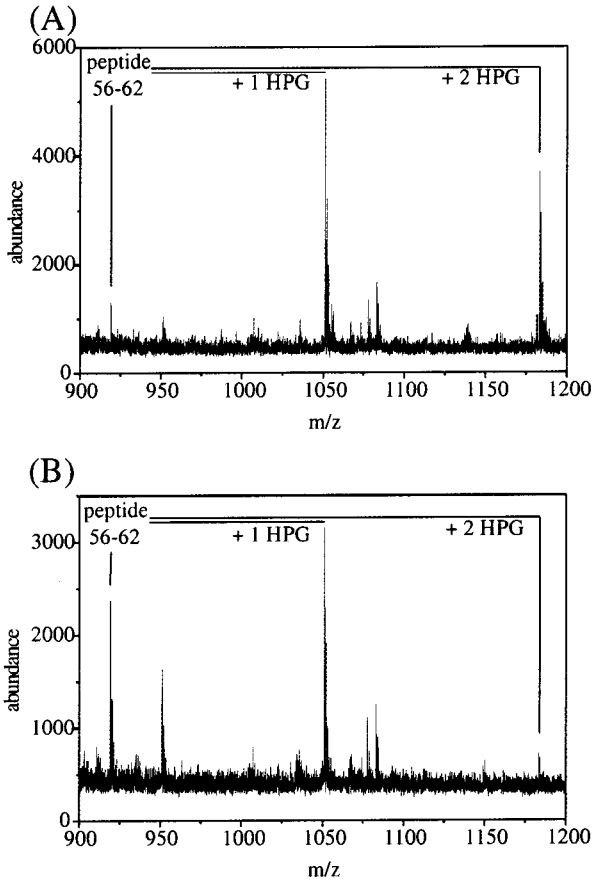


FIGURE 4: Representative MALDI mass spectra of the peptide mixture of derivatized (A) $sCD4_{free}$ and (B) $sCD4_{bound}$ after the digest with endoproteinase AspN.

which covered 66% of the cDNA-derived sequence of CD4 (amino acid 1−369). In $sCD4_{bound}$, the peptides detected in the MALDI mass spectra could be assigned to 74% of the sequence. To identify modified peptides, an increase in mass corresponding to the label was used as an indication. A mass shift of 132 Da was observed for only those peptides that contained at least one arginine residue (average mass accuracy of 0.02%). In the peptide mixture from $sCD4_{free}$, four arginine-containing peptides were found in both their unmodified and modified forms. A shift in mass was observed for peptide 218−229 containing R219, peptide 230−243 containing R240, and peptide 277−322 containing R293. Peptide 56−62 contains two arginines at positions 58 and 59, and a singly derivatized as well as a doubly derivatized peptide could be detected. In the peptide mixture of $sCD4_{bound}$, peptides 56−62, 218−229, 230−243, and 277−322 were present as unmodified and singly derivatized peptides. The ion corresponding to doubly derivatized peptide 56−62 could only be observed at a very low relative abundance (Figure 4). Peptide 53−55 with R54 could not be detected because its singly charged ion of *m*/*z* 361.18 Da was too low for detection by MALDI/MS because of chemical background.

As the digest of the proteins with AspN resulted also in several arginine-containing peptides that had a relatively high mass (>6000 Da), further cleavage into smaller fragments was necessary to obtain additional information on these peptides. Therefore, the peptide mixture obtained from the endoproteinase AspN digest was further digested with

