The primary structure of a soluble form of the CD4 receptor (sCD4) expressed in Chinese hamster ovary cells has been confirmed by mass spectrometric peptide mapping and tandem mass spectrometry. These studies corroborated 95% of the 369-amino acid-long sequence and established the fidelity of translation of the NH₂ and COOH termini including the absence of "ragged ends." The arrangement of the three disulfide bonds in recombinant sCD4 was also established by mass spectrometry and comparative high performance liquid chromatography mapping and shown to be identical to that expected from previous studies of intrachain disulfide bonding in T4 antigens derived from sheep and mouse. No other arrangements of disulfides were detected. Carbohydrate mapping by mass spectrometry was used to establish that both potential Asn-linked glycosylation sites in sCD4 (Asn²⁷¹ and Asn³⁰⁶) have oligosaccharides attached. Structural characterization by mass spectrometry and methylation analysis of the heterogeneous family of oligosaccharides at each of the specific attachment sites indicates that the major components of both families of oligosaccharides have the following biantennary structures:

\[
\begin{align*}
\text{NeuAc}_2 \rightarrow & \text{Gal} \rightarrow \text{GlcNAc}_1 \rightarrow \text{Man}_1 \rightarrow 6 \\
\text{Man}_1 \rightarrow & \text{GlcNAc}_1 \rightarrow \text{GlcNAc}
\end{align*}
\]

where \( m + n = 0-2 \), and \( x = 0.1 \). Minor carbohydrate components having three N-acetyleneuraminic acid (NeuAc) groups and an additional hexose-hexosamine unit were detected by high performance anion-exchange chromatography.

The human CD4 receptor is a 55-kilodalton glycoprotein found predominantly on a subset of mature, thymus-derived (T) lymphocytes and to a lesser extent on monocyte and macrophage related cells. T lymphocytes are involved in the recognition of antigens presented by class II major histocompatibility complex molecules, and substantial evidence indicates that the CD4 receptor directly interacts with class II major histocompatibility complex antigens thereby mediating an efficient immune response (1-3). In man, CD4 also serves as the receptor for the human immunodeficiency virus (HIV) (4-7). Numerous studies have provided evidence for such direct interaction. Certain monoclonal antibodies directed against CD4, such as Leu3a and OKT4A, block HIV infection of T cells and prevent formation of giant multinucleated cells (syncytia) that form by fusion of HIV-infected cells with CD4-bearing (CD4⁺) cells that are uninfected (4, 5). Furthermore, human cell lines lacking the CD4 receptor (CD4⁻) cannot be infected by HIV, but these cells become susceptible to infection when a cDNA encoding CD4 is introduced and expressed (7).

Interaction of CD4⁺ cells with HIV-1 is mediated by gp120, the surface glycoprotein component of the viral envelope glycoprotein gp120/160, which CD4 binds with high affinity (\( K_d \approx 10^{-9} \) M) (8). The CD4 receptor glycoprotein can be co-purified with anti-gp120 antibodies from lysed cells infected with HIV-1, and, conversely, gp120 can be precipitated with anti-CD4 monoclonals (6). These and other data strongly indicate that interaction of CD4 with gp120 is a critical step in viral infection and destruction of CD4⁺ lymphocyte populations.

We and others (8-13) have hypothesized that inhibition of HIV binding by the interaction of viral gp120 with a recombinant soluble form of the CD4 (sCD4) receptor will block virus binding, infection, and virus-mediated cell fusion (8-13). A variety of recombinant sCD4 molecules lacking the membrane-spanning and intracellular domains have been cloned and expressed and shown to inhibit the binding of HIV-1 to CD4⁺ cells, prevent formation of syncytia, and block HIV-1 infectivity in vitro (8-13). More recently, sCD4 has been shown to block diverse strains of HIV-1, HIV-2, and the...
simian virus SIV; however, infection of certain brain and muscle cell lines in vitro could not be blocked by either sCD4 or anti-CD4 antibodies, suggesting that the virus may infect these cells by a mechanism not involving direct interaction of gp120 with CD4 (14).

The amino acid sequence of human CD4 has been deduced by sequencing of the cDNA coding for the protein. The CD4 precursor consists of an NH₂-terminal hydrophobic signal domain, an extracellular domain 370 amino acids in length that has limited sequence homology to the immunoglobulin variable and joining regions, a hydrophobic transmembrane domain, and a charged intracellular domain consisting of 38 amino acids (15, 16). There are 6 cysteine residues in the extracellular domain of CD4 that, by analogy to the reported arrangement of intrachain disulfide bonds in mouse and sheep CD4 (17), are expected to form three disulfide bonds between successive pairs of cysteines. One or more of these disulfide bonds is critical for the binding of gp120 to CD4 presumably due to stabilization of the tertiary structure of the binding region located in the NH₂-terminal region of the extracellular portion of the protein (18–22). The protein also contains 2 asparagine residues in the consensus sequence (Asn-X-Ser/Thr, where X = any amino acid except Pro) required for attachment of carbohydrate (23).

