Cell Line and Site Specific Comparative Analysis of the N-Linked Oligosaccharides on Human ICAM-1des454−532 by Electrospray Ionization Mass Spectrometry

James W. Bloom,* Melanie S. Madanat, and Manas K. Ray

Department of Process Sciences, Bayer Corp., Berkeley, California 94701

Received October 2, 1995; Revised Manuscript Received November 17, 1995[⊗]

ABSTRACT: Sialylated oligosaccharide structures were determined by the technique of electrospray ionization mass spectroscopy at seven of eight N-linked glycosylation sites of recombinant human ICAM-1des454−532 [tICAM(453)] purified from the tissue culture fluid of Chinese hamster ovary, human embryonic kidney, and mouse myeloma cell lines. The number of structures at each site depended on the cell line and ranged from 8 to 34. *N*-Glycolylneuraminic acid, a human oncofetal antigen, was found at all sites of all three cell line derived forms of tICAM(453). Tetraantennary complex structures containing one and/or two galactose-β1,4 *N*-acetylglucosamine repeats, characteristic of membrane bound proteins, were found on soluble tICAM(453) primarily at Asn-379. Asn-379, located between the D4 and D5 domains, is believed to be located close to the membrane surface in membrane bound ICAM-1. It has been proposed that the extent of N-linked glycosylation at Asn-240 and Asn-269 in the third domain of ICAM-1 may regulate the binding avidity of ICAM-1 to Mac-1 [Diamond, M. S., Staunton, D. E., Marlin, S. D., & Springer, T. A. (1991) *Cell 65*, 961−971]. In the present study the tICAM(453) Asn-269 site was found to contain predominantly one oligosaccharide structure that is conserved in all three cell lines. On the other hand, the Asn-240 site was found to contain cell line dependent oligosaccharide structural heterogeneity particularly in the degree of sialylation.

Intercellular adhesion molecule-1 (ICAM-1) is a ligand for the leukocyte integrins, lymphocyte function-associated antigen-1 (LFA-1) (Rothlein et al., 1986; Marlin & Springer, 1987) and Mac-1 (Smith et al., 1989; Diamond et al., 1990). ICAM-1 has also been shown to be an endothelial cell adhesion receptor for *Plasmodium falciparum*-infected erythrocytes (Berendt et al., 1989). In addition, ICAM-1 is subverted as a cellular receptor by a large family of human rhinoviruses that are the major causative agents of the common cold (Greve et al., 1989; Staunton et al., 1989; Tomassini et al., 1989a).

Amino acid sequence homology suggests that ICAM-1 is an integral membrane protein with an extracellular region of 453 residues containing five immunoglobulin-like domains (D1−D5, numbered sequentially from the amino end), a transmembrane portion, and a small cytoplasmic domain (Simmons et al., 1988; Staunton et al., 1988). The primary binding sites for LFA-1, human rhinoviruses and *P. falciparum*-infected erythrocytes are located in domain 1 of ICAM-1. Domain 2 is believed to play an essential role in the conformation of domain 1 and may play an additional role in some or all of these binding activities (Staunton et al., 1988; Staunton et al., 1990; McClelland et al., 1991; Lineberger et al., 1990; Fegister et al., 1991; Kolatkar et al., 1992; Olson et al., 1993; Berendt et al., 1992; Ockenhouse et al., 1992). Mac-1, on the other hand, has been shown to bind to domain 3 of ICAM-1 (Diamond et al., 1991). Two soluble forms of ICAM-1, one corresponding to the entire extracellular portion (ICAM-1des454−532) (Marlin et al., 1990; Greve et al., 1991) and one corresponding to the two N-terminal immunoglobulin-like domains (ICAM-1des186−532) (Greve et al., 1991), have been shown to prevent rhinovirus infection.

Eight potential N-glycosylation sites (Asn-Xaa-Ser/Thr, where "Xaa" is any amino acid except proline) are evident in the amino acid sequence of ICAM-1 (Staunton et al., 1988) in domains D2, D3, and D4. Enzyme digestion was used to show that 30% of the molecular mass of ICAM-1 is composed of complex-type oligosaccharides, and in those studies seven N-linked glycosylation sites were detected but not identified (Tomassini et al., 1989b). The potential role of glycosylation of ICAM-1 in human rhinovirus binding has been explored by mutating each of the four sites of N-linked glycosylation in domain 2. The partially deglycosylated receptor bound virus indistinguishably from the wild-type molecule, suggesting no direct role of N-linked glycosylation in HRV binding (McClelland et al., 1991). In contrast to these studies, Lineberger et al. (1990) concluded that glycosylation was involved in virus binding on the basis of the lack of virus binding to ICAM-1 made in the presence of tunicamycin. A possible explanation for these apparently conflicting results is that glycosylation of domain 2 is required for the proper folding and transport of ICAM-1 to the cell surface (McClelland et al., 1991).

The present N-linked glycosylation structural characterization study was undertaken to better understand the role that oligosaccharides play in the biological activity of ICAM-1. In order to distinguish between control of glycosylation by the polypeptide backbone or expression of selective glycosylation capacity by the cell synthesizing that polypeptide, we compared tICAM(453)[1] produced by three different cell lines: Chinese hamster ovary, human embryonic kidney, and a murine myeloma.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1996.

0006-2960/96/0435-1856$12.00/0

MATERIALS AND METHODS

Reagents. Sequencing grade trypsin was obtained from Promega (Madison, WI). *N*-Glycosidase F (PNGase), recombinant, was obtained from Boehringer Mannheim (Indianapolis, IN).

Three tICAM(453)-secreting cell lines were used: Chinese hamster ovary (CHO), human embryonic kidney (HEK), and a mouse myeloma line (NS0). Recombinant tICAM(453) was purified from tissue culture fluid by monoclonal antibody affinity chromatography. The tissue culture fluid was passed over the affinity column equilibrated with 20 mM imidazole, 0.5 M NaCl, pH 6.8 buffer. After sample application, the column was washed with 20 mM imidazole, 0.5 M NaCl, 1.0 M $MgCl_2$, pH 6.8 buffer to remove nonspecifically bound proteins. The tICAM(453) was eluted with 50 mM diethanolamine, pH 11.5, collected and neutralized to pH 7.0−7.5 by the addition of 1.0 M HCl. The pool was concentrated using an Amicon YM 30 membrane (Amicon, Inc., Beverly, MA) and then dialyzed against (two) 2 L volumes of phosphate-buffered saline, pH 7.2 overnight at 4 °C. The dialyzed tICAM(453) solution was then sterile filtered and frozen at −70 °C. The purified protein appeared as a diffuse single band by the technique of SDS−PAGE with Fast Stain (Zoion Biotech, Newton, MA).