Table 1: Relative Abundances (Mean Value ± Standard Error, in Percent) of Arginine-Containing Peptides 56−62 and 230−243 in Their Unmodified and Modified Forms as Determined by ESI-QTOF/MS and MALDI/MS

| | $sCD4_{free}$ | | $sCD4_{bound}$ | |
|---|---|---|---|---|
| peptide | ESI-QTOF | MALDI | ESI-QTOF | MALDI |
| 56−62 | 6.4 ± 0.7 | 5 ± 3 | 36.6 ± 0.1 | 40 ± 4 |
| 56−62 + 1 HPG | 60.6 ± 0.8 | 58 ± 3 | 61.7 ± 0.2 | 58 ± 4 |
| 56−62 + 2 HPG | 33.0 ± 0.1 | 37 ± 4 | 1.7 ± 0.1 | 2 ± 2 |
| 230−243 | 65.1 ± 6.9 | 62 ± 3 | 56.7 ± 2.3 | 58 ± 4 |
| 230−243 + 1 HPG | 34.9 ± 6.9 | 38 ± 3 | 43.3 ± 2.3 | 42 ± 4 |

endoproteinase LysC, which cleaves C terminal of lysine residues. The sequence coverage resulting from the ions detected in the MALDI mass spectra was 56% for $sCD4_{free}$ and 69% for $sCD4_{bound}$, respectively. In addition to the fragment 56−62, which had also been observed in the AspN digest, peptides 329−344 with R329, 207−224 with R219, 283−299 with R293, and 105−136 with R131 and R134 were detected. In $sCD4_{free}$ and $sCD4_{bound}$, all peptides containing one arginine residue were modified once; in both samples, peptide 105−136, which contains two arginine residues, was detected as a singly and a doubly derivatized peptide. Again, the spectrum of peptide 56−62 with R58 and R59 showed signals for the singly and doubly derivatized peptide for $sCD4_{free}$, whereas in $sCD4_{bound}$, the doubly derivatized peptide was barely detectable above background (signal-to-noise ratio of 2.25). Taking the information provided by the AspN digest and the AspN/LysC double digest together, a combined sequence coverage of 97% for $sCD4_{free}$ and 96% for $sCD4_{bound}$ was obtained (Figure 2). Only the N-terminal peptide 1−9 and the peptide 53−55 with R54 were not observed in the MALDI spectra of either sample. A total of eight arginine residues could be identified as sites of modification. However, only the relative reactivity of R58 and R59 in peptide 56−62 seemed to be different in $sCD4_{free}$ versus $sCD4_{bound}$.

*Determination of the Relative Abundance of Peptides.* To determine if the doubly derivatized peptide 56−62 of *m*/*z* 1183.4 Da was reduced significantly in $sCD4_{bound}$ as compared to $sCD4_{free}$, we determined the relative abundance of various peptides in their modified and unmodified forms. The relative abundances of peptides from the AspN digest which have been analyzed by ESI-QTOF/MS were compared to the data obtained by MALDI/MS analysis (Table 1). Although the standard errors are much larger for small values in the MALDI/MS measurements than in the ESI-QTOF/MS measurements, the relative distribution of modified and unmodified peptides obtained by both techniques corresponded well. Therefore, the data from the AspN digest were supplemented by the MALDI/MS results found for the AspN/LysC double digest (Figure 5). Peptides with one arginine residue showed a higher abundance for their unmodified form than for their modified form, with values ranging from 50.3 ± 1.2% to 74.3 ± 4.3% of unmodified peptide in $sCD4_{free}$ and 58.0 ± 4.2% to 86.3 ± 4.2% in $sCD4_{bound}$. However, values are comparable for a specific peptide in $sCD4_{free}$ and $sCD4_{bound}$. While peptide 329−344 showed almost no change in its relative reactivity in $sCD4_{free}$ and $sCD4_{bound}$, binding of sCD4 to gp120 increased the relative reactivity for R240 in peptide 230−243 by ∼4%. In the noncovalent complex, the abundance of modified