To date, only limited structural analyses of human CD4-related proteins have been reported. Here we provide the first detailed structural characterization of a recombinant sCD4 in which mass spectrometry and tandem mass spectrometry have been used to corroborate the primary structure, including disulfide bond arrangement, determine the location and extent of Asn-linked glycosylation, and characterize the major structural classes of carbohydrate at the two specific attachment sites. Detailed sequence and stereochemical analysis by exoglycosidase digestion of the major and minor glycoforms present at the two attachment sites will be presented elsewhere.¹

**EXPERIMENTAL PROCEDURES**²

RESULTS AND DISCUSSION

Soluble CD4 was expressed by dihydrofolate reductase complementation in Chinese hamster ovary (CHO) cells as described previously (9). The protein was purified to apparent homogeneity (Fig. 1) (see "Experimental Methods") in the Miniprint with an overall recovery of 5–10 mg of sCD4/liter of serum-free conditioned medium. Measured amino acid composition and NH₂-terminal sequence were consistent with the predicted sequence of the molecule (29). Carbohydrate composition analysis employing the method of Chaplin (44) indicated the presence of mannose, galactose, N-acetylglucosamine, fucose, and N-acetylgalactosamine. No N-acetylgalactosaminic acid was detected, suggesting our preparation of recombinant sCD4 contains only N-linked (versus O-linked) oligosaccharides.

Peptide Mapping by FABMS—Peptide molecular weight determination by fast atom bombardment mass spectrometry (FABMS, also referred to as liquid secondary ion mass spectrometry), coupled with sequence analysis of specific peptides by tandem MS, is an ideal complement to Edman degradation for structural characterization of proteins (24-28). The strat-egy as it was applied to recombinant sCD4 is illustrated in Fig. 2, A and B. In the FABMS peptide mapping procedure, the molecular weights of peptides in digests are determined by FABMS and fitted (based on established rules for cleavage by the specific enzyme or chemistry employed) to the known or deduced sequence of the protein. This fitting is accomplished with the aid of computer programs which, given the reaction conditions employed and the predicted sequence of the protein as input, produce lists of molecular weights and sequence locations of the expected peptides. The protein is usually reduced and alkylated prior to proteolysis (unless the intent is to assign disulfide bonds, see below) in order to increase its susceptibility to cleavage. The proteases employed in the present study were trypsin and Staphylococcus aureus V8, individually or in combination. These enzymes were also used on samples that had not been reduced previously and alkylated in order to define the disulfide bond arrangements (see Fig. 2B and below) and on samples of sCD4 diglycosylated with peptide N-glycosidase F to define carbohydrate attachment sites (see below). Following analysis of the complex proteolytic digests by FABMS (for example, see Fig. 9) the mixtures were fractionated by reversed phase HPLC (Figs. 3 and 8) and the resulting fractions reanalyzed by FABMS. Signals for peptides that are not detected by direct analysis of the entire digest are often observed following HPLC fractionation resulting in greater coverage of the amino acid sequence. In addition, signals for peptides observed previously in the complex mixture are often much stronger in the simplified mixtures. The FABMS methods presented here are, in general, sufficiently sensitive and reliable to detect alterations present at >5% in the regions of the protein sequence mapped (i.e. the 95% mapped in the present work) provided that such changes result in a mass shift from that predicted based on the DNA or cDNA sequence (24, 25, 28).

FABMS peptide and carbohydrate mapping data for recombinant sCD4 are summarized in Fig. 4. The amino acid sequence defined by these studies corresponds to residues +3 to +371 of the sequence by Maddon et al. (15) but with a Lys at +3 consistent with the NH₂-terminal sequence of human CD4 (11, 16, 29). In addition, we have remumbered the sequence of our recombinant sCD4 beginning with Lys⁴-Lys⁵-Val⁶... to be consistent with the NH₂-terminal sequence of the mature, expressed protein identified in these studies by MS and by Edman sequence analysis in other studies (11, 29). Altogether, approximately 95% of the primary structure was confirmed in the present studies. Approximately 79% of the sequence was corroborated by FABMS of the 6-h tryptic digest of reduced and carboxymethylated (RCM) sCD4 and HPLC fractions derived therefrom (Fig. 3). Tryptic peptides derived from regions of sCD4 containing either Asn⁷¹ or Asn⁹⁰, the two potential Asn-linked glycosylation sites, were not detected until peptide N-glycosidase F was used to release carbohydrate from the glycoprotein (see below). All significant signals observed in these FABMS data could be assigned to the deduced sequence of the glycoprotein (solid underlines, Fig. 4). Additional coverage of the glycoprotein was obtained by FABMS of S. aureus V8 digested RCM-sCD4 and cyanogen bromide-cleaved RCM-sCD4 (Fig. 4, II—II, and O—O, respectively; only peptides yielding additional coverage or confirming of the COOH terminus are shown). The former digest yielded the NH₂-terminal tripeptide of MH₁ = 1401.7 (subscript 1 = chemical average mass, see "Experimental Methods"). Extended or NH₂-terminally modified forms of this peptide were not detected by MS or NH₂-terminal sequence analysis,¹ indicating that the expected NH₂-terminal sequence is the

¹ C.-T. Yuen, S. A. Carr, and T. Feizi, manuscript in preparation.
² Portions of this paper (including "Experimental Procedures," Figs. 1–3 and 6–10, and Footnote 7) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.
only one signal present in this preparation of the glycoprotein. Four peptide signals at m/z 999.6, 1442.7, 3073.5, and 4241.9 correspond to (M + H)+; they are calculated as monoisotopic values unless labeled with a subscript n in which case they are calculated as chemical average values.