Enzyme Digestion. A 100 μL aliquot of a solution of approximately 20 mg of tICAM(453)/mL was added to 900 μL of 6 M guanidine hydrochloride−0.05 M Tris, pH 8.6. Dithiothreitol (25 μL of a 2 M solution) was added, and the sample was heated at 100 °C for 1 h. Iodoacetic acid (60 μL of a 2 M solution) was added, and the sample was incubated in the dark for 30 min. Samples were then dialyzed extensively against 0.1 M NH_4HCO_3. The samples were digested by adding 10 μg of trypsin to each 2 mg sample of tICAM(453) and incubating at 37 °C for 18 h. The reaction was stopped by heating at 100 °C for 5 min. In some instances, the sample was then split into two 500 μL aliquots. One aliquot was designated as a control and 765 μL of 0.08 M sodium phosphate, 1 mM EDTA, 0.15 M NaCl, 50% glycerol, pH 7 buffer was added. For removal of N-linked oligosaccharides, 15 units of PNGase in 75 μL of solution was added to the other aliquot followed by 690 μL of the above buffer. The samples were then incubated for 18 h at 37 °C. The reaction was stopped by heating at 100 °C for 5 min.

High-Performance Liquid Chromatography (HPLC). Trypsin-digested samples of tICAM(453) (approximately 400 μg per injection) were analyzed by reversed-phase HPLC using a Hewlett Packard 1090 Series M system with an integrated diode-array detector. Hewlett Packard Chemstation Rev. A. 02.00 software was used for data analysis. A Vydac 218TP52 (C18, 5 μm, 250 mm × 2.1 mm i.d.) column was used with the following chromatographic conditions: flow rate, 0.15 mL/min; oven temperature, 40 °C; detection, 210 nm; solvent A, 0.1% TFA; solvent B, 0.1% TFA−60% acetonitrile; gradient, 0 min 10% B, 5 min 10% B, 140 min 65% B, 145 min 100% B at 0.30 mL/min, 150 min 10% B, 165 min 10% B. Peptide peaks to be further analyzed were collected by hand and stored frozen at −70 °C.

Electrospray Ionization Mass Spectrometry (ESI-MS). The HP1090 HPLC column effluent was mixed with 20% propionic acid/isopropyl alcohol (0.15 mL/min from an HP 1050 HPLC) and introduced directly into a Hewlett Packard model 59987A API-Electrospray LC/MS Interface with a model 5989B MS Engine (Goodley et al., 1994). The quadrupole was scanned from 450 to 2000 *m*/*z* in 2 s in the positive ion mode. Conditions used for data acquisition were as follows: integration time, 100 μs; samples, 4; step size, 0.1 amu; mass response filter, Gaussian FWHM 0.300 amu; and time response filter, Gaussian FWHM 0.050 min. The instrument was tuned with a mixture of cytochrome *C* and gramicidin S at a peak width of approximately 1.2 amu. Data analysis was accomplished with an HP ChemStation data system with HP G1047A LC/MS Software. Peptide molecular masses were calculated from the amino acid sequence using MacBioSpec software (Perkin Elmer).

Sequence Analysis. Automated Edman degradation was performed with an Applied Biosystems model 473A pulsed liquid protein sequencer system, and the data was analyzed with the model 610A program.

ESI-MS Oligosaccharide Structural Analysis. The tICAM-(453) N-linked oligosaccharide site specific structures were determined by ESI-MS analysis of reversed phase HPLC trypsin digest peptide maps. The observed ESI-MS spectra *m*/*z* peak values at each of the N-linked sites were entered into a spreadsheet with several likely Z values (in the range 2−5), and the masses of the putative glycopeptides were calculated using the formula: $[(m/z) \times z] - z$. The known calculated mass for the peptide component of each trypsin digest peptide containing an N-linked oligosaccharide was then subtracted. The resulting calculated mass was compared to a database of masses of known oligosaccharide structures. The oligosaccharide structures in the databank were variations of structures shown by other methods to be present on tICAM(453) obtained from the three different cell lines (data not presented in this report). The methodology used to arrive at these structures consisted of high-performance anion exchange chromatography with pulsed amperometric detection analysis and exoglycosidase sequencing of oligosaccharides removed from the tICAM(453) by hydrazinolysis. The majority of the ESI-MS experimentally determined site specific oligosaccharide masses agreed within 1 amu of the values in the databank. A difference greater than 3 amu was considered not to be a match. The abundance of each structure was calculated from the normalized peak heights in the *m*/*z* spectrum.

RESULTS

N-Glycosylation Site Identification. Reversed phase HPLC trypsin digest peptide maps of CHO cell line derived tICAM-(453) with and without PNGase treatment were used to determine the number and location of peptides containing N-linked oligosaccharides. A comparison of the two peptide maps (not shown) showed that several peaks present in the trypsin only treated tICAM(453) were missing in the PNGase treated material. These results suggested that the "missing" peaks contained peptides with N-linked oligosaccharides and that removal of the oligosaccharides with PNGase had

[1] Abbreviations: tICAM(453), ICAM-1des454−532; CHO, Chinese hamster ovary; HEK, human embryonic kidney; PNGase, *N*-glycosidase F; ESI-MS, electrospray ionization mass spectrometry; NeuAc, *N*-acetylneuraminic acid; NeuGc, *N*-glycolylneuraminic acid; *N*-acetyllactosamine, galactose-β1,4 *N*-acetylglucosamine; Man, mannose; Fuc, fucose; Gal, galactose; GlcNAc, *N*-acetylglucosamine.

Table 1: Identification of Seven N-Glycosylated Peptides in a Reversed Phase HPLC Peptide Map of a Trypsin Digest of CHO Cell Line Derived tICAM(453)

glycosylation site	t_R[a]	Edman sequence[b]	residues	calcd. mass
Asn-103	35.1	(K)_LTLR	103−107	615.7
Asn-118	78.0	(R)Ā_LTVVLLR	117−125	998.2
Asn-175	120.0	(R)TĒLDLRPQGLELFE_...(R)	161−198	4267.9
Asn-240	54.2	(R)LNPTVTYG_DSFSAK̄	232−246	1613.8
Asn-269	114.0	(R)LT_AVILG_Q̄SQETL...(K)	261−304	4778.5[c]
Asn-358	88.0	(R)SFS̄_SATLĒVAGQLI...(R)	341−361	2348.1[d]
Asn-379	102.3	(R)D_PG̅_WTWPENSQQT...(K)	375−403	3444.7[e]
	104.0			

[a] Retention time in minutes. [b] The peptides containing Asn-175, Asn-269, Asn-358, and Asn-379 were not completely sequenced. [c] Calculated average mass with carboxymethylated Cys at residue 263. [d] Calculated average mass with carboxymethylated Cys at residue 344. [e] Calculated average mass with carboxymethylated Cys at residue 376.

resulted in a shift in the elution position of the now "naked" peptides. To confirm this hypothesis, the peptides in question were collected and subjected to Edman sequencing. Seven N-linked glycosylation sites were determined by this technique: Asn-103, -118, -175, -240, -269, -358, and -379. The peptides containing Asn-103, -118, and -240 were completely sequenced, and the others were only partially sequenced (Table 1). These data provide indirect evidence of glycosylation at these positions since glycosylated Asn does not appear in the amino acid chromatogram generated by the Edman sequencer. The peptide containing Asn-358 was not sequenced through this amino acid. The same Asn-379 containing peptide sequence was observed in two peaks collected from the tICAM(453) trypsin digest, suggesting that they contained different glycoforms of the peptide with differing hydrophobicities. A peptide containing Asn-156 could not be located in the peptide map. Subsequent to these studies, a reversed phase HPLC trypsin digest peptide map of CHO cell line derived tICAM(453) was generated using a different gradient: 0 min 0% B at 0.30 mL/min, 5 min 10% B at 0.15 mL/min, 150 min 60% B, 150.1 min 60% B at 0.30 mL/min, 155 min 100% B at 0.30 mL/min, 157 min 0% B, 165 min 0% B. Under these chromatography conditions a glycosylated peptide containing Asn-156 (residues 151−160) was eluted at a retention time of 30.0−31.3 min. The N-linked oligosaccharides at Asn-156 were not characterized.