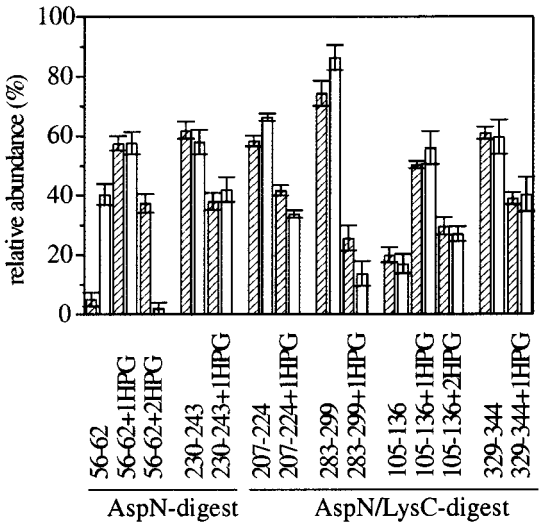


FIGURE 5: Relative abundances (mean value ± standard error, in percent) of arginine-containing peptides from $sCD4_{free}$ (dashed bars) and $sCD4_{bound}$ (white bars) in their unmodified and modified forms, as determined by MALDI/MS.

peptides 207−224 and 283−299 was decreased by ∼8% and ∼12%, respectively. For peptide 105−136 which contains R131 and R134, the monoderivatized peptide dominated. Ratios for the abundances of the peptide with 0, 1, and 2 HPG labels were ∼20:50:30 for $sCD4_{free}$ and ∼17:56:27 for $sCD4_{bound}$, indicating similar surface accessibility of the arginine residues before and after interaction with gp120. For peptide 56−62 containing arginines at positions 58 and 59, significant changes in the relative abundance could be detected; in $sCD4_{free}$, the unmodified form of the peptide contributed only ∼5%, the monoderivatized form ∼58%, and the diderivatized form ∼37%. Upon binding of gp120 to sCD4 the relative abundance of the peptides changed to ∼40% for the peptide without HPG label, ∼58% for the peptide with one label and only ∼2% for the peptide with two labels.

To evaluate the significance of the values determined for the changes in relative reactivity, the ratios of the relative abundance of a specific peptide derived from $sCD4_{free}$ to the relative abundance of the corresponding peptide derived from $sCD4_{bound}$ were calculated. This calculation should compensate for any systematic experimental differences in the two sets of data (Figure 6). A ratio of 1.00 indicates that the surface accessibility was not affected upon binding gp120 to sCD4. Peptides 230−243, 207−224, 105−136, and 329−344 showed a ratio of 0.987 ± 0.116, 1.058 ± 0.253, 1.058 ± 0.144, and 0.994 ± 0.038 (mean value ± standard deviation) for their unmodified and modified forms, with a total of 1.028 ± 0.128 averaged for all of these peptides. These ratios indicate that there were no differences in the surface accessibility of the arginine residues contained in these peptides. The small standard deviations indicate that there were no unexpected experimental errors introduced in the course of the reactions. A moderate change was observed for the modified peptide 283−299 with a ratio of 1.876 and for the unmodified peptide 56−62 with a ratio of 0.124. The ratio for peptide 56−62 in its doubly derivatized form with 18.7 indicated a significant alteration of the accessibility of the arginine residues; one of the two arginine residues was highly protected from the derivatization agent in the noncovalent complex sCD4/gp120.

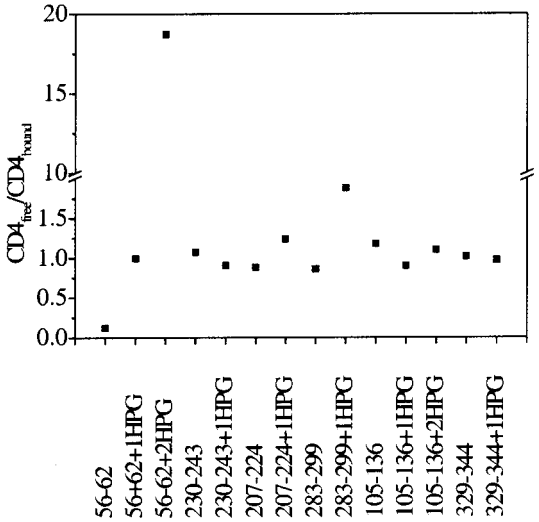


FIGURE 6: Ratio of the relative abundances of a specific peptide derived from $sCD4_{free}$ to the relative abundance of the corresponding peptide derived from $sCD4_{bound}$, as calculated from the data shown in Figure 5.