The sequence of the COOH-terminal nonapeptide was established by tandem high performance mass spectrometry with a VG ZAB-SE 4F four-sector double focusing mass spectrometer. The signal at m/z 999.6 was mass selected from the complex digest or an HPLC fraction containing this peptide using the first double focusing mass spectrometer. The mass selected parent was fragmented by high energy (10 keV, laboratory frame of reference) collisions with helium in the collision region between the two mass spectrometers. Daughter ion spectra are obtained at the final collector at the magnetic (B) field/electric (E) field is maintained constant. The sequence Val-Leu-Pro-Thr-Trp-Ser-Thr-Pro-Val (where xLeu = Ile or Leu) could be defined by interpretation of the resulting daughter ion spectrum (Fig. 5). The interested reader is referred to Ref. 24 for a detailed discussion of the fragmentation processes observed for peptides by tandem high performance mass spectrometry. Several minor signals caused by our handling of the protein were also detected in these studies. Specifically, 515% of Met235, Met314, and Met342 (but not Met349) in sCD4 were confirmed that Val369 is the COOH terminus of our recombinant peptide (two dipeptides, one tripeptide, two tetrapeptides, and a single pentapeptide) and an amino acid. In general, small hydrophilic peptides have been identified by composition and Edman sequence analysis of early eluting HPLC fractions of the tryptic digest (data not shown).
Identification of Glycosylation Sites and Structural Classification of Carbohydrates at Specific Attachment Sites—The sites of attachment of Asn-linked oligosaccharides in sCD4 glycoprotein were determined by FABMS carbohydrate mapping (30). In this technique, peptides containing Asn-linked carbohydrate are detected by comparing the FAB mass spectra obtained before and after treatment of the glycoprotein with peptide:N-glycosidase F (PNGase F (31, 32)) which cleaves the ⍺-aspartylglycosylamine linkage of all known types of glycosylation site peptides determined by carbohydrate fingerprinting (34). Potential glycopeptides were identified by comparing the reversed phase HPLC profile of the tryptic digest of RCM-sCD4 with that of a sample of RCM-sCD4 that had been sequentially digested with PNGase F, then trypsin (Fig. 3). Peaks in the tryptic digest that disappeared or were greatly attenuated in the chromatogram of the sample digested with the glycosidase are likely to be glycopeptides with carbohydrate linked to Asn.

Four putative glycopeptide-containing fractions were preparatively isolated from the tryptic digest and analyzed by FABMS. Signals for intact glycopeptides were observed in each of these fractions by FABMS (for example, see Fig. 6), as indicated by high-mass signals separated by 146 and 291 Da, the in-chain masses of deoxyhexose and NeuAc. The glycosylation site peptides determined by carbohydrate mapping, above, the in-chain masses of the respective oligosaccharide moieties are obtained by subtracting the (M + H)⁺ of the peptides from the observed (M + H)⁺ of the glycopeptides; the values used for the (M + H)⁺ of the peptides are 1

peak heights of the signals corresponding to oxidized versus nonoxidized peptide during the carboxymethylation procedure. These side products gave rise to low abundance satellite signals 16 Da above the Met-containing tryptic peptides Leu27-Lys251, Leu213-Lys218, Glu300-Lys300, and Ala332-Lys330 (Fig. 4). Absence of these satellite signals in the FABMS data obtained on digests of native (not reduced and carboxymethylated) sCD4 (see below) demonstrated that these oxidation products are not present in the protein as purified.

FIG. 5. Tandem mass spectrum of the COOH-terminal tryptic nonapeptide of sCD4. The resolution of MS-1 was set to select only the monoisotopic (M + H)⁺ and MS-2 was set at R ~ 800. The parent was attenuated by ~75% with helium collision gas. See Ref. 24 and 26 for discussions of the fragmentation processes.

4 A. Jurewicz, unpublished observation.
Da less than the observed masses noted above since the oligosaccharide attachment site Asn has been converted to Asp. Possible carbohydrate compositions were then generated by computer using windows of 0-10 units each of hexose and N-acetylhexosamine and 0-5 units each of NeuAc and deoxyhexose (dHex). The only reasonable carbohydrate compositions for the glycopeptides present in HPLC fractions 32 and 45 are shown in Table I. The data indicate that both attachment sites in recombinant sCD4 (Asn\textsuperscript{Z71} and Asn\textsuperscript{3'2}) have nearly identical families of oligosaccharides attached with the following general composition: NeuAc\textsubscript{2+3}Gal\textsubscript{4}GlcNAc\textsubscript{1+2}Man\textsubscript{1+4} where \(m = 0-2\), and \(n = 0.1\). The relative intensities of the parent signals in the FABMS data suggest that the most abundant oligosaccharides have \(m = 1\) and \(n = 1\) at each site, but that the Asn\textsuperscript{300} site has proportionately more deoxyhexose (Table I).