N-Glycosylation Site Oligosaccharide Analysis by ESI-MS. A reversed phase HPLC peptide map of a trypsin digest of CHO cell derived tICAM(453) with on-line analysis by ESI-MS is shown in Figure 1 [peptide maps of HEK and myeloma cell line derived tICAM(453) were also determined but are not shown]. The 210 nm UV and total ion current chromatograms are compared. As might be expected the two different techniques for peptide detection resulted in quite different chromatograms. The seven N-linked glycosylation site peptide-containing peaks are identified in the UV chromatograms. In several instances no apparent corresponding peak was observed in the total ion current chromatogram; a high degree of carbohydrate heterogeneity on the peptides contained within these peaks could account for this low signal to noise phenomenon. The peptide map chromatograms for tICAM(453) from the three different cell lines were similar and differed primarily only in the shape and number of the glycopeptide-containing peaks. These differences were probably the result of peptide glycoforms with differing hydrophobicities eluting at different retention times.

The oligosaccharide structures at each of the seven N-linked sites were determined by analyzing the ESI-MS *m*/*z* spectra obtained from the total ion current chromatograms. As an example, the ESI-MS *m*/*z* spectra of the

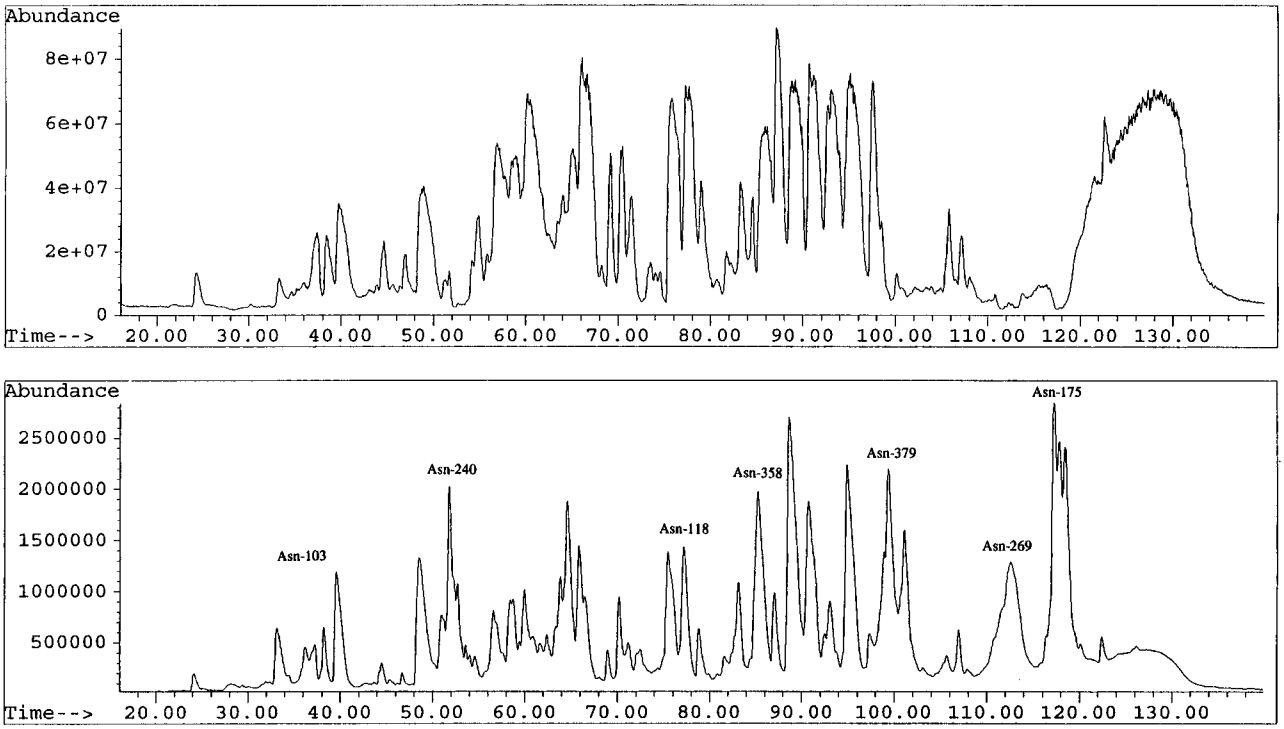


FIGURE 1: Reversed phase HPLC peptide map of a trypsin digest of CHO cell line derived tICAM(453) with UV monitoring at 210 nm (lower trace) and total ion current as measured by ESI-MS. The UV chromatogram *Y*-axis units (labeled abundance) were software generated, and 2 000 000 units equal approximately 1000 mAU at 210 nm. The labeled peaks contain the N-linked oligosaccharide binding site glycopeptides.