*Determination of the Derivatization Site in the Monoderivatized Peptide 56−62.* As shown for $sCD4_{free}$, both arginine residues R58 and R59 in peptide 56−62 are accessible to HPG, which results in double derivatization. To determine the actual site of derivatization on the monoderivatized peptide 56−62, these peptides from $sCD4_{free}$ and $sCD4_{bound}$ were analyzed by tandem mass spectrometry, selecting the doubly charged ion of *m*/*z* 526.26 as the precursor ion. In $sCD4_{free}$, the b-series ions identified the sequence RRSLW as part of the peptide $^{56}$DSRRSLW$^{62}$ (Figure 7A). The b ions of $b_4$ and higher had a mass that was shifted by 132 Da as compared to the corresponding ions of the underivatized peptide 56−62. The $b_3$ ion corresponding to the fragment $^{56}$DSR$^{58}$ was observed in an unmodified and a modified form. The analysis of y ions revealed the presence of the $y_4$ ion ($^{59}$RSLW$^{62}$) with and without an HPG label. These data show that the ions of *m*/*z* 1051.52 and 526.26 from $sCD4_{free}$ represent a mixture of two peptides, one of them carrying a label at position R58 and the other one with a modification at arginine residue R59. The diversity of the $b_3$ and $y_4$ ions (modified and unmodified) also explains the relatively low abundance of these signals in the MS/MS spectra.

In $sCD4_{bound}$, the b-series ions corresponding to the sequence RRSLW were observed in accordance to the predicted amino acid sequence of the peptide (Figure 7B). The $b_3$ ion was found exclusively as the HPG-derivatized form $b_3$* at *m*/*z* of 491.20 (calculated *m*/*z* = 491.19). The singly charged ion of $y_4$ was detected only as the unmodified fragment of *m*/*z* 561.32 (calculated *m*/*z* = 561.31). Signals for the underivatized ion of $b_3$ and the derivatized ion of $y_4$* were not observed. Therefore, the singly charged ion of *m*/*z* 1051.52 in $sCD4_{bound}$ reflects a uniformly modified peptide with the derivatization localized only at position R58.

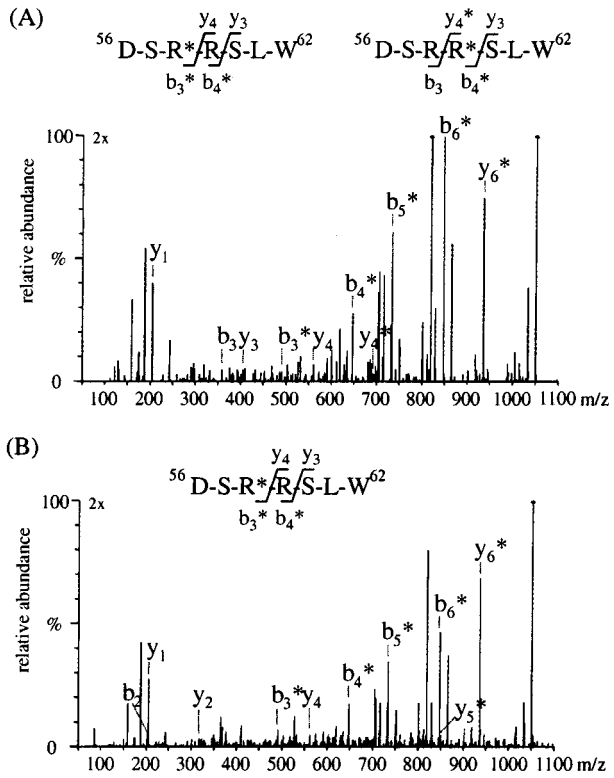


FIGURE 7: Tandem mass spectrum (LC/MS/MS) of the monoderivatized peptide 56−62 of (A) $sCD4_{free}$ and (B) $sCD4_{bound}$. The spectra were deconvoluted so that all ions are singly charged. The asterisk indicates a mass shift of 132 Da due to a derivatized arginine residue.