The compositions of the oligosaccharides at each of the specific attachment sites were further verified by permethylation of the PNGase F-treated HPLC-derived glycopeptide fractions and FABMS of the extracted derivatized carbohydrates (for example, see Fig. 7). The molecular species observed (Table I) confirm the general compositions outlined above; higher molecular weight carbohydrates were not detected at either of the two attachment sites in these MS studies. Several structurally informative fragment ions were observed in these FAB mass spectra. Loss of the reducing terminal HO-HexNAc-OCH\(_3\) or HO-HexNAc-dHex-OCH\(_3\) from (M + H)* at \(m/z\) 2410 or 2584, respectively, yields a nonreducing terminal oxonium ion (A-type cleavage (36)) at \(m/z\) 2133 (Fig. 7). Absence of a signal corresponding to the loss of HO-HexNAc-OCH\(_3\) from the nonreducing terminus of an antenna lacking a NeuAc endcap. Peaks at 376 and 344 are due to NeuAc. Peaks indicating the presence of a bisecting HexNAc or of polylactosamine-type structures were not detected.

Methylation analysis of the permethylated oligosaccharides from each attachment site gave linkage types indicative of biantennary complex carbohydrates (Table II). Together with compositions provided by the FABMS analyses of the underivatized glycopeptides and permethylated oligosaccharides from each attachment site, we conclude that the major carbohydrate structures at both Asn\textsuperscript{271} and Asn\textsuperscript{300} are biantennary complex sugars heterogeneous in NeuAc and Fuc content:

NeuAc\textsubscript{2+3}Gal\textsubscript{4}GlcNAc\textsubscript{1+2}Man\textsubscript{1+4} where \(m + n = 0-2\), and \(x = 0.1\). The NeuAc is linked entirely 2,3 to Gal, as 2,6-linked Gal was not detected. Trace levels of 1,2,4-linked Man and terminal Man suggest the possible presence of minor amounts of triantennary or higher structures and oligomannose structures, respectively. Our findings conflict with an earlier report that suggested, based on gly-
binant sCD4 expressed in CHO cells have triantennary carbohydrate has recently been shown to be the major (95%) because -S-S- bridges of disulfide-linked peptides are prone to potentially digested with trypsin and s. peptides. Reduced forms of the peptides are also often observed between difference observed reduced peptide and disulfide-bonded peptides. Fractionation of PNGase F-released oligosaccharides using high performance anion-exchange chromatography (see "Experimental Methods") also indicated that each site has minor experimental Methods") also indicated that each site has minor amounts of carbohydrates bearing three sialic acid groups, consistent with the presence of trace amounts of triantennary structures (region S3, Fig. 10). The molecular weights and fragmentation observed in the FAB mass spectra of the major components in the neutral, mono-, and disialylated fractions (Fig. 10) are consistent with the structures outlined above. Furthermore, the FABMS analyses demonstrated that biantennary oligosaccharides containing fucose attached to the reducing-end GlcNAc elute earlier than their nonfucosylated analogs. This observation is consistent with the reported order of elution using high performance anion-exchange chromatography of fucosylated versus nonfucosylated oligosaccharides in which the fucose is linked α-1,3 to GlcNAc in the antennae (41).

**Location of Disulfide Bonds**—The FABMS peptide mapping approach has also been used to identify the locations of disulfide bonds in sCD4. The method (Fig. 2B) involves cleavage of the protein under conditions known to minimize disulfide reduction and reshuffling (37, 38). In the absence of free thiols, procedures involving acidic or mildly basic conditions such as trypsin digestion can be used with minimal concern for disulfide rearrangement. The goal is to obtain disulfide-linked peptides each containing only a single -S-S- bridge. The FAB mass spectra of these mixtures exhibit (M + H)⁺ for intact, inter-, and intramolecularly disulfide-bonded peptides. Reduced forms of the peptides are also often observed in these spectra, even in the absence of added base, because -S-S- bridges of disulfide-linked peptides are prone to reduce in the liquid matrix under FABMS conditions. If signals for the constituent thiol peptides are not detected, the sample is reduced in the dithiothreitol/dithioerythritol or thioglycerol matrix by addition of triethylamine or ammonia hydroxide. Comparison of the spectra prior to and after incubation with dithiothreitol at basic pH (Fig. 9). To check for the presence of peptides in other disulfide-bonded arrangements, the trypsin/V8 digest was analyzed by reversed phase HPLC and compared with the HPLC of the same digest after incubating for two hours with dithiothreitol at basic pH (Fig. 8). Regions of the HPLC map of the nonreduced protein digest that disappeared after reduction were preparatively fractionated by HPLC and analyzed by FABMS. The individual fractions were also reduced on the probe to determine the masses of the constituent peptides. The data (Table III) indicate the following arrangements of disulfides: Cys⁴₁₅, Cys⁴₅₆, Cys⁴₇₄₀, Cys⁴₅₆, and Cys⁴₅₆₋₁₅ (Fig. 4). This arrangement is identical to that expected from previous studies of intrachain disulfide bonding in T4 antigens derived from sheep and mouse (17). No signals corresponding to disulfide-bonded peptides were detected in the mass range 800–4500 Da for fractions 52–54 (Fig. 8). These fractions most likely consist of large, incompletely digested disulfide-bonded peptides with molecular weights beyond the mass range analyzed.

