Proc. Natl. Acad. Sci. USA
Vol. 74, No. 6, pp. 2301–2305, June 1977
Biochemistry

Location and structural significance of the oligosaccharides in human IgA1 and IgA2 immunoglobulins

(carbohydrate/A2m allotypes/protein structure/amino acid sequence/evolution)

ALFREDO TORAÑO*, YASUHARU TSUZUKIDA†, YU-SHENG VICTOR LIU, AND FRANK W. PUTNAM‡

Department of Biology, Indiana University, Bloomington, Indiana 47401

Contributed by Frank W. Putnam, March 18, 1977

ABSTRACT **The location, number, and kinds of oligosaccharides in human IgA1 and IgA2 immunoglobulins have been determined by amino acid sequence analysis of the α heavy chains. Both A2m allotypes of the α2 chain of IgA2 have two GlcN oligosaccharides that are absent in the α1 chain, but they lack GalN. The A2m(2) allotype has a fifth GlcN oligosaccharide. The α chains of IgA proteins also have subclass-specific and allotype-specific differences in amino acid sequence. Although other classes of human immunoglobulins differ in the number and kind of oligosaccharides, the sites are often homologous and are related to the immunoglobulin domain structure. Evolutionary preservation of the tripeptide acceptor sequence for GlcN probably indicates both a structural and biological role for carbohydrate.**

Although the five classes of human immunoglobulins differ greatly in their content of carbohydrate and in its distribution along the polypeptide chain, the carbohydrate often is in homologous positions and usually seems to lie in between the compact domains of the heavy chain or on the surface of a domain. With rare exceptions governed by the presence of a signal acceptor tripeptide sequence, the carbohydrate is present only on the heavy chain and is confined to the constant (C) region of the heavy chain. We earlier reported the location and nature of the five oligosaccharides of the μ chain of human IgM (1, 2). In contrast, the γ chain of human IgG has a single oligosaccharide (3), and the ε chain of human IgE is reported to have six (4). In all these cases the oligosaccharide contains glucosamine (GlcN), which is attached to asparagine by an *N*-glycosidic linkage. The IgA1 subclass is unusual among glycoproteins in having two types of linkage to the polypeptide chain, the *N*-glycosidic linkage of GlcN to asparagine and an *O*-glycosidic linkage of galactosamine (GalN) to serine (5, 6). Recently we reported (7) the complete amino acid sequence of a human IgA1 immunoglobulin (designated Bur), including the location of five GalN carbohydrates in the hinge region, two GlcN oligosaccharides in the Fc region, and an adventitious GlcN oligosaccharide in the variable (V) region of this α1 heavy chain. We have now determined the amino acid sequence of the C region of two α2 heavy chains representing different allotypes of the IgA2 subclass, namely, the A2m(2) allotype (IgA2 protein But) and the A2m(1) allotype (IgA2 protein Lan). Both α2 chains lack galactosamine but share the two GlcN oligosaccharides present in the α1 chain. Both A2m allotypes have two additional GlcN oligosaccharides lacking in the α1 chain, and the A2m(2) allotype has a fifth GlcN oligosaccharide missing in the A2m(1) allotype and in the α1 chain. Despite the apparently disparate distribution of carbohydrate in immunoglobulin classes, subclasses, and allotypes, attachment of GlcN oligosaccharides is dictated by the presence of the tripeptide acceptor sequence Asn-X-Thr/Ser (where X represents any amino acid except possibly proline, and Ser/Thr means serine or threonine) (8). This signal sequence appears in homologous positions in a number of human heavy chains, and also in some animal heavy chains. The serological determinants of the IgA1 and IgA2 subclasses and of the A2m allotypes probably do not lie in the differences of carbohydrate, but rather reside in differences in amino acid sequence that are characteristic of the subclass and allotype.

