RED CELL MEMBRANE GLYCOPROTEIN*: AMINO ACID SEQUENCE OF AN INTRAMEMBRANOUS REGION

Jere P. Segrest, Richard L. Jackson** and Vincent T. Marchesi
Laboratory of Experimental Pathology, NIAMDD

Robert B. Guyer and William Terry
Immunology Branch, NCI

National Institutes of Health, Bethesda, Maryland 20014

Received September 27,1972

SUMMARY: Human erythrocyte glycophorin has been purified and partially sequenced. The portion of the sequence reported here (51 residues) represents a unique region of glycophorin that, on the basis of labeling experiments reported elsewhere, is probably associated with the hydrophobic interior of the red cell membrane. This sequence has the anticipated hydrophobic characteristics for interaction with the hydrocarbon interior of a membrane and, as far as we are aware, represents the first reported sequence from a hydrophobic region of an integral membrane protein. There is a distribution of charged and hydrophobic residues topographically similar to a cross-section of a phospholipid bilayer, a structure many now accept as the basic feature of a biological membrane. We propose that the hydrophobic sequence of approximately 23 residues comprises the intramembranous domain of human erythrocyte glycophorin and is intimately and possibly specifically associated with lipids of the membrane.

Membrane glycoproteins have received increasing attention recently because of a growing awareness of the involvement of these proteins in processes of cellular recognition[2,3,4,5]. Cellular recognition or its failure appears to play a role in such phenomena as embryological development[3] and oncogenesis[6].

Probably the most extensively studied of membrane glycoproteins is that isolated from red blood cells (erythrocyte glycophorin[1]) by a variety of techniques[7,8,9]. This glycoprotein contains many receptors,[1] and may serve as a model for other cell surface glycoproteins.

In addition to its receptor-bearing function, a major feature of erythrocyte glycophorin, probably shared by other membrane glycoproteins, is the tenaci of its binding to the cell membrane, requiring the use of vigorous solvents or detergents for solubilization[10]. Membrane proteins with this characteristic have been called integral proteins by Singer,[11] and it has been suggested that such proteins play a role in membrane structural integrity by pene-

---

**Present Address: Department of Internal Medicine, Baylor College of Medicine and the Methodist Hospital, Houston, Texas

*Human erythrocyte glycophorin[1]

trating at least partially into the hydrophobic interior of the cell membrane[7,8,11]. It has been anticipated that that portion of an integral protein in the domain of the membrane interior would contain unique features, such as an amino acid sequence reflecting the hydrophobic nature of its environment.

Human erythrocyte glycophorin has been purified[12] and partially sequenced. The portion of the sequence reported here represents a unique region of glycophorin that is probably associated with the hydrophobic interior of the red cell membrane. This sequence has the anticipated hydrophobic characteristics for interaction with the hydrocarbon interior of a membrane and, as far as we are aware, represents the first reported sequence from a hydrophobic region of an integral membrane protein.

METHODS AND RESULTS: Glycophorin, isolated from human red cell ghosts by the lithium diiodosalicylate (LIS) procedure,[12] is approximately 60% carbohydrate and 40% protein by weight and appears to have a monomeric molecular weight of around 50,000[1,13]. However, the exact size of glycophorin is not known with certainty since the presence of large amounts of carbohydrate complicates analytical studies[13].

Cyanogen bromide (CNBr) cleavage produces five fragments,[1,14] but only three of these fragments have been identified as unique portions of a single polypeptide chain by overlap with tryptic peptides and by partial amino acid sequence analysis[14,15]. Two, designated C-5 and C-1, are sialoglycopeptides which form the N-terminal half of the molecule. The third, C-2, is the C-terminal fragment. The remaining fragments were isolated in low yields, and it is not yet clear whether they are unique portions of the same polypeptide chain[14]. However, the uncertainty as to the molecular weight of the monomer and the exact number of CNBr fragments does not affect the interpretation of the results reported here.

One of the overlapping tryptic peptides was found to have special properties. When isolated erythrocyte glycophorin is treated with trypsin for 24 hours at 37°C, an insoluble precipitate is formed[8,14]. This precipitate, after washing several times in 15% formic acid, gives a single band on polyacrylamide gel electrophoresis in sodium dodecyl sulfate (SDS) and has an apparent molecular weight of approximately 4,000[14]. This insoluble peptide, T(is), was found by amino acid analysis to have a high content of non-polar amino acids and a single methionine[14].

The C-terminal CNBr fragment, designated C-2, also has a high content of hydrophobic amino acids,[14] suggesting that C-2 and T(is) overlap; this proved to be the case, as shown by partial amino acid sequence analysis.