The work described here illustrates the unique strengths of mass spectrometry and tandem mass spectrometry for sequence analysis of recombinant proteins and for characterization of posttranslational modifications such as disulfide bonds and carbohydrates. The methodology employed is particularly useful for rapid characterization of the class (hybrid, oligomannose, complex) and branching type (biantennary, triantennary, etc.) of the carbohydrates at specific sites in glycoproteins and can be used to rapidly (on the order of a few weeks) compare and define the structural class of carbohydrates on the same recombinant glycoprotein expressed in different cell lines or under different cell growth and harvesting conditions. Although cloning and sequencing of the gene coding for virtually any desired protein can now be accomplished with tremendous speed and efficiency, structural characterization of the recombinant protein product is often a relatively slow step. Amino-terminal and carboxyl-terminal sequencing of the protein product are required to establish identity, but these are only starting points, and internal sequences must be verified to be sure of the fidelity of the sequence. Automated Edman sequence analysis is not sufficient, since many native or induced protein modifications cannot be identified by this technique. Clearly, fast, sensitive, and reliable procedures are necessary to bridge the analytical gap between protein chemistry and molecular genetics. Mass spectrometry and tandem mass spectrometry when used in conjunction with conventional chemical and biochemical approaches can help to fill this gap, particularly when each approach is used so as to take advantage of its unique strengths.

**Acknowledgements**—We wish to thank V. Dodia, R. Inacker, A.

---

**Table III**

| Disulfide-bonded peptides | (M + H)⁺ Signal | Component peptide (M + H)⁺ |
|---------------------------|----------------|--------------------------|
| 14–21(Cys⁴₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋˓→<sup>+</sup> | 2037.2 | 2037.2 |
| 14–22(Cys⁴₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋˓→<sup>+</sup> | 2165.5 | 2165.5 |
| 120–131(Cys⁴₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋˓→<sup>+</sup> | 3022.1 | 3022.1 |
| 300–304(Cys⁴₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋₋₋₄₋˓→<sup>+</sup> | 2428.8 | 2428.8 |

* Signals observed after reduction with diethiothreitol; (M + Na)⁺ also observed in all fractions.
* Observable (M + 2K - H)⁺.
* Observable (M + H)⁺.
* Peptide not observed; assignment made by mass difference observed reduced peptide and disulfide-bonded peptides.
* No signals corresponding to disulfide-bonded peptides were detected in the mass range 800–4500 Da.

---

<sup>a</sup> S. A. Carr and J. R. Barr, unpublished observations.
Jurewicz, M. Mentzer, G. Roberts, and M. Vettese for technical assistance.