## DISCUSSION

The initial step of infection of human blood cells by HIV is formation of the complex between the proteins human CD4 and HIVgp120. Although the data from X-ray crystallography of gp120 in complex with sCD4 and the antigen-binding fragment of a neutralizing antibody provide insight into the interactions of these proteins, the information is limited to deglycosylated gp120 with deleted variable loops V1, V2, and V3 and a modified CD4 consisting of only the first two of its four extracellular domains (*17*, *21*). On the basis of the structural data of this truncated gp120 and the characterization of 25 of 26 conserved carbohydrate moieties of recombinant expressed HIV gp120 strain SF2, Zhu et al. (*20*) proposed a model of glycosylated full-length gp120. Although asparagine residues as origin points for N-glycosylation are relatively far away from the CD4 binding site (e.g., 17.14 Å for Cα of N370 to Cα of W394), some carbohydrate chains (e.g., the 9-mannose chain of N370) are long enough to reach the binding site of CD4 (e.g., 4.10 Å for the closest atom of the carbohydrate moiety to Cα of W394; R. Bienstock; unpublished data). To gain additional information on the structure and, especially, the effect of the carbohydrate moieties of gp120 on the interaction with the primary receptor CD4, we chemically modified arginine residues in CD4 free in solution and as part of the noncovalent complex with gp120.

In the mass spectrometric analyses, eight arginine residues could be detected as sites of derivatization with HPG in sCD4 modified free in solution. Therefore, R58, R59, R131, R134, R219, R240, R293, and R329 are exposed on the surface of the protein. A comparison with the calculated surface accessibilitiy of arginine residues based on the X-ray crystallographic structure of CD4 with amino acids 1−369 (*16*) confirms these results. Moreover, the crystal structure showed that R54 would not be accessible to HPG as this arginine residue is partly buried and the guanidino group is pointing to the core of the protein; the mass of the peptide from the

digests containing R54 (*m*/*z* 361.18 for the unmodified form) was too low for MALDI/MS detection because of chemical background and, therefore, evaded analysis.

The relative abundances of peptides in their unmodified and modified forms showed that the derivatization reaction does not derivatize all arginine residues to the full extent, despite a large molar excess of HPG and a long incubation time of 15 h. Consequently, the difference in mass between modified $sCD4_{free}$ and $sCD4_{bound}$ does not permit the precise calculation of the number of arginine residues shielded from the derivatizing agent but rather gives an average number of residues that are not modified. The calculation of 2.7 arginine residues reflects, therefore, an average number of arginines that were unmodifed and probably overestimates the number of residues that are truely inaccessible as a consequence of being buried in the core of the protein or of interaction with gp120. Therefore, the calculation of the ratio of the relative abundance of a specific peptide derived from $sCD4_{free}$ to the relative abundance of the corresponding peptide derived from $sCD4_{bound}$ helps to determine the experimental error. Arginine residues R131, R134, R219, R240, and R329 showed only minor changes in their ratios, which leads to the conclusion that the exposure of these amino acids was barely or not at all influenced by the binding of gp120 to sCD4. The crystal structure of CD4 with the extracellular domains D1−D4 shows that R219, R240, and R329 are exposed on the surface (*16*); however, data on the accessibility of these amino acids in CD4 bound to gp120 are not available for comparison.

