Glycosylation Site of Band 3, the Human Erythrocyte Anion-Exchange Protein†

Daniel G. Jay

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Received August 21, 1985

ABSTRACT: The band 3 protein has a single glycosylation site on the carboxy-terminal 55 000-dalton tryptic fragment that defines a sequence of the polypeptide on the extracytoplasmic surface of the cell. To locate this site, a novel procedure involving end labeling of the 55 000-dalton tryptic fragment was used. Peptides resulting from partial proteolysis of the end radiolabeled glycoprotein were separated by lectin–Sepharose chromatography and analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and autoradiography. The smallest fragment observed defined the distance between the glycosylation site and the amino terminus. The procedure was first tested on a protein for which the location of the glycosylation site is known, HLA-B7 antigen. It was then used to show that the glycosylation site of human band 3 is 28 000 ± 3000 daltons from the carboxy terminus of the protein.

Band 3, the anion-exchange protein of the human erythrocyte, is a 95 000-dalton membrane glycoprotein (Fairbanks et al., 1971) known to have a single carbohydrate chain (Drickamer, 1978) that binds to wheat germ agglutinin (Tsuji et al., 1980). The glycosyl moiety has been extensively studied, and its structure has been determined for both fetal (Fukuda et al., 1984a) and adult forms of band 3 (Fukuda et al., 1984b). In individuals who lack the erythrocyte En^a antigen, the band 3 protein has an altered carbohydrate composition (Tanner & Anstee, 1976). By proteolytic analysis the location of the glycosylation site has been restricted to the COOH-terminal 28 000-dalton fragment of the band 3 polypeptide resulting from papain digestion (Jennings et al., 1984). The precise location of the glycosylation site within this fragment is not known.

The sequence of murine band 3 has recently been determined (Kopito & Lodish, 1985), and the similarity with regions for which sequence is known for human band 3 suggests that these two proteins are highly homologous. Mouse band 3 has two potential glycosylation sites at Asn-611 and Asn-660. Residue 611 corresponds to a site in human band 3 that has been shown not to be glycosylated (Brock et al., 1983), implying that mouse band 3 is glycosylated at Asn-660. The homology between mouse and human band 3 would suggest that this may also be the glycosylation site of human band 3, but there has been no direct evidence for this.

A procedure to label proteins exclusively at the amino terminus and locate specific amino acids in the sequence has recently been reported (Jay, 1984; Jue & Doolittle, 1985). A consequence of this procedure is the ability to locate any sites of particular interest in the primary sequence of the protein. Since one position (the amino terminus) is defined by radiolabel, the location of any second site that can be specifically recognized is defined by the smallest fragment that contains both labels. This procedure was used to locate the glycosylation site of band 3, the anion-exchange protein of the human erythrocyte.

Materials and Methods

Purified papain-solubilized HLA-B7 antigen was a generous gift of Ken Parker (Department of Biochemistry and Molecular Biology, Harvard University). The carboxy-terminal 55 000-dalton tryptic fragment (Tr55) and the carboxy-terminal 35 000-dalton chymotryptic fragment (Ch35) were prepared as previously described (Reithmeier, 1979) and were purified by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) followed by gel elution in 50 mM $NaHCO_3$–0.2% SDS, overnight at 37 °C with rocking, followed by extensive dialysis against the same buffer [to remove tris(hydroxymethyl)aminomethane (Tris) and glycine] and lyophilization. The Ch35 fragment was further purified by affinity chromatography with wheat germ agglutin (WGA)–Sepharose. The proteins were end-labeled and digested by the method of Jay (1984). The protein (100 μg in 200 μL of 500 mM $NaHCO_3$, pH 9.8, and 2% SDS) was incubated at 50 °C for 1 h with 10 μL of phenyl isothiocyanate (PITC; Pierce) under nitrogen with occasional vortexing. The protein was precipitated by the addition of 1 mL of acetone at room temperature, collected by centrifugation, and washed 5 times with 0.5 mL of acetone per wash and dried. The pellet was redissolved in 100 μL of trifluoroacetic acid (TFA; Pierce) and incubated under nitrogen at 50 °C for 5 min. Sodium borate (100 μL of a 100 mM solution at pH 8.5, with 1% SDS) was added, and the sample was dialyzed overnight against the same buffer. The PITC/TFA-treated protein was labeled with ^{125}I-labeled Bolton–Hunter reagent (Bolton & Hunter, 1973).