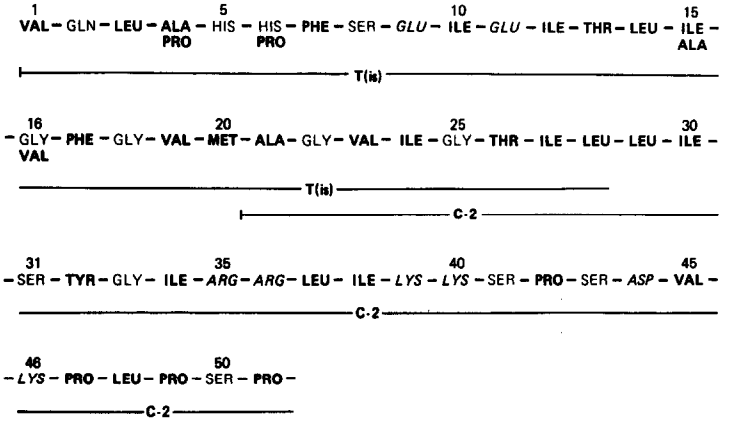

Figure 1 Amino acid sequence analysis of the intramembranous domain of human erythrocyte glycophorin. The tryptic peptide T(is) and CNBr fragment C-2 were subjected to amino acid sequence analysis by the Edman procedure utilizing an automated Beckman Sequenator. The PTH-amino acid derivatives were analyzed by gas chromatography before and after sialylation on a Varian series 1800 gas chromatograph and after acid hydrolysis with HI and HCl by amino acid analysis on a Beckman 121 automated analyzer. Yields based on amino acid analysis for the first step of the Edman degradation were approximately 25% for T(is) and 55% for C-2. For T(is) it was necessary to select a major residue at each step from an elevated background (see Table 1). The background was low enough in one preparation (less than 10%) to make it easy, in general, to distinguish major residues, but, at four positions, 4,6, 15 and 16, the background level was high and for this reason possible alternate residues are listed below the major residue. In each of these positions polypeptide microheterogeneity[14] is the most likely explanation for the low and nearly equal yields of two residues. The low degree of background in the overlapping fragment, C-2, supports the interpretation that background in T(is) is due primarily to contamination. The portion of the sequence of 51 residues obtained from the combined fragments is indicated, together with an eight residue overlap. Amino acid residues which are more hydrophobic than glycine according to the criteria of Tanford[16] are indicated by bold-faced type (Thr, Ala, Pro, Tyr, Val, Met, Leu, Ile, Trp, Phe). This is an operational definition and includes residues (e.g., Thr) that some investigators might not consider hydrophobic. Charged residues (Asp, Glu, Lys, Arg) are in italics and others (Gln, Asn, His, Ser, Gly) are in normal case. See Table 1 for a listing of significant background residues for each step in the analysis of peptide T(is).

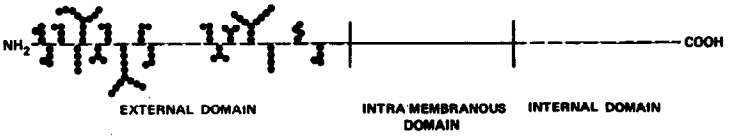

Figure 2 Schematic diagram of the glycophorin molecule which illustrates the relative positions of the three chemically distinct divisions of the monomeric polypeptide chain. Approximate locations of carbohydrate side chains are represented by chains of dots.

Figure 1 shows the results of these sequence analyses. Note the overlap of 8 residues found between T(is) and C-2. Hydrophobic residues[16] are represented by bold face type and charged residues (asp, glu, lys, arg) by italics. There are two prominent aspects to this sequence. (a) The sequence is extremely hydrophobic being particularly rich in leucine and isoleucine. No charged residues occur between residue 11 and residue 35 which represents a stretch of 23 amino acids with a particularly high density of hydrophobic residues. This region has been referred to elsewhere as the hydrophobic domain[14] of glycophorin. (b) There is a clustering of charged residues at both ends of the 23 residue domain. These features together represent a distribution of charged and hydrophobic residues topographically similar to a cross-section of a phospholipid bilayer, a structure many now accept as the basic feature of a biological membrane[11].

DISCUSSION: The approximate location of the hydrophobic domain along the polypeptide backbone of glycophorin is known from the alignment and size of the CNBr fragments[14]. Based upon the amino acid and carbohydrate composition of these fragments,[14] a unique "molecular topography" has been proposed for human erythrocyte glycophorin[14] and is shown in Figure 2. The main feature is that glycophorin can be divided into three chemically distinct regions, an N-terminal half containing all of the carbohydrate and receptor sites, the hydrophobic domain itself, and a C-terminal portion which is hydrophilic and rich in proline and contains no detectable carbohydrate[14].

Bretscher has published evidence that the major glycoprotein of the human red cell membrane (glycophorin) spans the full thickness of the membrane[17]. We have shown by labeling experiments reported elsewhere that the receptor domain of glycophorin is exposed to the external aqueous environment of the cell while the C-terminal segment is exposed to the cytoplasm of the cell[14]. This finding suggests strongly that the hydrophobic domain is that portion of the molecule which passes through the center of the membrane and represents an intramembranous segment of this molecule.