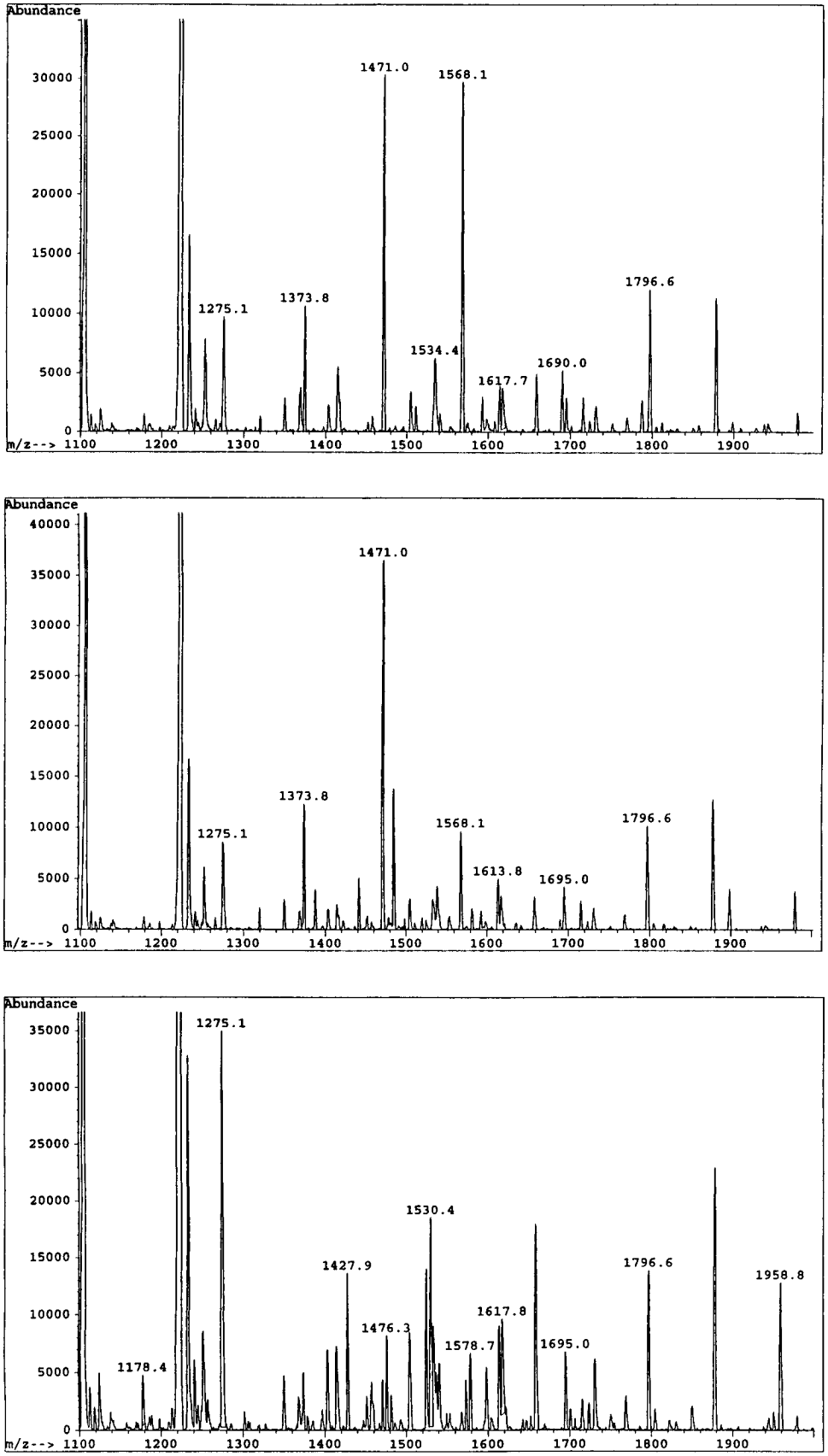
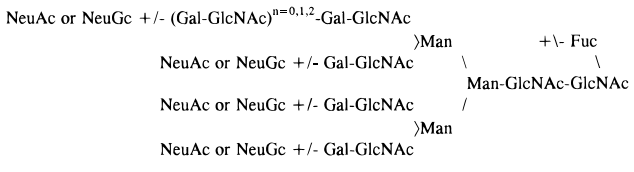


FIGURE 2: ESI-MS *m*/*z* spectra of the Asn-358 N-linked oligosaccharide binding site glycopeptides from CHO (top panel), HEK (middle panel) and mouse myeloma (bottom panel) cell line derived tICAM(453). A few of the more prominent spectral peaks that were identified with an oligosaccharide structure are labeled with their *m*/*z* values.

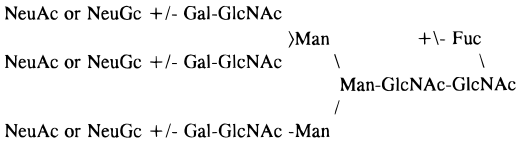
tICAM(453) Asn-358 N-linked oligosaccharide binding site trypsin digest glycopeptides obtained from the three cell lines are shown in Figure 2. A few of the more prominent spectral peaks in Figure 2 that were identified with an oligosaccharide structure are labeled with their *m*/*z* values. A number of spectral peaks were not identified and may represent either other peptides that eluted from the reversed phase HPLC column along with the glycopeptides or unknown oligosac-

Table 2: Schematic Diagrams of the General Types of Oligosaccharide Structures Found on tICAM(453)

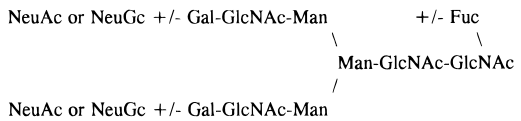
Tetraantennary Complex Structure with or without *N*-Acetyllactosamine



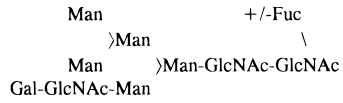
Triantennary Complex Structure



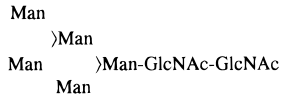
Biantennary Complex Structure



Hybrid Structure



Oligomannosyl Structure (Man5)



charide structures N-linked to the Asn-358 peptide. The proposed oligosaccharide structures (schematically illustrated in Table 2) determined by ESI-MS for the Asn-358 site are listed in order of abundance in Tables 3−5. For the purpose of conciseness, the structures determined for all the sites are not presented but rather are summarized in terms of structural types in Table 6. The relative mass abundance percentages of nonsialylated and fully and partially sialylated oligosaccharide structures are presented in Table 7. Individual sites are discussed below.

Asn-103. The most abundant structure at Asn-103 for CHO (16 total) and myeloma (28 total) cell line derived tICAM(453) was Man5+Fuc+2GlcNAc+2Gal+NeuGc while the most abundant structure for HEK (18 total) cell line derived tICAM(453) was Biant+Fuc+NeuAc. The majority of the oligosaccharide structures for all three cell lines were of the biantennary type (approximately 40%) with the rest of the structures approximately evenly divided between hybrid, tetraantennary and triantennary types.

Asn-118. The most abundant structures at Asn-118 for each cell line are as follows: CHO (19 total), Biant+Fuc+2NeuAc; HEK (29 total), Triant-2Gal+Fuc and Biant+NeuAc; and myeloma (34 total), Man3+Fuc+GlcNAc+Gal. The majority of the oligosaccharide structures for the CHO and HEK cell line derived tICAM(453) were of the triantennary and biantennary type, while the myeloma cell line derived protein contained primarily triantennary and hybrid structures.

Asn-175. The most abundant structure at Asn-175 for CHO (18 total) and myeloma (22 total) cell line derived tICAM(453) was Triant-2Gal+Fuc+NeuAc while the most abundant structure for HEK (22 total) cell line derived tICAM(453) was Man3+Fuc+GlcNAc+Gal+NeuGc. The majority of the oligosaccharide structures for the CHO and myeloma cell line derived tICAM(453) were of the triantennary type while the HEK cell line derived protein contained primarily hybrid plus triantennary structures.