REFERENCES

1. Gay, D., Maddon, P., Sekaly, R., Talle, M. A., Godfrey, M., Long, E., Goldstein, G., Chess, L., Axel, R., Kappler, J., and Marrack, P. (1987) Nature 328, 626–629
2. Doyle, C., and Strominger, J. L. (1987) Nature 330, 256–259
3. Sattentau, Q., and Weiss, R. A. (1988) Cell 52, 631–633
4. 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
5. Klatzman, D., Champagne, E., Chamaret, S., Gruet, J., Guettard, D., Herceg, T., Gluckman, J.-C., and Montagnier, L. (1984) Nature 312, 767–768
6. McDougal, J. S., Kennedy, M. S., Sligh, J. M., Cort, S. P., Mawie, A., and Nicholson, J. K. (1988) Science 231, 382–385
7. Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., and Axel, R. (1986) Cell 47, 333–348
8. Smith, D. H., Byrn, R. A., Marsters, S. A., Gregory, T., Groopman, J. E., and Capon, D. J. (1987) Science 238, 1704–1707
9. Deen, K. C., McDougal, J. S., Inacker, R., Folela-Wasserman, G., Arthos, J., Rosenberg, J., Maddon, P. J., Axel, R., and Sweet, R. W. (1988) Nature 331, 82–84
10. Fisher, R. A., Bertonis, J. M., Meier, W., Johnson, V. A., Costopoulos, D. S., Liu, T., Tizard, R., Walker, B. D., Hirsch, M. S., Schooley, R. T., and Flavell, R. A. (1988) Nature 331, 76–78
11. Hussey, R. E., Richardson, N. E., Kowalski, M., Brown, N. R., Chang, H. C., Siliciano, R. F., Dorfman, T., Walker, B., Sodroski, J., and Reinhart, E. L. (1988) Nature 313, 78–81
12. Traunecker, A., Lüke, W., and Karjalainen, R. (1988) Nature 331, 84–86
13. Berger, E., Fuerst, T. R., and Moss, B. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2357–2361
14. Clapham, P. R., Weber, J. N., Whitby, D., McIntosh, K., Dalgleish, A. G., Maddon, P. J., Deen, K. C., Sweet, R. W., and Weiss, R. A. (1989) Nature 337, 368–370
15. Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L., and Axel, R. (1985) Cell 42, 83–104
16. Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J., and Williams, A. F. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1649–1653
17. Clausson, B. J., Tsagaratos, J., McKenzie, I. P., and Walker, I. D. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 4499–4503
18. Bedinger, P., Moriarty, A., von Borstel, II, R. C., Donovan, N. J., Steimer, K. S., and Littman, D. R. (1988) Nature 334, 162–165
19. Richardson, N. E., Brown, N. R., Hussey, R. E., Vaid, A., Matthews, T. J., Bolognesi, D. P., and Reinhart, E. L. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6102–6106
20. Peterson, A., and Seed, B. (1988) Cell 54, 65–72
21. Chao, B. H., Costopoulos, D. S., Curiel, T., Bertonis, J. M., Chahlohim, P., Williams, C., Schooley, R. T., Rosa, J. J., Fisher, R. A., and Maragamore, J. M. (1989) J. Biol. Chem. 264, 5812–5817
22. Arthos, J., Deen, K. C., Chaiken, M. A., Formwald, J. A., Sattentau, Q. J., Clapham, P. R., Weiss, R. A., McDougal, J. S., Pietropaolo, C., Maddon, P. J., Truneh, A., Axel, R., and Sweet, R. W. (1989) Cell 57, 469–481
23. Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 831–864
24. Biemann, K. (1988) Biomed. Environ. Mass Spectrom. 16, 99–111
25. Carr, S. A., Hemling, M. E., and Roberts, G. D. (1988) in Macromolecular Sequencing and Synthesis: Selected Methods and Applications (Schlesinger, D. H., ed) pp. 85–99, Alan R. Liss, New York
26. Anderegg, R. J., Carr, S. A., Huang, I. Y., Hüppakka, R. A., Chang, C., and Liao, S. (1988) Biochemistry 27, 4214–4221
27. DeWolfe, W. E., Jr., Carr, S. A., Varrichio, A., Goodhart, P. J., Mentzer, M. A., Roberts, G. D., Southan, C., Dolle, R. E., and Kruse, L. I. (1988) Biochemistry 27, 9093–9101
28. Carr, S. A. (1990) Adv. Drug Delivery Rev., in press
29. Deen, K. C., Arthos, J., McDougal, J. S., Chaiken, M., Folema-Wasserman, G., Maddon, P. J., Axel, R., and Sweet, R. W. (1989) UCLA Symp. Mol. Cell. Biol. New Ser. 90, 22–34
30. Carr, S. A., and Roberts, G. D. (1986) Anal. Biochem. 157, 396–406
31. Plummer, T. H., Jr., Elder, J. H., Alexander, S., Phelan, A. W., and Tarentino, A. L. (1984) J. Biol. Chem. 259, 10700–10704
32. Tarentino, G. L., Gomez, C. M., and Plummer, T. H., Jr. (1985) Biochemistry 24, 4665–4671
33. Hirani, S., Bernasconi, R. J., and Rasmussen, J. R. (1987) Anal. Biochem. 162, 485–492
34. Carr, S. A., Roberts, G. D., Jurewicz, A., and Frederick, B. (1988) Biochimie 70, 1445–1454
35. Conradt, H. S., Egge, H., Peter-Katalinic, J., Reiser, W., Siklosi, T., and Schaper, K. (1987) J. Biol. Chem. 262, 14600–14605
36. Dell, A. (1987) Adv. Carbohydr. Chem. Biochem. 45, 19–72
37. Morris, H. R., and Pucci, P. (1985) Biochem. Biophys. Res. Commun. 126, 1122–1126
38. Yazdanparast, R., Andrews, P. C., Smith, D. L., and Dixon, J. E. (1987) J. Biol. Chem. 262, 2507–2513
39. Pohl, G., Kallstrom, M., Bergedorf, N., Wallen, P., and Jornvall, H. (1984) Biochemistry 23, 3701–3707
40. Finne, J., Krustus, T., and Rauvala, H. (1980) Carbohydr. Res. 80, 336–339
41. Hardy, M. R., and Townsend, R. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 3289–3293
42. Fenouillet, E., Clerget-Raslain, B., Gluckman, J. C., Guertard, D., Montagnier, L., and Bahrnou, E. (1989) J. Exp. Med. 169, 807–822
43. Ciucanu, I., and Kerek, F. (1984) Carbohydr. Res. 131, 209–217
44. Chaplin, M. F. (1982) Anal. Biochem. 123, 336–341
isolation and purification of CD4 receptor: Ten to 20-fold ~70 °C freeze concentration of serum-free conditioned medium (b) containing soluble CD4 receptor (sCD4) were thawed at 4 °C, and diluted 3-fold with 80 ml 2-mercaptoethanol (ME), Sigma, pH 8.0 buffer. Purify was assessed at this and later stages using SDS-PAGE with silver staining techniques (Fig. 1). The eluted solution was treated with 0.1 M PIPES (p-amino-benzylphenethylamine hydrochloride, Calbiochem-Behring) and loaded onto a column of S-Sepharose fast flow (Pharmacia) which was previously equilibrated with 50 mM sodium phosphate, pH 6.8 buffer. The column was washed with equilibration buffer, and eluted with a linear gradient of 0.5 M NaCl in 50 mM sodium phosphate, pH 6.8 buffer containing sCD4 were pooled, treated with 0.1 M PIPES, pH 7.0 and dialyzed against 50 mM sodium phosphate, pH 6.8 buffer.