End-labeled HLA B7 antigen in 0.25% SDS–25 mM sodium borate, pH 8.5, was digested at a protease:substrate ratio of 20:1 at 37 °C in a total volume of 20 μL. The incubation time was dependent on the protein sample and the protease used. Proteolysis was terminated by the addition of an equal volume of 2× Laemmli sample buffer and boiling for 2 min. A 5-μL aliquot of each sample was removed for direct loading onto SDS–PAGE, and the remainder of the sample was diluted with 1 mL of incubation buffer (1% Triton X-100, 50 mM Tris-HCl, pH 7.4). Concanavalin A–agarose (Bio-Rad), 25 μL, was added to each tube and incubated overnight at 4 °C with gentle rocking. The lectin–agarose-bound glycopeptides were collected by centrifugation and washed 5 times with 1 mL of incubation buffer and once with 1 mL of 0.1% SDS–50 mM Tris-HCl, pH 7.4, and eluted with 25 μL of Laemmli sample buffer. Partial digests and lectin–agarose-adsorbed partial digests were applied to SDS–PAGE (Laemmli, 1970) and visualized by autoradiography. The same procedures were applied to the Tr55 and Ch35 fragments except that glyco-

†This work was supported by Grant HL 08893 from the National Institutes of Health to G. Guidotti, Harvard University.

0006-2960/86/0425-0554$01.50/0

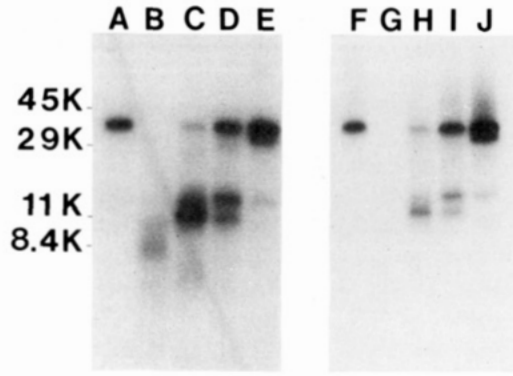

FIGURE 1: Mapping the glycosylation site of HLA B7 antigen. End-labeled HLA B7 antigen was digested for 5 min with chymotrypsin, 2 min with elastase, 30 min with trypsin, and 30 min with *S. aureus* protease. Partial digests and lectin–agarose-adsorbed partial digests were applied to 18% polyacrylamide Laemmli SDS–PAGE to directly compare the respective samples and visualized by autoradiography. (Lane A) Undigested HLA B7; (lane B) chymotrypsin-digested HLA B7; (lane C) elastase-digested HLA B7; (lane D) trypsin-digested HLA B7; (lane E) *S. aureus* protease digested HLA B7; (lane F) con A–agarose adsorbed undigested HLA B7; (lane G) con A–agarose-adsorbed chymotrypsin-digested HLA B7; (lane H) con A–agarose-adsorbed elastase-digested HLA B7; (lane I) con A–agarose-adsorbed trypsin-digested HLA B7; (lane J) con A–agarose-adsorbed *S. aureus* protease digested HLA B7.

peptides were isolated on WGA–Sepharose.

Results and Discussion

The location of the glycosylation site of a glycoprotein can be determined by subjecting an end-labeled glycoprotein to partial proteolysis, isolating the peptides that bind to a lectin by affinity chromatography, and analyzing these peptides by SDS–PAGE and autoradiography. Every band in the ladder of bands observed has both the amino terminus and the glycosyl moiety with the smallest fragment observed defining the maximum distance of the sugar from the amino terminus. The largest radiolabeled fragment not retained by affinity chromatography defines the minimum distance between the carbohydrate and the amino terminus.