MATERIALS AND METHODS

Human IgA was prepared from the serum of patients with multiple myeloma who produced large amounts of monoclonal IgA. The procedure involved ammonium sulfate precipitation, ion exchange chromatography, and gel filtration techniques (9, 10). The purity of the IgA was tested by immunodiffusion and immunoelectrophoresis with monospecific antisera and by gel electrophoresis and ultracentrifugation (9, 10). The purified IgA proteins were reduced at room temperature under N_2 with 0.01 M dithiothreitol for 1 hr followed by alkylation with iodoacetic acid (10% molar excess) or ethylene imine. The heavy and light chains were separated on a Sephadex G-100 column. The reduced and alkylated α heavy chains were digested with trypsin, α chymotrypsin, or thermolysin. The resulting peptides were isolated by gel filtration on Sephadex G-25 or G-50, by ion exchange chromatography, and by high-voltage paper electrophoresis. Amino acid analyses were done with the Beckman amino acid analyzer (model 120B and 120C). Hexose determination was done with sulfuric acid and phenol (11). For quantitative determination of hexosamine and tryptophan, the peptides were hydrolyzed with 3 M *p*-toluenesulfonic acid and analyzed with the amino acid analyzer (12). The sequences of some small tryptic, chymotryptic, and thermolytic peptides were determined by manual Edman degradation and dansylation techniques (13). For most peptides, sequential Edman degradation was performed in a Beckman sequencer (model 890B, updated). Since significant improvements of sequence methodology were continuously reported during the period of this investigation, not all peptides were automatically sequenced in the same fashion. The programs used in this study were a fast peptide dimethylallylamine program (090872) and a modified program (111374); both were provided by the Beckman Co. The phenylthiohydantoin derivatives of the amino acids were identified by gas chromatography (14), amino acid analysis (15), or thin-layer chromatography (16).

Abbreviations: Abbreviations for classes, fragments, regions, and domains of immunoglobulins accord with official WHO recommendations for human immunoglobulins published in *Biochemistry* (1972) 11, 3311–3312.

* Present address: Department of Zoology, University College, London WC1E 6BT, United Kingdom.
† Present address: Research Institute for Atomic Energy, Osaka University, Osaka 558, Japan.
‡ To whom reprint requests should be addressed.