Asn-240. The most abundant structures at Asn-240 for each cell line are as follows: CHO (10 total), Biant+Fuc+2NeuAc; HEK (17 total), Biant+ Fuc+NeuAc;

Table 3: Proposed Oligosaccharide Structures Determined by ESI-MS for the Asn-358 N-Linked Oligosaccharide Site for CHO Cell Line Derived tICAM(453) Listed in Order of Abundances

m/z[a]	measd mass[b]	calcd. mass[c]	abundance[d]	structure
1471.0^{3+}	2061.8	2060.9	0.219	Biant+Fuc+NeuAc
1568.1^{3+}	2353.1	2352.2	0.213	Biant+Fuc+2NeuAc
1796.6^{3+}	3038.6	3040.7	0.086	Triant+Fuc+NeuAc+2NeuGc
1373.8^{3+}	1770.2	1769.6	0.077	Biant+Fuc
1275.1^{4+}	2748.2	2749.5	0.070	Triant+Fuc+2NeuGc
1534.4^{4+}	3785.4	3786.4	0.042	Tetra+GlcNAc+4NeuGc
1690.0^{3+}	2718.8	2717.5	0.039	Triant+Fuc+2NeuAc
1613.9^{4+}	4103.4	4104.7	0.029	Tetra+Fuc+2GalGlcNAc+3NeuAc
1368.4^{3+}	1754.0	1752.6	0.028	Biant-Gal+NeuAc
1617.7^{3+}	2501.9	2500.3	0.028	Tetra+Fuc
1504.5^{4+}	3665.8	3665.3	0.026	Tetra+Fuc+4NeuAc
1695.1^{3+}	2734.1	2733.5	0.020	Triant+Fuc+NeuAc+NeuGc
1787.1^{3+}	3010.1	3010.8	0.020	Tetra+Fuc+GlcNAc+NeuGc
1403.3^{4+}	3261.0	3259.9	0.017	Tetra+NeuAc+2NeuGc or Tetra-Gal+Fuc+3NeuGc
1592.8^{3+}	2427.2	2426.2	0.013	Triant+Fuc+NeuAc
1113.3^{4+}	2101.0	2101.9	0.011	Triant-2Gal+Fuc+NeuAc
1178.4^{4+}	2361.4	2362.2	0.011	Tetra-3Gal+GlcNAc+NeuAc
1540.9^{4+}	3811.4	3812.4	0.011	Tetra+GlcNAc+GalGlcNAc+2NeuAc+NeuGc
1768.9^{3+}	2955.5	2953.3	0.009	Tetra+Fuc+Gal+NeuAc
1319.8^{3+}	1608.2	1607.5	0.009	Biant-Gal+Fuc
1265.7^{3+}	1445.9	1445.4	0.007	Biant-2Gal+Fuc
1598.4^{3+}	2444.0	2442.2	0.007	Triant+Fuc+NeuGc
1898.4^{3+}	3344.0	3343.0	0.006	Tetra-2Gal+Fuc+GlcNAc+GalGlcNAc+NeuAc+NeuGc

[a] Mass-to-charge ratio. [b] The oligosaccharide measured mass equals $[(m/z) \times z] - z$ minus the calculated total mass of the amino acids in the tryptic peptide (2348.2 amu). [c] Calculated by summing the monosaccharide average residue masses. The average residue mass is calculated from the atomic weights of the elements and is minus H_2O. [d] Calculated from the normalized peak heights in the m/z spectrum.

Table 4: Proposed Oligosaccharide Structures Determined by ESI-MS for the Asn-358 N-Linked Oligosaccharide Site for HEK Cell Line Derived tICAM(453) Listed in Order of Abundances

m/z[a]	measd mass[b]	calcd. mass[c]	abundance[d]	structure
1471.0^{3+}	2061.8	2060.9	0.254	Biant+Fuc+NeuAc
1484.7^{3+}	2102.9	2101.9	0.096	Triant-2Gal+Fuc+NeuAc
1373.8^{3+}	1770.2	1769.6	0.085	Biant+Fuc
1796.6^{3+}	3038.6	3040.7	0.070	Triant+Fuc+NeuAc+2NeuGc
1568.1^{3+}	2353.1	2352.2	0.066	Biant+Fuc+2NeuAc
1275.1^{4+}	2748.2	2749.5	0.060	Triant+Fuc+2NeuGc
1441.5^{3+}	1973.3	1972.8	0.034	Triant-Gal+Fuc
1613.9^{4+}	4103.4	4104.7	0.034	Tetra+Fuc+2GalGlcNAc+3NeuAc
1538.7^{3+}	2264.9	2264.1	0.030	Triant-Gal+Fuc+NeuAc
1387.5^{3+}	1811.3	1810.7	0.028	Triant-2Gal+Fuc
1695.0^{3+}	2733.8	2733.5	0.028	Triant+Fuc+NeuAc+NeuGc
1898.4^{3+}	3344.0	3343.0	0.028	Tetra-2Gal+Fuc+GlcNAc+GalGlcNAc+ NeuAc+NeuGc
1617.7^{3+}	2501.9	2500.3	0.023	Tetra+Fuc
1504.6^{4+}	3666.2	3665.3	0.021	Tetra+Fuc+4NeuAc
1533.9^{4+}	3783.4	3786.4	0.021	Tetra+GlcNAc+4NeuGc
1319.7^{3+}	1607.9	1607.5	0.015	Biant-Gal+Fuc
1403.3^{4+}	3261.0	3259.9	0.014	Tetra+NeuAc+2NeuGc or Tetra-Gal+Fuc+3NeuGc
1581.8^{3+}	2394.2	2395.2	0.014	Tetra-Gal+GlcNAc
1113.3^{4+}	2101.0	2101.9	0.013	Triant-2Gal+Fuc+NeuAc
1368.3^{3+}	1753.7	1752.6	0.013	Biant-Gal+NeuAc
1592.8^{3+}	2427.2	2426.2	0.013	Triant+Fuc+NeuAc
1768.9^{3+}	2955.5	2953.3	0.011	Tetra+Fuc+Gal+NeuAc
1178.6^{4+}	2362.2	2362.2	0.008	Tetra-3Gal+GlcNAc+NeuAc
1265.5^{3+}	1445.3	1445.4	0.008	Biant-2Gal+Fuc
1689.9^{3+}	2718.5	2717.5	0.008	Triant+Fuc+2NeuAc
1598.7^{3+}	2444.9	2442.2	0.005	Triant+Fuc+NeuGc

[a] Mass-to-charge ratio [b] The oligosaccharide measured mass equals $[(m/z) \times z] - z$ minus the calculated total mass of the amino acids in the tryptic peptide (2348.2 amu). [c] Calculated by summing the monosaccharide average residue masses. The average residue mass is calculated from the atomic weights of the elements and is minus H_2O. [d] Calculated from the normalized peak heights in the *m*/*z* spectrum.