The procedure was tested on a glycoprotein in which the location of the glycosylation site is known. HLA class I antigen B7 has a single carbohydrate chain that binds to concanavalin A, and this group is attached to Asn-86 (Orr et al., 1979). The result of experiments to locate the glycosylation site by end radiolabeling and lectin–agarose affinity chromatography are shown in Figure 1. Lanes A–E show the autoradiograph of the SDS gel of end-radiolabeled intact HLA-B7 and HLA-B7 subjected to partial proteolyses; lanes F–J show the autoradiograph of the SDS gel of the respective samples isolated by lectin–agarose chromatography. All the high molecular weight species are retained by the lectin-conjugated agarose while the smaller species are not. The binding is efficiently competed out by the presence of 0.2 M methyl α-mannoside, the specific ligand for concanavalin A (data not shown). The smallest radiolabeled fragment that is retained has an M_r of 11 000, while the largest radiolabeled fragment that is not retained by lectin–agarose chromatography has an M_r of 8400. These fragments define a range around the location of the glycosylation site. By assuming an average molecular weight of 110 per residue, the glycosylation site of HLA-B7 is 90 ± 10 residues from the amino terminus, which is very close to the actual value of 86. It should be noted that the largest fragment retained contains carbohydrate, which introduces additional uncertainty in the estimation of the size of this peptide. In general, the glycosyl moiety will decrease