The eluate was treated with 0.1 M thioglycolic acid (Sigma), adjusted to pH 5.0, and loaded onto a column of Q-Sepharose fast flow (Pharmacia) that was previously equilibrated in 50 mM Tris-HCl, pH 8.0 buffer. The column was washed with equilibration buffer and eluted with the same buffer containing 1 M NaCl. Soluble CD4 receptor was recovered in the unbound solution which was immediately adjusted to pH 6.8. The Q-Sepharose product was concentrated in an Amicon stirred cell fitted with a YM-3 membrane and further purified by chromatography on a column of Supersose 12, pre grade, using 50 mM sodium phosphate, 0.1 M NaCl, pH 7 buffer as the mobile phase.

Cell-free Dyalysis: Removal of the native CD4 receptor was performed in an Aminco Centriprep-30 microconcentrator (30 Kd cutoff). One ml of a 4.65 mg/ml solution was placed in each of four plastic syringes, each sampler was then dialyzed for four times by adding 100 mM NaHCO3 solution to bring the total volume in each microconcentrator to 2 ml and spinning to concentrate the protein reis to 0.4 m in each vial.

Reduction and Carbohydrate-methylation: Reduction of the sCD4 protein was performed in 0.5 M Tri-HCl buffer, pH 7.0, with 0.5 M NaBH4 (50% aqueous solution) in a total volume of 1 ml. The solution was then dehydrated by being dialyzed against 0.5 M NaHCO3 solution for 1 h at 4 °C. The sample was diluted at 3.5 and 0.5 M NaHCO3, for 1 h at 4 °C.

Selective fragmentation of Proteins, Glycoproteins, and Peptides: For tryptic digests RCM-CD4 protein was dissolved in 250 μM ammonium bicarbonate buffer at pH 8.5. An aliquot of tryptic (TPCK-treated, Calbiochem-Behring, La More, Pennsylvania) was added to the same buffer solution to give an enzyme substrate ratio of 1:100. Tryptic digests were incubated at 37 °C for 16 h. The reaction was stopped by the addition of 0.2 M HCl and the samples were then lyophilized and redissolved in 50 mM sodium phosphate, pH 7.3 buffer.

Digestion with Glycosidases: An aliquot of the proteolytic digest of the glycoprotein is dissolved with the aid of 6N HCl (1 ml, 1:100 dilution) and after heating at 100 °C for 1 h, 0.5% SDS (sodium dodecyl sulfate) solution was added to bring the concentration to 0.5% with dithiothreitol (DTT) and 0.5% with diisourea (I). The samples were then digested with pony (Sigma, St. Louis, Missouri) and analyzed by SDS-PAGE.

HPLC: Peptide mixtures derived from digests of 0.3 to 20 nanograms of protein were analyzed and fractionated by high-performance liquid phase chromatography (HPLC) on a Beckman System Gold equipped with a model 570A variable wavelength detector. An Abbott 100 M adsorbent was used to control the system and to obtain a complete trace of the data. The fractionation was done on a 10x10 cm reversed-phase column using 0.1% aqueous TFA and 0.1% aqueous TFA. The samples were collected by lyophilization and then digested to the desired pH.

Acetylation and Permethylation of Oligosaccharides: The dried PF-HPLC treated HPLC fractions in 1 ml methanol/conc. HCl were acetylated and/or permethylated as previously described (34) except that sodium hydroxide in dimethylformamide was used as the acetylation reagent (35).

Fractionation of PF-HPLC Released Oligosaccharides: A high performance anion exchange (DPA) chromatofocusing system from Dionex was used for the fractionation of PF-HPLC released oligosaccharides (46). The Dionex BIOAC system consists of a gradient pump, a pulsed amperometric detector (PAD) with a gold electrode and a computerized interface with an AI/AL controller. A Dionex DM column was coupled to a Dionex 1000 data collector and controlled by a Dionex Chromatography Software. PF-HPLC separated fractions were subjected to 42 cycle (50 mM NaHCO3, Na2CO3, NaCl, 0.1 NaCl on TFA). At the beginning of each run, gradient was set to pH 7.0 for 30 min and then stepwise to pH 5.0.