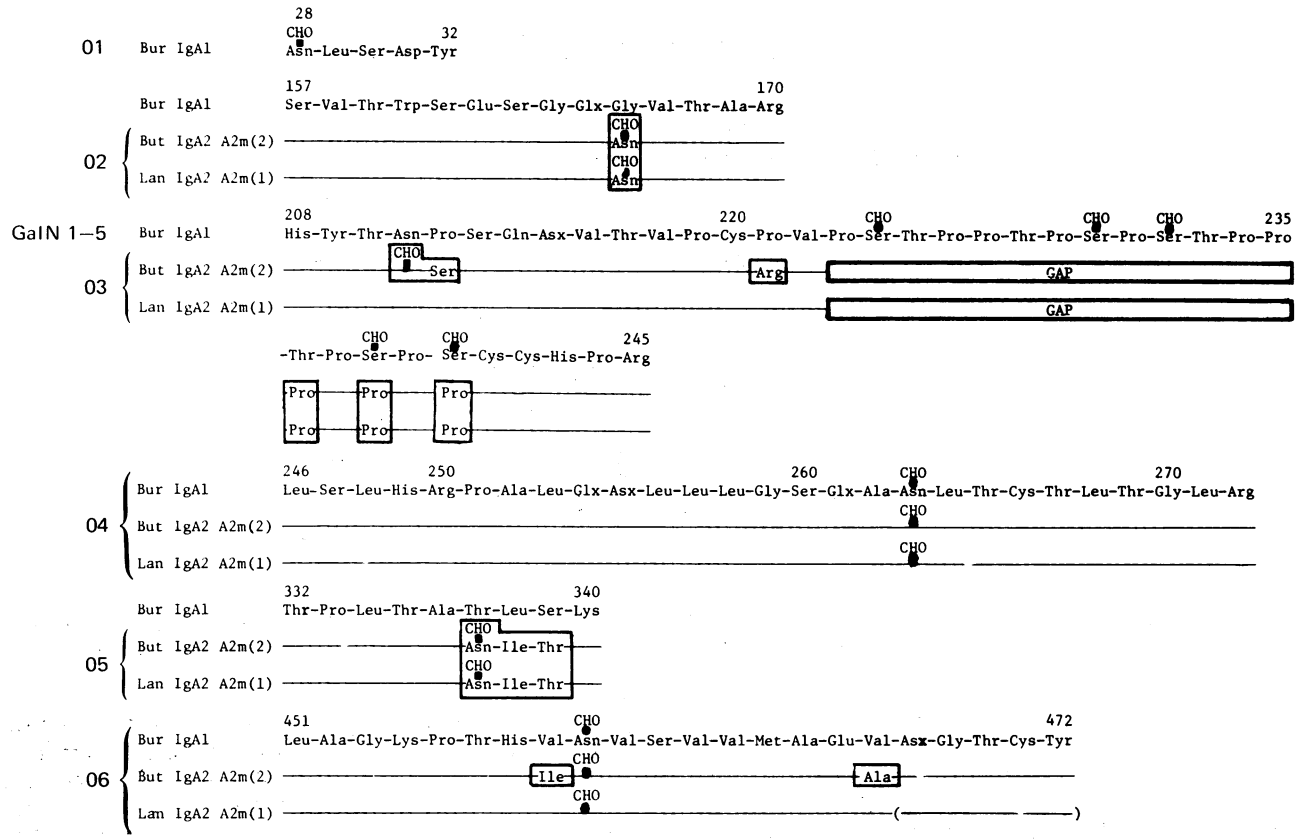

FIG. 1. Amino acid sequence of glycopeptides and location of carbohydrate (CHO) in subclasses and allotypes of human immunoglobulin A. The sequences are numbered according to the numbering system of the IgA1 protein Bur (7). Proteins Lan and But were typed as the A2m(1) and A2m(2) allotypes, respectively. A horizontal line indicates the A2m allotype sequences are identical to IgA1; differences are denoted by residues enclosed in boxes. The gap in the IgA2 sequences represents the deletion of 13 residues (223–235) in the hinge region of the $\alpha 2$ chains; however, the same numbering system is used for the $\alpha 1$ and $\alpha 2$ chains to facilitate their comparison. The symbols 01–06 in the left margin designate the six GlcN oligosaccharides. GalN 1–5 denotes the five GalN oligosaccharides in the $\alpha 1$ hinge region.

RESULTS

The location of the carbohydrates in the $\alpha 1$ and $\alpha 2$ chains was established by isolation of and determination of the sequence of the glycopeptides from the tryptic, chymotryptic, and thermolytic digests and by placement of the glycopeptides in the composite sequence which included many overlapping peptides. Some of the results have been documented (7, 9, 10, 17), and articles giving the complete proof of the $\alpha 1$ and $\alpha 2$ chain sequences are in preparation. To clarify presentation for the three proteins we summarize the results of the analysis of sequences containing glycopeptides in Fig. 1, and show the locations of the oligosaccharides in the polypeptide chain structure schematically in Fig. 2. Although $\alpha 1$ and $\alpha 2$ chains differ in length because of the deletion in the $\alpha 2$ hinge region, the numbering system for the C region of the $\alpha 1$ chain is used for homologous residues in the $\alpha 2$ chain in order to facilitate comparison.

Carbohydrate is not a usual characteristic of the V region of heavy chains, but a GlcN oligosaccharide designated 01 is linked to Asn-28 in the V region of the Bur $\alpha 1$ chain (Fig. 1). Attachment of this adventitious carbohydrate is due to the recognition by glycosyltransferases of the signal acceptor sequence Asn-Leu-Ser that occurs by chance in the V region of the Bur $\alpha 1$ chain.

A subclass-specific change from Gly-166 in the $\alpha 1$ chain to Asn-166 in both $\alpha 2$ chains generates a new tripeptide acceptor sequence, Asn-Val-Thr, to which the glucosamine oligosaccharide denoted 02 is linked. As seen in Fig. 2, 02 occurs within the first constant region domain of the $\alpha 2$ chain.

Identification of the asparagine to which oligosaccharide 03 is linked was difficult because of the repetition of two potential acceptor sites in the sequence Asn-Pro-Ser-Gln-Asx-Val-Thr in the $\alpha 1$ chain and in the A2m(1) allotype and the mutation to Asn-Ser-Ser-Gln-Asx-Val-Thr in the A2m(2) allotype. However, no glucosamine was detected in this sequence in the $\alpha 1$ chain or in the A2m(1) allotype, whereas the peptide from the A2m(2) allotype contained glucosamine. Proof that the carbohydrate is linked to Asn-211 rather than to Asx-215 in the A2m(2) allotype was obtained by repeated analysis with the automatic sequencer; this showed a blank at Asx-211, but aspartic acid was regenerated at Asx-215 by acid hydrolysis. In automatic sequencing of glycopeptides with Asn-linked carbohydrate, no recovery of asparagine is obtained. Furthermore, staphylococcal protease, when used under conditions specific for cleavage after aspartic acid and glutamic acid (18), cleaved this peptide after Asp-215 in the A2m(2) allotype. Since this established that Asp-215 was in the acid rather than the amide form, carbohydrate could not be attached to this position.