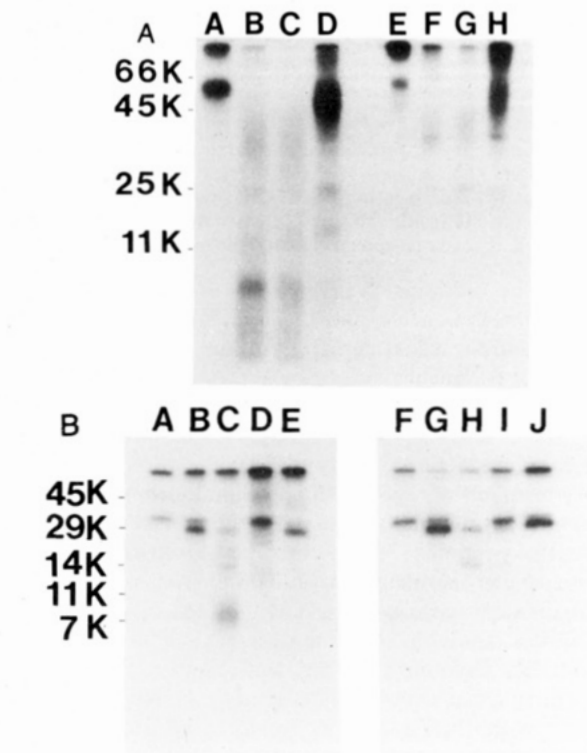

FIGURE 2: (A) Mapping the glycosylation site of band 3 from the carboxy-terminal 55 000-dalton tryptic fragment (Tr55). Tr55 was digested for 15 min with chymotrypsin, 5 min with elastase, and 30 min with trypsin. The samples were applied to a 15% polyacrylamide Laemmli gel and visualized by autoradiography. (Lane A) Undigested Tr55; (lane B) chymotrypsin-digested Tr55; (lane C) elastase-digested Tr55; (lane D) trypsin-digested Tr55; (lane E) WGA–Sepharose-adsorbed undigested Tr55; (lane F) WGA–Sepharose-adsorbed chymotrypsin-digested Tr55; (lane G) WGA–Sepharose-adsorbed elastase-digested Tr55; (lane H) WGA–Sepharose-adsorbed trypsin-digested Tr55. (B) Mapping the glycosylation site of band 3 from the carboxy-terminal 35 000-dalton chymotryptic fragment (Ch35). Ch35 was digested for 15 min with chymotrypsin, 5 min with elastase, 30 min with trypsin, and 30 min with *S. aureus* protease. The samples were applied to an 18% polyacrylamide Laemmli gel and visualized by autoradiography. (Lane A) Undigested Ch35; (lane B) chymotrypsin-digested Ch35; (lane C) elastase-digested Ch35; (lane D) trypsin-digested Ch35; (lane E) *S. aureus* protease digested Ch35; (lane F) WGA–Sepharose-adsorbed undigested Ch35; (lane G) WGA–Sepharose-adsorbed chymotrypsin-digested Ch35; (lane H) WGA–Sepharose-adsorbed elastase-digested Ch35; (lane I) WGA–Sepharose-adsorbed trypsin-digested Ch35; (lane J) WGA–Sepharose-adsorbed *S. aureus* protease digested Ch35.

the migration of the fragment, giving it an anomalously high apparent molecular weight, and the extent of this aberration will depend upon several factors including size of the glycosyl moiety, porosity of the gel, and the carbohydrate composition. Taking into account the difficulty in resolution of fragments of this size on SDS–PAGE and the uncertainty in size due to the glycosyl moiety itself, this estimation is quite good and demonstrates the predictive value of the method.

By use of the same analysis, the glycosylation site of band 3 was determined. Since the amino terminus of band 3 is blocked (Drickamer, 1978) and the end-labeling procedure must be done on species with a free amino terminus (Jay, 1984), it was necessary to use proteolytic fragments of band 3, which contain the glycosylation site. The analysis was carried out on the COOH-terminal 55 000-dalton tryptic fragment. The results of this experiment are shown in Figure 2A. Lanes A–D show the autoradiograph of samples directly loaded onto the gel while lanes E–H show the autoradiograph of the gel with the corresponding samples isolated by WGA–Sepharose chromatography. This binding is completely pre-

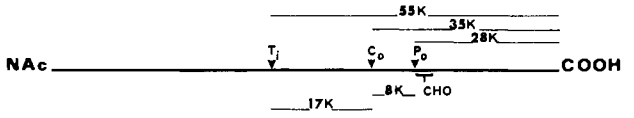

FIGURE 3: Summary of proteolytic cleavages used in the band 3 polypeptide and the location of the glycosylation site: NAc, *N*-acetylmethioninyl amino terminus; COOH, carboxy terminus; T_i, trypsin cleavage site inside the cell; C_o, chymotrypsin cleavage site outside the cell; P_o, external papain cleavage site outside the cell; CHO, glycosylation site.

vented by the presence of 0.2 M *N*-acetylglucosamine, the specific ligand for wheat germ agglutinin (data not shown). The smallest fragment retained by WGA–Sepharose has an M_r of 27 000 while the largest fragment not retained has an M_r of 15 000. This suggests that the glycosylation site of band 3 is located between 40 000 and 28 000 daltons from the carboxy terminus of the band 3 polypeptide. Jennings et al. (1984) have established that the glycosyl moiety is localized to the carboxy-terminal 28 000-dalton papain fragment. These two data suggest that the glycosylation site is located just after the papain cleavage site (Figure 3).

A similar analysis was done on the COOH-terminal 35 000-dalton chymotryptic fragment (Figure 2B). A comparison of a partial elastase digest of end-radiolabeled protein (lane C) with this sample isolated by WGA–Sepharose chromatography (lane H) demonstrates that a fragment of 14 000 daltons is retained by WGA–Sepharose while the next smallest fragment (7000 daltons) is not. This implies that the glycosylation site is between 21 000 and 28 000 daltons from the carboxy terminus. This finding is consistent within experimental error with the result obtained from the 55 000-dalton tryptic fragment.

Finally, the location is in agreement with that postulated by Kopito & Lodish (1985) for murine band 3, which is homologous to human band 3. The two analyses and the sequence information from a highly homologous protein argue strongly that the glycosylation site of band 3 is located 28 000 daltons from the COOH terminus (Figure 3).