Table 5: Proposed Oligosaccharide Structures Determined by ESI-MS for the Asn-358 N-Linked Oligosaccharide Site for Mouse Myeloma Cell Line Derived tICAM(453) Listed in Order of Abundances

m/z[a]	measd mass[b]	calcd. mass[c]	abundance[d]	structure
1275.1^{4+}	2748.2	2749.5	0.163	Triant+Fuc+2NeuGc
1530.4^{3+}	2240.0	2238.0	0.087	Biant+2NeuGc
1525.1^{3+}	2224.1	2222.0	0.066	Biant+NeuAc+NeuGc
1796.6^{3+}	3038.6	3040.7	0.065	Triant+Fuc+NeuAc+2NeuGc
1427.9^{3+}	1932.5	1931.5	0.064	Biant+Fuc+Gal
1958.8^{3+}	3525.2	3522.2	0.059	Tetra+Fuc+2GalGlcNAc+NeuAc
1617.8^{3+}	2502.2	2500.3	0.044	Tetra+Fuc
1613.9^{4+}	4103.4	4104.7	0.041	Tetra+Fuc+2GalGlcNAc+3NeuAc
1504.4^{4+}	3665.4	3665.3	0.040	Tetra+Fuc+4NeuAc
1476.3^{3+}	2077.7	2076.9	0.039	Biant+Fuc+NeuGc
1403.3^{4+}	3261.0	3259.9	0.033	Tetra+NeuAc+2NeuGc or Tetra-Gal+Fuc+3NeuGc
1695.0^{3+}	2733.8	2733.5	0.032	Triant+Fuc+NeuAc+NeuGc
1578.7^{3+}	2384.9	2384.2	0.032	Biant+Fuc+2NeuGc
1540.9^{4+}	3811.4	3812.4	0.028	Tetra+GlcNAc+GalGlcNAc+2NeuAc+NeuGc
1598.4^{3+}	2444.0	2442.2	0.025	Triant+Fuc+NeuGc
1373.8^{3+}	1770.2	1769.6	0.024	Biant+Fuc
1178.4^{4+}	2361.4	2362.2	0.022	Tetra-3Gal+GlcNAc+NeuAc
1470.9^{3+}	2061.5	2060.9	0.021	Biant+Fuc+NeuAc
1573.4^{3+}	2369.0	2368.2	0.021	Biant+Fuc+NeuAc+NeuGc
1113.2^{4+}	2100.6	2101.9	0.017	Triant-2Gal+Fuc+NeuAc
1768.7^{3+}	2954.9	2953.3	0.014	Tetra+Fuc+Gal+NeuAc
1481.7^{3+}	2093.9	2093.0	0.014	Biant+2Gal+Fuc
1368.1^{3+}	1753.1	1752.6	0.014	Biant-Gal+NeuAc
1257.0^{3+}	1419.8	1420.3	0.012	Man5+GlcNAc
1849.8^{3+}	3198.2	3197.9	0.010	Tetra-Gal+Fuc+GlcNAc+GalGlcNAc+NeuAc
1700.9^{3+}	2751.5	2749.5	0.008	Triant+Fuc+2NeuGc
1568.1^{3+}	2353.1	2352.2	0.008	Biant+Fuc+2NeuAc

[a] Mass-to-charge ratio. [b] The oligosaccharide measured mass equals $[(m/z) \times z] - z$ minus the calculated total mass of the amino acids in the tryptic peptide (2348.2 amu). [c] Calculated by summing the monosaccharide average residue masses. The average residue mass is calculated from the atomic weights of the elements and is minus H_2O. [d] Calculated from the normalized peak heights in the *m*/*z* spectrum.

and myeloma (22 total), Biant+Fuc+Gal+NeuGc. The majority of the oligosaccharide structures for all three cell lines were composed of biantennary and triantennary types.

Asn-269. The most abundant structure at Asn-269 for the CHO (8 total), HEK (10 total), and myeloma (13 total) cell lines was Biant+NeuAc+NeuGc. The majority of the oligosaccharide structures for all three cell lines were composed of biantennary types.

Asn-358. The most abundant structure at Asn-358 for CHO (23 total) and HEK (26 total) cell line derived tICAM-

Table 6: Relative Mass Abundance Percentages of Different N-Linked Oligosaccharide Structures at Each of Seven Sites on CHO, HEK, and Mouse Myeloma Cell Line Derived tICAM(453) as Determined by ESI-MS

structure[a]	cell line	Asn-103	Asn-118	Asn-175	Asn-240	Asn-269	Asn-358	Asn-379
Tetra+2LN	CHO	0	0	3	0	1	3	29
	HEK	0	0	2	10	1	3	29
	myeloma	0	0	2	6	2	10	29
Tetra+LN	CHO	2	6	0	7	0	2	12
	HEK	0	3	0	7	1	3	9
	myeloma	0	1	0	3	2	4	9
Tetra	CHO	14	0	8	1	6	15	27
	HEK	28	1	9	2	6	11	27
	myeloma	20	0	10	0	10	15	22
Triant	CHO	14	32	63	15	2	25	8
	HEK	15	40	24	21	2	38	8
	myeloma	20	45	58	22	9	31	6
Biant	CHO	41	50	10	70	91	55	24
	HEK	44	34	15	49	90	44	26
	myeloma	34	10	7	49	77	39	34
hybrid	CHO	29	8	13	7	0	0	0
	HEK	13	16	46	10	0	0	1
	myeloma	26	28	16	21	0	1	0
oligomannosyl	CHO	0	5	3	0	0	0	0
	HEK	0	8	3	0	0	0	0
	myeloma	0	16	5	0	0	0	0

[a] LN is an abbreviation for *N*-acetyllactosamine.

Table 7: Relative Mass Abundance Percentages of Nonsialylated and Fully and Partially Sialylated N-Linked Oligosaccharide Structures at Each of Seven Sites on CHO, HEK, and Mouse Myeloma Cell Line Derived tICAM(453)

degree of sialylation	cell line	Asn-103	Asn-118	Asn-175	Asn-240	Asn-269	Asn-358	Asn-379
complete	CHO	19	24	6	53	89	37	22
	HEK	3	6	29	34	86	18	30
	myeloma	14	11	9	22	72	32	31
partial	CHO	59	35	74	29	10	51	62
	HEK	76	36	32	43	13	61	53
	myeloma	68	38	66	38	22	54	59
neutral	CHO	22	41	20	18	1	12	16
	HEK	21	58	39	23	1	21	17
	myeloma	18	51	25	40	6	14	10

(453) was Biant+Fuc+NeuAc, while the most abundant structure for the myeloma (27 total) cell line derived tICAM-1 was Triant+Fuc+2NeuGc. The majority of the oligosaccharide structures for all three cell lines were composed of biantennary and triantennary types.

Asn-379. The most abundant structures at Asn-379 for each cell line are as follows: CHO (29 total), Tetra+NeuGc; HEK (27 total), Biant+2NeuAc; and myeloma (26 total), Biant+Fuc+NeuAc+NeuGc. The majority of the oligosaccharide structures for all three cell lines were distributed evenly between tetrantennary+*N*-acetyllactosamine, tetrantennary, and biantennary types.

DISCUSSION

In this study, oligosaccharides at seven out of eight N-linked glycosylation sites of CHO, HEK, and murine myeloma cell line derived tICAM(453) were characterized. The analytical technique used was electrospray ionization mass spectroscopy combined with on-line reversed phase HPLC chromatography (Hemling et al., 1990; Ling et al., 1991; Carr et al., 1991, 1993; Guzzetta et al., 1993). Although sialylated oligosaccharide structures can be deduced from the molecular weights of glycopeptides, the linkage, sequence, and branching information cannot be determined directly by this technique. To the best of our knowledge, ESI-MS cannot be used to directly determine the absolute concentration of any given oligosaccharide in a complex mixture bound to a peptide. However, our experience with tICAM(453) and one other glycoprotein (unpublished results) suggests that, for a given glycopeptide, the relative intensities of peaks in the mass spectra can be used in a comparative manner to provide information as to the relative abundances of oligosaccharide structures.