The five galactosamine-containing oligosaccharides that are characteristic of the $\alpha 1$ hinge region are missing in both $\alpha 2$ allotypes because of the deletion in the hinge region. Here both $\alpha 2$ allotypes lack the repeating octapeptide sequence Pro-Ser-Thr-Pro-Pro-Thr-Pro-Ser in which galactosamine is attached *N*-glycosidically to the serine residues, as well as to the fifth serine in the hinge. In place of this duplication, the $\alpha 2$ allotypes have an unusual pentaproline sequence Val-Pro-Pro-Pro-Pro-Pro-Cys-Cys, which must greatly affect the conformation of the $\alpha 2$ hinge. Placement of the GalN oligosac-

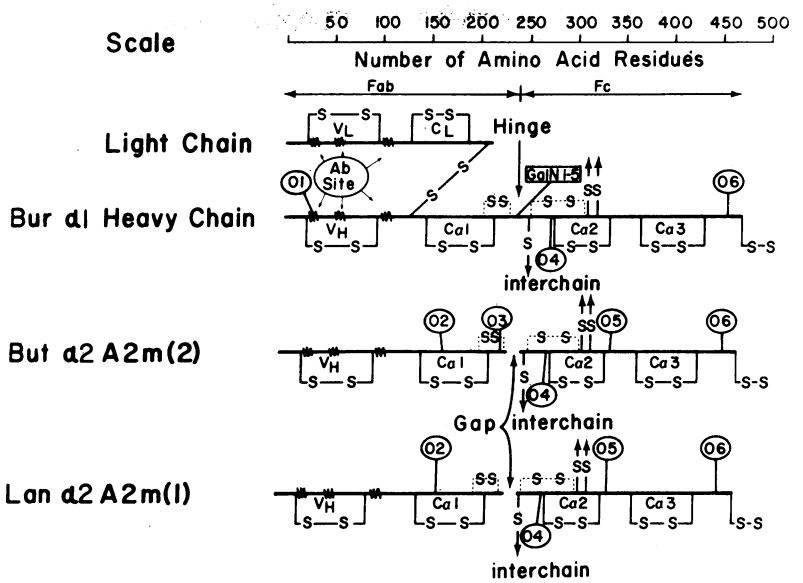

FIG. 2. Schematic structural models of a light chain and of the $\alpha 1$ and $\alpha 2$ heavy chains drawn to scale (*Top*). The disulfide-bridged domains are identified by standard symbols. The location is shown for the six GlcN oligosaccharides (01–06) and the five GalN oligosaccharides (GalN 1–5) identified in Fig. 1.

charide units in the $\alpha 1$ hinge was done by Baenziger and Kornfeld (5, 6).

The GlcN oligosaccharide designated 04 is common to both $\alpha 1$ and $\alpha 2$ chains. This leucine-rich glycopeptide was first reported by Moore and Putnam (17). In contrast, the GlcN oligosaccharide designated 05 is characteristic of the $\alpha 2$ chains. The glycopeptide was identified in three $\alpha 2$ chains by Despont and Abel (19), but the allotype of the IgA2 proteins was not specified. Attachment of carbohydrate 05 between the C_H2 and C_H3 domains (Fig. 2) appears to be characteristic of the $\alpha 2$ chain; the linkage results from the substitution of the acceptor triplet Asn-Ile-Thr in $\alpha 2$ chains for the sequence Thr-Leu-Ser in $\alpha 1$ chains. With the exception of the hinge deletion, this is the most prominent subclass-specific difference between $\alpha 1$ and $\alpha 2$ chains, for it involves substitution of three consecutive amino acids as well as addition of carbohydrate.

The GlcN oligosaccharide 06, which is in the COOH-terminal tail of both subclasses of α chains, has a counterpart in the COOH-terminal tail of the μ chain. Two amino acid substitutions occur in this region in the A2m(2) allotype protein. Besides the fact that the corresponding residues are identical in the $\alpha 1$ chain and in the A2m(1) allotype, we have additional unpublished evidence that these substitutions contribute structurally to the allotypic determinants identified serologically by Loghem *et al.* (20).

DISCUSSION

Although other workers had measured the carbohydrate composition of human IgA immunoglobulins (21) and had identified some of the glycopeptides (5–7, 17, 19, 22), in this report all of the glycopeptides of an IgA1 molecule and of both A2m allotypes of IgA2 have been isolated, their sequences have been determined, and they have been placed in the primary structure of the α chains. We believe these results validly represent the oligosaccharide distribution in the IgA1 subclass and in the two allotypes of the IgA2 subclass. Although the complete amino acid sequence of only a single protein representative of each type was determined, we (7) and other workers (5, 6, 19) have identified all of the characteristic glycopeptides of the $\alpha 1$ chain in other IgA1 proteins. Two of the $\alpha 2$ glycopeptides (05 and 06) were also identified by Despont and Abel (19).