CONCLUSIONS

The location of the glycosylation site of band 3, which must be on the extracytoplasmic surface of the cell, is supported by the evidence that there is an exofacial papain cleavage site here (Figure 3) (Jennings & Adams, 1981). Ramjeesingh et al. (1983), on the basis of the arrangement of chymotryptic fragments of the carboxy-terminal 28 000-dalton papain fragment, have suggested that the carbohydrate moiety is closer to the carboxy terminus than the two cysteine residues in this region of the sequence. The data presented here coupled with the locations of the cysteines at positions 861 and 903 in the sequence of murine band 3 (Kopito & Lodish, 1985), a highly homologous protein, imply that the arrangement of Ramjeesingh is incorrect.

We have demonstrated a method by which the location of a glycosylation site in the sequence of a glycoprotein can be determined and have illustrated how important second sites in general may be mapped in the sequence of proteins. The procedure is limited by the resolution of SDS–PAGE and for glycoproteins by the anomalous binding of SDS. This may be especially true for small glycopeptides resulting from glycoproteins whose carbohydrate moiety is near the amino terminus. Also, the localization of the site of interest is dependent on the presence and proximity of proteolytic cleavage sites. Although not impossible, the large variety of proteolytic reagents makes it unlikely that cleavage will not be effected near the site of interest. It should also be noted that for glycoproteins that have multiple glycosylation sites only the site nearest the amino terminus can be located. These limitations aside, site mapping with end-labeled proteins may be a useful tool in the determination of protein structure.

ACKNOWLEDGMENTS

I thank Guido Guidotti in whose laboratory this work was conducted for his enthusiastic support and advice, Chung Wang for his generous gift of WGA–Sepharose, Ken Parker for his generous gift of HLA B7 antigen, Walter Gilbert and Jonathan Lytton for helpful discussion, and Ron Kopito and Harvey Lodish for prior communication of their manuscript.

REFERENCES

Bolton, A. E., & Hunter, W. M. (1973) *Biochem. J. 133*, 529–539.

Brock, C. J., Tanner, M. J. A., & Kempf, C. (1983) *Biochem. J. 213*, 577–586.

Drickamer, L. K. (1978) *J. Biol. Chem. 253*, 7242–7248.

Fairbanks, G., Steck, T. L., & Wallach, D. F. H. (1971) *Biochemistry 10*, 2606–2617.

Fukuda, M., Dell, A., & Fukuda, M. N. (1984a) *J. Biol. Chem. 259*, 4782–4791.

Fukuda, M., Dell, A., Oates, J. E., & Fukuda, M. N. (1984b) *J. Biol. Chem. 259*, 8260–8273.

Jay, D. G. (1984) *J. Biol. Chem. 259*, 15572–15578.

Jennings, M. L., & Adams, M. F. (1981) *Biochemistry 20*, 7118–7123.

Jennings, M. L., Adams-Lackey, M., & Denny, G. H. (1984) *J. Biol. Chem. 259*, 4652–4660.

Jue, R. A., & Doolittle, R. F. (1985) *Biochemistry 24*, 162–170.

Kopito, R. R., & Lodish, H. F. (1985) *Nature (London) 316*, 234–238.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Orr, H. T., Lopez de Castro, J. A., Lancet, D., & Strominger, J. L. (1979) *Biochemistry 18*, 5711–5719.

Ramjeesingh, M., Gaarn, A., & Rothstein, A. (1983) *Biochim. Biophys. Acta 729*, 150–160.

Reithmeier, R. A. F. (1979) *J. Biol. Chem. 254*, 3054–3060.

Tanner, M. J. A., & Anstee, D. J. (1976) *Biochem. J. 153*, 271–277.

Tsuji, T., Irimura, T., & Osawa, T. (1980) *Biochem. J. 187*, 677–686.