One of the interesting findings of this study is that the oligosaccharide structures and the degree of sialylation at different N-linked sites in the tICAM(453) molecule can be grossly different. For example, the structures at Asn-269 are predominantly fully sialylated biantennary while those at Asn-379 are predominantly partially sialylated tetraantennary complex. These results suggest that protein structure influences glycosylation of a given site. The crucial role that protein structure plays in specifying N-linked oligosaccharide primary structure has been demonstrated in numerous other studies of mammalian recombinant and non-recombinant glycoproteins: Sindbis virus glycoproteins, four N-linked sites (Weitzman et al., 1979; Hsieh et al., 1983; Hubbard, 1988); HLA-DR antigen, three sites (Neel et al., 1987); an artificial avidin−glycan complex, one site (Shao

et al., 1989); immunoglobulin G, one site (Lee et al., 1990); ovoinhibitor, three sites; ovotransferrin, one site; orosomucoid, five sites; antitrypsin, three sites (Yet & Wold, 1990); ovalbumin, one site (Sheares & Robbins, 1986); human erythropoietin, three sites (Sasaki, 1988); human tissue plasminogen activator, three sites (Pohl et al., 1987; Pfeiffer et al., 1989; Parekh et al., 1989; Spellman et al., 1989); human interferon-β1, one site (Kagawa et al., 1988), and; human corticosteroid-binding globulin, six sites (Avvakumov & Hammond, 1994). The uniqueness of the present study is that tICAM(453) contains more N-linked sites than any other previously studied glycoprotein and the analytical technique used does not require desialylation before analysis.

The number of oligosaccharide structures at a given N-linked site of tICAM(453), depending on the cell line, vary from eight to 34. This inter- and intra-cell line diversity of oligosaccharide structures at a given site is believed to be related primarily to the presence, concentration, kinetic characteristics, and compartmentalization of the individual glycosyltransferases and glycosidases found in a given cell line (Rademacher et al., 1988). In addition, for cells in culture, the glycosylation apparatus is sensitive to numerous environmental perturbations (Goochee & Monica, 1990).

Tetraantennary complex structures containing one and/or two *N*-acetyllactosamine repeats were found on tICAM(453) primarily at Asn-379 (approximately 40% of the structures at this site). Although one or two *N*-acetyllactosamine repeats have been found on secretory proteins such as α_1-acid glycoprotein (Yoshima et al., 1981) and erythropoietin (Sasake et al., 1987), it is believed that these oligosaccharides are most likely to be found attached to membrane glycoproteins (Fukuda et al., 1988). It has been shown that the secretory form of human chorionic gonadotropin-α chain contains typical N-linked complex-type oligosaccharides while a membrane-bound form of the protein contains *N*-acetyllactosamine repeats (Fukuda et al., 1988). The authors concluded that membrane anchoring increases the accessibility of the N-linked glycans to the enzymes involved in *N*-acetyllactosamine addition and hypothesized that the N-linked glycans to which *N*-acetyllactosamine is added are located near the membrane. The hydrodynamic properties of tICAM(453) indicate it to be highly asymmetric and elongated (Greve et al., 1991). In addition, electron micrographs of the molecule show it as a thin rod in two forms: either straight or with a single bend between the D2 and D3 or the D3 and D4 domains. These experiments have led to a model of the ICAM-1 molecule as a straight rod oriented perpendicular to the membrane surface (Staunton et al., 1990). This model would place Asn-379, located between the D4 and D5 domains, close to the membrane surface. In light of the chorionic gonadotropin-α subunit results it is interesting to speculate that the location of Asn-379 near the C-terminal in the rod-like tICAM(453) molecule influenced the attachment of *N*-acetyllactosamine to oligosaccharides at this site. In addition, membrane-bound ICAM-1 may contain more *N*-acetyllactosamine residues at Asn-379 than soluble tICAM(453).

It has been observed that ICAM-1 isolated from B lymphoblastoid and myelomonocytic cells differ in the extent of N-linked carbohydrate processing (Dustin et al., 1986). Thus it has been suggested that the level and type of N-linked glycosylation on ICAM-1 may regulate biological interaction with cells bearing specific leukocyte integrins: myeloid cells may accumulate if ICAM-1 is relatively deglycosylated, whereas lymphocytes may localize if ICAM-1 is more fully glycosylated (Diamond et al., 1991). Mutational analysis and the use of agents that interfere with carbohydrate processing have provided evidence that the size of the N-linked oligosaccharide side chains at Asn-240 and Asn-269 in the third domain of ICAM-1 determine the binding avidity of ICAM-1 to Mac-1 but not to LFA-1 (Diamond et al., 1991). In the present study, the tICAM(453) Asn-269 site was found to contain predominantly one oligosaccharide structure that is conserved in all three cell lines. On the other hand, the Asn-240 site was found to contain cell line dependent oligosaccharide structural heterogeneity particularly in the degree of sialyation.

Prediction of the in vivo behavior of tICAM(453) was one of the goals of oligosaccharide structural analysis. It is known that the sialic acid NeuGc is an oncofetal antigen in humans (Fujii et al., 1982; Nakarai et al., 1990; Hirabayashi et al., 1987) and that the degree of glycoprotein sialylation influences circulatory system clearance in vivo due to interaction with hepatic asialoglycoprotein receptors (Ashwell & Harford, 1982). We have demonstrated that, depending on the cell line and N-linked oligosaccharide binding site, tICAM(453) glycoforms of differing sialic acid content exist which may clear at different rates from the circulatory system as well as expose differing amounts of NeuGc to the immune system. Knowledge of the protein surface topology of the oligosaccharides at each N-linked site would add another dimension to the prediction of the in vivo behavior of tICAM(453).

ACKNOWLEDGMENT

We thank M. Dumas and W. Dang for supplying us with purified tICAM(453). We also acknowledge that Dr. S. Cho, M. Tran, and H. Yee generated the CHO and HEK cell lines and that Dr. L. Miller generated the NS0 cell line.

REFERENCES

Ashwell, G., & Harford, J. (1982) *Annu. Rev. Biochem. 51*, 531−554.

Avvakumov, G. V., & Hammond, G. L. (1994) *Biochemistry 33*, 5759−5765.

Berendt, A. R., Simmons, D. L., Tansey, J., Newbold, C. I., & Marsh, K. (1989) *Nature 341*, 57−59.

Berendt, A. R., McDowall, A., Craig, A. G., Bates, P. A., Sternberg, M. J. E., Marsh, K., Newbold, C. I., & Hogg, N. (1992) *Cell 68*, 71−81.

Carr, S. A., Hemling, M. E., Bean, M. F., & Roberts, G. D. (1991) *Anal. Chem. 63*, 2802−2824.

Carr, S. A., Huddleston, M. J., & Bean, M. F. (1993) *Protein Sci. 2*, 183−196.

Diamond, M. S., Staunton, D. E., deFougerolles, A. R., Stacker, S. A., Garcia-Aguilar, J., Higgs, M. L., & Springer, T. A. (1990) *J. Cell Biol. 111*, 3129−3139.