Although it has been suggested (19) that the glycopeptides of the α chain do not seem homologous to those of other heavy chains, significant homology does appear when the chain structures are aligned with the deletion of the extra domain of the μ and ϵ chains (Fig. 3). There is no counterpart in IgG, IgM, and IgE of the IgA1 galactosamine oligosaccharides (although there may be in IgD). There are at least 11 locations for GlcN carbohydrate in the C region of the four classes of human heavy chains whose sequences have been determined (γ, μ, ϵ, and α). However, some of the oligosaccharides are in almost exactly homologous positions despite the fact that the γ chain has only one oligosaccharide, the μ chain has five, the ϵ has six, and α chains have two, four, or five according to the subclass and allotype. Because of differences in chain length, the homology is not at first apparent from the position of the oligosaccharides in the linear sequence. However, when the chains are aligned as in Fig. 3, there are seven sets of homologous positions. These are: (*i*) within the C_H1 domain ($\alpha 2$, μ, and ϵ); (*ii*) just before the hinge [$\alpha 2$ A2m(2) and ϵ]; (*iii*) within the hinge (GlcN in μ and GalN in $\alpha 1$); (*iv*) just inside the first domain of Fc (04 in $\alpha 1$ and $\alpha 2$); (*v*) 40 residues down the first domain of Fc ($\gamma 1$, μ, and ϵ); (*vi*) between the first and second domains of Fc ($\alpha 2$ allotypes); and (*vii*) 14 residues from the COOH-terminus of the extended chain ($\alpha 1$, $\alpha 2$, and μ).

With the exception of the Asn-Pro-Ser sequence in $\alpha 1$ and in the $\alpha 2$ A2m(1) allotype, all Asn-X-Ser/Thr acceptor sequences in the primary structures published for human α, γ, μ, and ϵ C regions have GlcN carbohydrate attached. An acceptor sequence with carbohydrate attached is conserved in comparable classes of animal immunoglobulins. For example, in animal species the complete amino acid sequence of only the C region of the γ chain has been determined, but the single carbohydrate present is linked to the same position (Asn-297) as in the human γ chain, i.e., d11 allotype of rabbit γ (23), guinea pig antibody $\gamma 2$ (24), and mouse MOPC 173 myeloma $\gamma 2a$ chain (25). This highly conserved carbohydrate is missing in the human $\alpha 1$ and $\alpha 2$ chains because the homologous position lacks the acceptor sequence, which is replaced by a Cys-Gly-Cys sequence, leading to a highly strained conformation (22). These cysteines must be on the surface of the molecule for two reasons: (*i*) they are complementary to the $\gamma 1$ carbohydrate, which is