If this portion of the molecule is in an α-helical conformation while in situ, a stretch of 23 residues would be about 35 Å long, a distance compatible with most estimates of the thickness of the hydrocarbon portion of a membrane lipid bilayer[18]. Arguments in favor of the idea that the intramembranous portion of this molecule is in a helical conformation while in the membrane are described elsewhere[14].

We propose, then, that the hydrophobic sequence of approximately 23 residues comprises the intramembranous domain of human erythrocyte

Table 1. Significant Amino Acid Residues Encountered at Each Step of the Automated Sequence Analysis of Tryptic Peptide T(is)

Step	Residue* Major	Minor
1	Val	Phe
2	Gln	Thr
3	Leu	Val
4	Ala,Pro	
5	His	Ala
6	His,Pro	
7	Phe	Val
8	Ser	Val
9	Glu	Ile
10	Ile	Pro
11	Glu	Leu
12	Ile	
13	Thr	Val
14	Leu	
15	Ile,Ala	
16	Gly,Val	Ala
17	Phe	Val
18	Gly	Ala
19	Val	Ala
20	Met	Ala
21	Ala	
22	Gly	
23	Val	
24	Ile	
25	Gly	
26	Thr	
27	Ile	
28	Leu	

*Clearly identified major residues at each step are listed under the major residue heading. In the three steps (6,15,16) that no major residue is obvious, two residues have been listed in the major residue column. Residues not clearly identified as major that are listed in this column are underlined. Potentially significant background residues at each step are included under a minor residue column, except for the case of the overlap region between T(is) and C-2 (steps 21-28), where no possibility of ambiguity exists because of the very low level of background to C-2. Note that all the background residues are hydrophobic,[16] so that even if errors have occurred in assignment of the sequence shown in Figure 1, the physicochemical nature of the "hydrophobic domain" will be unaltered.

glycophorin and is intimately and possibly specifically[19] associated with lipids of the membrane.

REFERENCES

1. Marchesi, V. T., Tillack, T. W., Jackson, R. L., Segrest, J. P., and Scott, R. E., Proc. US Nat. Acad. Sci., 69, 1445 (1972).
2. Gesner, B. M., and Ginsburg, V., Proc. US Nat. Acad. Sci., 52, 750 (1964).
3. Margoliash, G., Schenk, J. R., Margie, M. P., Burokas, S., Richter, W. R., Barlow, G. H., and Moscona, A. A., Biochem. Biophys. Res. Comm., 20, 283 (1965).

4. Spiro, R. G., Ann. Rec. Biochem., 39, 732 (1970).
5. Dische, Z., in Glycoproteins of Blood Cells and Plasma, G. Jamieson and T. J. Greenwalt (editors), (J. B. Lippincott Co., Philadelphia, 1971).
6. Janson, V. K., and Burger, M. M., in Glycoproteins of Blood Cells and Plasma, G. Jamieson and T. J. Greenwalt (editors), (J. B. Lippincott Co., Philadelphia, 1971).
7. Morawiecki, A., Biochim. Biophys. Acta, 83, 339 (1964).
8. Winzler, R. J.,in Red Cell Membrane: Structure and Function, G. A. Jamieson, and T. J. Greenwalt (editors), (J. B. Lippincott, Philadelphia, p. 157, (1969).
9. Zvilichovsky, B., Gallop, P. M., and Blumenfeld, O. O., Biochem. Biophys. Res. Comm., 44, 1234 (1971).
10. Fairbanks, G., Steck, T. L., and Wallach, D. F. H., Biochemistry, 10, 2606 (1971).
11. Singer, S. J., in Structure and Function of Biological Membranes,L. I. Rothfield (ed), (Academic Press, New York, p. 145, 1971).
12. Marchesi, V. T., and Andrews, E. P., Science, 174, 1247 (1971).
13. Segrest, J. P., Jackson, R. L., Andrews, E. P., and Marchesi, V. T., Biochem. Biophys. Res. Comm., 44, 390 (1971).
14. Segrest, J. P., Kahane, I., Jackson, R. L., and Marchesi, V. T., Manuscript submitted for publication.
15. Jackson, R. L., Segrest, J. P., and Marchesi, V. T., Fed. Proc., 30, 1280, abstract (1971) and manuscript in preparation.
16. Tanford, C., J. Amer. Chem. Soc., 84, 4240 (1962).
17. Bretscher, M. S., Nature, 231, 229 (1971).
18. Urry, D. W., Goodall, M. C., Glickson, J. D., and Mayers, D. F., Proc. US Nat. Acad. Sci., 68, 1907 (1971).
19. Segrest, J. P., Manuscript in preparation.