Diamond, M. S., Staunton, D. E., Marlin, S. D., & Springer, T. A. (1991) *Cell 65*, 961−971.

Dustin, M. L., Rothlein, R., Bhan, A. K., Dinarello, C. A., & Springer, T. A. (1986) *J. Immunol. 137*, 245−254.

Fujii, I., Higashi, H., Ikuta, K., Kata, S., & Naiki, M. (1982) *Mol. Immunol. 19*, 87−94.

Fukuda, M., Guan, J. L., & Rose, J. K. (1988) *J. Biol. Chem. 263*, 5314−5318.

Goochee, C. F., & Monica, T. (1990) *BioTechnology 8*, 421−427.

Goodley, P. C., Fischer, S., Kuhlmann, F., & Apffel, A. (1994) *Protein Sci. 3* (Suppl. 1), 85.

Greve, J. M., Davis, G., Meyer, A. M., Forte, C. P., Yost, S. C., Marlor, C. W., Kamarck, M. E., & McClelland, A. (1989) *Cell 56*, 839−847.
Greve, J. M., Forte, C. P., Marlor, C. W., Meyer, A. M., Hoover-Litty, H., Wunderlich, D., & McClelland, A. (1991) *J. Virol. 65*, 6015−6023.
Guzzetta, A. W., Basa, L. J., Hancock, W. S., Keyt, B. A., & Bennett, W. F. (1993) *Anal. Chem. 65*, 2953−2962.
Hemling, M. E., Roberts, G. D., Johnson, W., & Carr, S. A. (1990) *Biomed. Environ. Mass Spectrom. 19*, 677−691.
Hirabayashi, Y., Kasakura, H., Matsumoto, M., Higashi, H., Kato, S., Kasai, N., & Naiki, M. (1987) *Jpn. J. Cancer Res. 78*, 251−260.
Hsieh, P., Rosner, M. R., & Robbins, P. W. (1983) *J. Biol. Chem. 258*, 2555−2561.
Hubbard, S. C. (1988) *J. Biol. Chem. 263*, 19303−19317.
Kagawa, Y., Takasaki, S., Utsumi, J., Hosoi, K., Shimizu, H., Kochibe, N., & Kobata, A. (1988) *J. Biol. Chem. 263*, 17508−17515.
Kolatkar, P. R., Oliveira, M. A., Rossmann, M. G., Robbins, A. H., Katti, S. K., Hoover-Litty, H., Forte, C., Greve, J. M., McClelland, A., & Olson, N. H. (1992) *J. Mol. Biol. 225*, 1127−1130.
Lee, S. O., Connolly, J. M., Ramirez-Soto, D., & Poretz, R. D. (1990) *J. Biol. Chem. 265*, 5833−5839.
Lineberger, D. W., Graham, D. J., Tomassini, J. E., & Colonno, R. J. (1990) *J. Virol. 64*, 2582−2587.
Ling, V., Guzzetta, A. W., Canova-Davis, E., Stults, J. T., Hancock, W. S., Covey, T. R., & Shushan, B. I. (1991) *Anal. Chem. 63*, 2909−2915.
Marlin, S. D., & Springer, T. A. (1987) *Cell 51*, 813−819.
Marlin, S. D., Staunton, D. E., Springer, T. A., Stratowa, C., Sommergruber, W., & Merluzzi, V. J. (1990) *Nature 344*, 70−72.
McClelland, A., DeBear, J., Yost, S. C., Meyer, A. M., Marlor, C. W., & Greve, J.M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 7993−7997.
Nakarai, H., Chandler, P. J., Kano, K., Morton, D. L., & Irie, R. F. (1990) *Int. Arch. Allergy Appl. Immunol. 91*, 323−328.
Neel, D., Merlu, B., Turpin, E., Rabourdin-Combe, C., Mach, B., Goussault, Y., & Charron, D. J. (1987) *Biochem. J. 244*, 433−442.
Ockenhouse, C. F., Betageri, R., Springer, T. A., & Staunton, D. E. (1992) *Cell 68*, 63−69.
Olson, N. H., Kolatkar, P. R., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S., & Rossmann, M. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 507−511.
Parekh, R. B., Dwek, R. A., Thomas, J. R., Opdenakker, G., & Rademacher, T. W. (1989) *Biochemistry 28*, 7644−7662.
Pfeiffer, G. Schmidt, M., Strube, K. H., & Geyer, R. (1989) *Eur. J. Biochem. 186*, 273−286.
Pohl, G., Kenne, L., Nilsson, B., & Einarsson, M. (1987) *Eur. J. Biochem. 170*, 69−75.
Rademacher, T. W., Parekh, R. B., & Dwek, R. A. (1988) *Annu. Rev. Biochem. 57*, 785−838.
Register, R. B., Uncapher, C. R., Naylor, A. M., Lineberger, D. W., & Colonno, R. J. (1991) *J. Virol. 65*, 6589−6596.
Rothlein, R., Dustin, M. L., Marlin, S. D., & Springer, T. A. (1986) *J. Immunol. 137*, 1270−1274.
Sasaki, H., Bothner, B., Dell, A., & Fukuda, M. (1987) *J. Biol. Chem. 262*, 12059−12076.
Sasaki, H., Ochi, N., Dell, A., & Fukuda, M. (1988) *Biochemistry 27*, 8618−8626.
Shao, M. C., Krudy, G., Rosevear, P. R., & Wold, F. (1989) *Biochemistry 28*, 4077−4083.
Sheares, B. T., & Robbins, P. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1993−1997.
Simmons, D., Makgoba, M. W., & Seed, B. (1988) Nature *331*, 624−627.
Smith, C. W., Marlin, S. D., Rothlein, R., Toman, C., & Anderson, D. C. (1989) *J. Clin. Invest. 83*, 2008−2017.
Spellman, M. W., Basa, L. J., Leonard, C. K., Chakel, J. A., O'Connor, J. V., Wilson, S., & van Halbeek, H. (1989) *J. Biol. Chem. 264*, 14100−14111.
Staunton, D. E., Marlin, S. D., Stratowa, C., Dustin, M. L., & Springer, T. A. (1988) *Cell 52*, 925−933.
Staunton, D. E., Merluzzi, V. J., Rothlein, R., Barton, R., Marlin, S. D., & Springer, T. A. (1989) *Cell 56*, 849−853.
Staunton, D. E., Dustin, M. L., Erickson, H. P., & Springer, T. A. (1990) *Cell 61*, 243−254.
Tomassini, J. E., Graham, D., DeWitt, C. M., Lineberger, D. W., Rodkey, J. A., & Colonno, R. J. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4907−4911.
Tomassini, J. E., Maxson, T. R., & Colonno, R. J. (1989b) *J. Biol Chem. 264*, 1656−1662.
Weitzman, S., Grennon, M., & Keegstra, K. (1979) *J. Biol. Chem. 254*, 5377−5382.
Yet, M. G., & Wold, F. (1990) *Arch. Biochem. Biophys. 278*, 356−364.
Yoshima, H., Matsumoto, A., Mizuochi, T., Kawasaki, T., & Kobata, A. (1981) *J. Biol. Chem. 256*, 8476−8484.

BI952354M