The calculation of the accessibility of arginine residues from the crystal structure of the complex (*17*, *21*) revealed that R131 and R134, respectively, are exposed on the surface to the same extent when calculations were performed for CD4 bound to gp120 and after reducing the structure to CD4 only. Although the crystal structure of gp120 neutralizing antibody and domains D1 and D2 of CD4 (*17*) limits the comparison to arginine residues 131 and 134, our data could be confirmed as both residues being exposed on the surface of CD4 bound to gp120. For peptide 283−299 with the surface arginine residue R293 of domain D4 (*16*), a moderate deviation from the average ratio of the relative abundances was observed upon binding to gp120. Complex formation of CD4 and gp120 not only triggers conformational changes in the viral envelope protein (*29*−*32*) but also in the primary receptor. Although gp120 binds to the CDR2-like loop, corresponding to the C′C″ $\beta$ hairpin and the end of strand D in domain D1 of CD4 only (*17*), rearrangements of the three-dimensional structure have been observed for various domains of soluble CD4 (*33*, *34*). The decreased accessibility of R293 for HPG might, therefore, reflect a gp120-induced conformational change.

Peptide 56−62 was doubly modified with HPG when sCD4 was free in solution, which shows that a label on one arginine residue does not impose steric hindrance to the derivatization of the adjacent arginine residue. The decrease in the relative abundance of doubly derivatized peptide 56−62 from $sCD4_{bound}$ reflects a significant alteration in the accessibility of one or both of the arginine residues upon binding to gp120. Steric interference caused by the viral protein would result in a relatively nonspecific derivatization of either R58 or R59. However, the tandem mass spectra showed a site-specific derivatization of R58, while R59 was unmodified. This leads to the conclusion that R59 was shielded from the derivatizing agent, not by an unspecific steric hindrance but by a specific interaction with gp120. In contrast, R58 was still exposed on the surface and, therefore, accessible to HPG. The close proximity of R58 to the binding pocket of gp120 results in partial steric hindrance. Consequently, the relative abundance of singly derivatized peptide 56−62 of $sCD4_{bound}$ was reduced and, simultaneously, the relative abundance of the unmodified peptide increased. Despite the high association constants of gp120 and CD4 with $1.4 \times 10^9$ $M^{-1}$ for HIV-$1_{HXB3}$ and $1.7 \times 10^9$ $M^{-1}$ for HIV-$1_{MN}$ (*35*), equilibrium between the two proteins free in solution and as part of the noncovalent complex may occur during the incubation time of 15 h. This small "contamination" of CD4 free in solution would explain the abundance of 2% of doubly derivatized peptide 56−62 in the sample $sCD4_{bound}$.

The crystallographic structure of the CD4/gp120/neutralizing antibody complex showed that the main interactions with gp120 are formed by the amino acids F43 and R59 of CD4 (*17*, *21*), with R59 of CD4 having multiple contact points with gp120. Conversely, the adjacent R58 does not participate in the actual binding site. These results were found for the terniary complex with gp120 from the laboratory-adapted strain HXBc2 (*17*) and from the primary isolate YU2 (*21*). Accessibility of R58 as determined from the crystal structures (*21*) is barely affected by binding of gp120 to CD4 (101 Å$^2$ for CD4 only as compared to 98 Å$^2$ for CD4 in complex with gp120); in contrast, the exposed surface of R59 is reduced to only half of the area when this arginine residue interacts with gp120 (126 Å$^2$ as compared to 60 Å$^2$). Moreover, R59 forms double hydrogen bonds with aspartate368 of gp120, thereby making the guanidino group of arginine not only structurally but also functionally inaccessible for HPG. The N-terminal domain of CD4 seems to be rigidly held in the deep binding pocket of gp120, because a superposition of the domains of HIV-$1_{HXBc2}$gp120 and HIV-$1_{YU2}$gp120 resulted in a position of CD4 which only differs by 0.5−1.0 Å. These data, however, were obtained with deglycosylated truncated gp120. Our study of sCD4 complexed with glycosylated full-length HIV-$1_{SF2}$gp120 showed that R59 of sCD4 is protected from modification and that R58 is accessible for the derivatizing agent HPG, which leads to the conclusion that R59 interacts with gp120, while R58 is still exposed on the surface. These results suggest that CD4 is, indeed, rigidly bound to gp120; neither the variable loops of gp120 nor any of the carbohydrate moieties in close proximity to the binding site for CD4 seem to impose any steric interference which would affect the structural orientation of the binding partners, thereby resulting in a significantly altered binding of CD4 to gp120. These structural similarities between sCD4 bound to the truncated deglycosylated $HIV_{HXBc2}$gp120 and the primary receptor bound to the full-length glycosylated $HIV_{SF2}$gp120 are further supported by thermodynamic measurements performed with deglycosylated truncated HIV-$1_{HXBc2}$gp120 (core) and recombinantly expressed glycosylated full-length HIV-$1_{WD61}$-gp120; the binding enthalpy changes were very similar (−62 kcal/mol and −63 kcal/mol CD4 for core and full-length gp120, respectively (*32*)). Moreover, similar temperature dependencies of the binding enthalpies were observed for both types of gp120. These data indicate a similar overall binding mechanism for CD4 to gp120 (*32*). Despite the similarities in the association of CD4 to deglycosylated