Reduction of Disulfide-linked Peptides: For mapping by HPLC, 2 nmol of PF-HPLC, Trypsin, and 3. aureus V8 digested CD4 were digested at 80 μl of sodium bicarbonate buffer, pH 6.5 was incubated with 18 μl of 15 μg/ml dithiothreitol as described (32, 34) except for 2 h at 37 °C. Putative disulfide-linked peptide fractions were also reduced on the HPLC probe by mixing a 1:50 of the peptide fraction with DTFF/TE (1:1) to which was added 1.5 μl of triethanolamine or a 4% aqueous solution of sodium hydroxide and allowing to react for 10 min.

Fast Atom Bombardment Mass Spectrometry: FAB mass spectra were obtained using (a) the first double focusing Boesl-1) of a VG ZAB SE-4F tandem magnetic deflection mass spectrometer equipped with a standard FAB source on source and high voltage Cs on gun on (b) a VG ZAB-HF equipped with a standard FAB source on source and high voltage Cs on gun on. The high voltage Cs on gun was operated at 65 kV and 45 kV, respectively. A 1:1 mixture of water and methanol was used as the data system to acquire isotope peaks. Approximately 1-3 of an enzymatic digest of digested protein in an aqueous solution containing 1:2 of 30% aqueous glycerol and 20% water was treated with 500 μl of glacial acetic acid in 10 mM NaOH, pH 8.0. Samples were then desalted and reconstituted in 10% methanol containing 0.1% TFA. The mass spectra were obtained using a low resolution (R = 800) FAB mass spectrometer (36). For samples above m/z 1000 in which the molecular ion cluster is unresolved, a mass value obtained in the corresponding chemical standard. Each sample is analyzed in separate experiments over two or more times and 100-150 scans were collected. Identification of the various components was made by computer (DMS) and a VG 7030 ESE equipped with a combination E Uses source (70, 100, 300) and a Waters Peltier 6000 GC equipped with a computerized data system.

Tandem Mass Spectrometry: FAB MS/MS experiments are conducted on a VG ZAB SE-4F or VG ZAB-HF. Tandem mass spectrometric conditions for FAB are the same as described above. The mass spectrum of m/z was selected to only monoisotopic (M+), the CDOM-terminal tryptic peptide, DAPG was produced by introducing into the ion collection cell located between MS1 and MS2. This MS1 transition was monitored by monitoring the pressure sufficient to reduce the selected parent ion beam by 75 to 80%. Daughter ion spectra were obtained at 23 nsec after the generated field ion pulse (exponential down, 25 nsec) at the entrance of MS2 such that the instrument was effectively operating in the linear mode. daughter ion spectra were obtained at 23 nsec after the generated field ion pulse (exponential down, 25 nsec) at the entrance of MS2. Daughter ion spectra were obtained at 23 nsec after the generated field ion pulse (exponential down, 25 nsec) at the entrance of MS2 such that the instrument was effectively operating in the linear mode.
Figure 3. (Top) HPLC trace of tryptic digested reduced and carboxymethylated (RCM) CD4 for comparison with (bottom) the HPLC trace of the tryptic digest of PNGase F treated RCM-CD4. Putative glycopeptide containing fractions are indicated (GP).

Figure 4. FAB mass spectra of ca. 1 nanomole each of glycopeptide-containing HPLC fraction 30 (top) and 42 (bottom) of the tryptic digest of RCM-CD4 (see Fig. 3). Molar mass are calculated as the chemical formula values.

Figure 5. FAB mass spectrum of ca. 200 picomoles of the permethylated carbohydrates released from the Gal¹⁴⁰ attachment site by PNGase F digestion of glycopeptide fraction 48 from the tryptic digest of RCM-CD4 (see Figs. 3, 4).
Structure of a Recombinant CD4 Receptor Glycoprotein

Figure 6. HPLC trace of PHGase F, trypsin, and Staph. V8 digested sT4 prior to (top) and after (bottom) reduction with DTT. The disulfide-linked peptide regions are indicated.

Figure 7. PAB mass spectra of the complex trypsin digest of native CD4 prior to (top) and after (bottom) addition of base to the sample on the probe (A); the spectra were monoisotopic. Addition of base to the reduction/marine deisulfurization of disulfide bonds (see text).

Figure 9. Separation of ca. 0.5 pmol each of PHGase F-released oligosaccharides from CD4 ApoF over ApoG5 sites using HPLC-PAD (Dionex). Oligosaccharides were eluted chromatographically on a 4.6 mm × 250 mm column in 0.1% trifluoroacetic acid 60% acetonitrile buffer at 1.0 ml/min. The eluate was monitored at 214 nm and at 252 nm for the tryptophan-containing peptides. The key indicates the regions of oligosaccharides eluting at 17.7 min (B) at 214 nm and 17.5 min at 252 nm (C) depending on their structure. The treated sample at 15.5 min is an artificial sample formed by incubation of an air bubble by the entry cap of the column.