truncated gp120 and glycosylated full-length gp120, CD4 dissociates more rapidly from core HIV-$1_{HXBc2}$gp120 than from full-length gp120 (strain HIV-$1_{BH10}$) with a 10-fold decrease in affinity for core versus full-length protein. These differences were assigned rather to the deletions of contact-forming segments than to the deglycosylation of the protein (*32*). In contrast to this, Ly and Stramatatos (*36*) found that the extent of glycosylation of loop V2 in HIV-$1_{SF162}$gp120 affects binding to CD4 as determined by infectivity and replication potential studies of various viruses with mutated sites of glycosylation in the V2 loop. These contradictory findings show the complexity of the problem: structural alterations, such as the deletion of the variable loops of gp120 or the removal of the carbohydrate moieties, result in changes of the kinetics, the stability, and the functionality of the complex of CD4 with gp120 (*37*). To understand the structure−function relationship in more detail, similar derivatization experiments should be carried out to address additional amino acids (e.g., acetylation of lysine residues) with the aim to characterize the interaction of K29, K35, and K46 of CD4 with gp120 as they are implicated in binding (*17*). Moreover, the structure of the glycosylated full-length gp120 needs to be characterized to complement the data obtained for glycosylated sCD4 and, finally, give a better insight into the interaction of the viral envelope protein and its human receptor. As examplified in this study on soluble CD4 with domains D1−D4 bound to glycosylated full-length HIV-$1_{SF2}$gp120, chemical modification of specific amino acids and subsequent analysis by mass spectrometry is an approach that can provide information on the structure of interacting post-translationally modified proteins in addition to crystallographic data or NMR studies.

## ACKNOWLEDGMENT

We thank Rachelle Bienstock and Joseph Krahn, Laboratory of Structural Biology, National Institute of Environmental Health Sciences, for the calculation of the distances of various atoms in the model of glycosylated $HIV_{SF2}$gp120 published by Zhu et al. (*20*) and the determination of the surface accessibilities of arginine residues from various crystal structures of CD4.

## REFERENCES

1. Gygi, S. P., and Aebersold, R. (2000) *Curr. Opin. Chem. Biol. 4*, 489−494.
2. Bennett, K. L., Matthiensen, T., and Roepstorff, P. (2000) in *Methods in Molecular Biology* (Chapman, J. R., Ed.), Vol. 146, pp 113−131, Humana Press, Totowa, NJ.
3. Suckau, D., Mak, M., and Przybylski, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5630−5634.
4. Ohguro, H., Palczewski, K., Walsh, K. A., and Johnson, R. S. (1994) *Protein Sci. 3*, 2428−2434.
5. Glocker, M. O., Nock, S., Sprinzl, M., and Przybylski, M. (1998) *Chem.−Eur. J. 4*, 707−715.
6. Fiedler, W., Borchers, C., Macht, M., Deininger, S.-O., and Przybylski, M. (1998) *Bioconjugate Chem. 9*, 236−241.
7. Hochleitner, E. O., Borchers, C., Parker, C., Bienstock, R. J., and Tomer, K. B. (2000) *Protein Sci. 9*, 487−496.
8. Steiner, R. F., Albaugh, S., Fenselau, C., Murphy, C., and Vestling, M. (1991) *Anal. Biochem. 196*, 120−125.
9. Wood, T. D., Guan, Z., Borders, C. L., Jr., Chen, L. H., Kenyon, G. L., and McLafferty, F. W. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 3362−3365.
10. Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) *Nature 312*, 763−767.
11. Klatzmann, D., Champagne, E., Charmaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., and Montagnier, L. (1984) *Nature 312*, 767−768.
12. Maddon, J. P., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L., and Axel, R. (1985) *Cell 42*, 93−104.
13. Deen, K. C., McDougal, J. S., Inacker, R., Folena-Wasserman, G., Arthos, J., Rosenberg, J., Maddon, P. J., Axel, R., and Sweet, R. W. (1988) *Nature 331*, 82−84.
14. Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990*) Nature 348*, 411−418.
15. Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N−h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990) *Nature 348*, 419−426.
16. Wu, H., Kwong, P. D., and Hendrickson, W. A. (1997) *Nature 387*, 527−530.
17. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) *Nature 393*, 648−659.
18. Starcich, B. R., Hahn, B. H., Shaw, G. M., McNeely, P. D., Modrow, S., Wolf, H., Parks, E. S., Parks, W. P., Josephs, S. F., Gallo, R. C., and Wong-Staal, F. (1986) *Cell 45*, 637−648.
19. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. (1990) *J. Biol. Chem. 265*, 10373−10382.
20. Zhu, X., Borchers, C., Bienstock, R. J., and Tomer, K. B. (2000) *Biochemistry 39*, 11194−11204.
21. Kwong, P. D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (2000*) Structure 8*, 1329−1339.
22. Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., and Luciw, P. A. (1985) *Science 227*, 484−492.
23. Borchers, C., Tomer, K. B. (1999) *Biochemistry 38* (36), 11734−11740.
24. Roepstorff, P., Larsen, M. R., Rahbek-Nielsen, H., and Nordhoff, E. (1998) in *Cell Biology: A Laboratory Handbook* (Celis J. E., Ed.) 2nd ed., Vol. 4, pp 556−565, Academic Press, New York.
25. Roepstorff, P., and Fohlman, J. (1984) *Biomed. Mass Spectrom. 11*, 601.
26. Biemann, K. (1988) *Biomed. Environ. Mass Spectrom. 16*, 99−111.
27. Bailey, S. (1994) *Acta Crystallogr. D 50*, 760−763.
28. Takahashi, K. (1968*) J. Biol. Chem. 243*, 6171−6179.
29. Sattentau, Q. J., and Moore, J. P. (1991) *J. Exp. Med. 174*, 407−415.
30. Sullivan, N., Sun, Y., Sattentau, Q., Thali, M., Wu, D., Denisova, G., Gershoni, J., Robinson, J., Moore, J., and Sodroski, J. (1998) *J. Virol. 72*, 4694−4703.
31. Zhang, W., Canziani, G., Plugariu, C., Wyatt, R., Sodroski, J., Sweet, R., Kwong, P., Hendrickson, W. A., and Chaiken, I. (1999) *Biochemistry 38*, 9405−9416.
32. Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J., and Doyle, M. L. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 9026−9031.
33. Denisova, G., Raviv, D., Mondor, I., Sattentau, Q. J., and Gershoni, J. M. (1997) *J. Immunol. 158*, 1157−1164.
34. Yachou, A., and Sékaly, R.-P. (1999) *Biochem. Biophys. Res. Commun. 265*, 428−433.
35. Layne, S. P., Merges, M. J., Dembo, M., Spouge, J. L., and Nara, P. L. (1990) *Nature 346*, 277−279.
36. Ly, A., and Stramatatos, L. (2000) *J. Virol. 74* (15), 6769−6776.
37. Poignard, P., Ollmann Saphire, E., Parren, P. W. H. I., and Burton, D. R. (2001) *Annu. Rev. Immunol. 19*, 253−274.

BI011626K