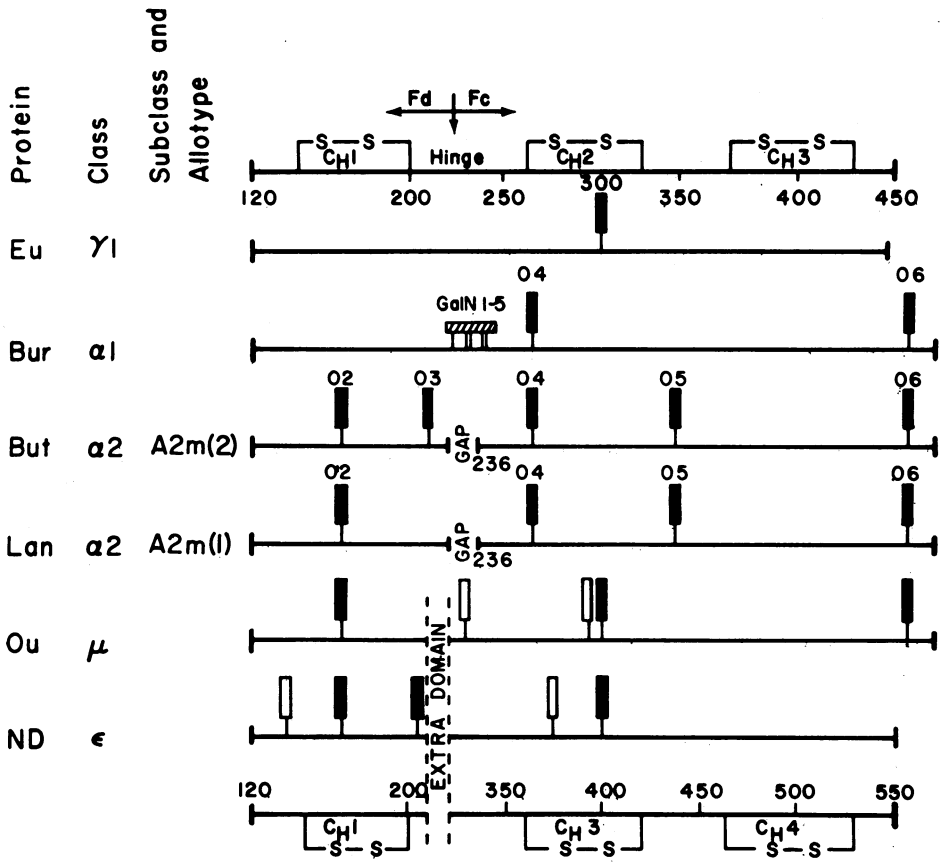

FIG. 3. Location of carbohydrate in human γ1, α1, α2, μ, and ε heavy chains. The vertical rectangles signify glucosamine oligosaccharides. Shading indicates those that have homologous positions in two or more chains. The numbers in the upper scale give the residue positions for the γ and α chains, those in the lower scale for the μ and ε chains. To facilitate comparison, the extra domain ($C_\mu 2$ and $C_\varepsilon 2$, respectively) has been omitted in the μ and ε chains. An oligosaccharide is present in $C_\varepsilon 2$, but not in $C_\mu 2$. The symbols 02–06 and GalN 1–5 denote α chain oligosaccharides identified in Fig. 1.

on the surface of the C_H2 domain according to crystallographic analysis of human Fc (26); and (*ii*) one of the α chain cysteines (Cys-301) makes an interchain bridge, and the other (Cys-299) makes an "extra" intrachain bridge to Cys-242. The latter is next to another interchain bridge (Cys-241), and these two cysteines are complementary to residues covered by carbohydrate in the γ1 chain (Phe-241 and Phe-243) (26). Thus, structural differences at this site probably contribute to major functional differences of the γ1 and α chains, such as the ability to bind complement.

In glycoproteins the Asn-Pro-Ser sequence is rarely glycosidated probably because the proline may interfere sterically (21). It is noteworthy that in the human α chains the Asn-Ser-Ser sequence of the A2m(2) allotype (which is glycosidated) and the Asn-Pro-Ser sequences of the A2m(1) allotype and the α1 chain (which are not glycosidated) all occur close to the proline-rich hinge region. However, Pro-221 just in front of the hinge is changed to arginine in the A2m(2) allotype. This substitution may impart a conformational change sufficient to expose the Asn-Ser-Ser sequence to glycosidation by the enzyme. The sequence Asn-X-Ser/Thr has low potential for the α-helix or β-sheet conformations and a high potential for a bend or turn (27). Thus, this tripeptide sequence tends to place carbohydrate at the surface.

Several possible roles have been proposed for the carbohydrate in immunoglobulins, but none has been established. It has been suggested (8, 21, 28) that carbohydrates increase solubility, facilitate secretion, and act as spacers between the domains. The presence of carbohydrate at homologous positions in human and animal immunoglobulins not only indicates that the acceptor sequence has been conserved in the face of great changes in the primary structure of immunoglobulins of different classes and species, but also suggests that evolutionary retention of carbohydrate has structural and functional significance.

We thank J. Dwulet, E. Dodge, J. Madison, and S. Dorwin for technical assistance, Dr. E. Osserman for plasma from patient Bur, Dr. G. M. Bernier for plasma from patient But, and Drs. H. G. Kunkel and M. S. Schanfield for serological typing of the IgA proteins. This work was supported by Grants CA08497 of the National Cancer Institute and IM-2D of the American Cancer Society, and by a fellowship to A.T. from the Fundación del Instituto de Industria, Madrid, Spain.

The costs of publication of this article were defrayed in part by the payment of page charges from funds made available to support the research which is the subject of the article. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

1. Shimizu, A., Putnam, F. W., Paul, C., Clamp, J. R. & Johnson, I. (1971) *Nature New Biol.* **231**, 73–76.
2. Putnam, F. W., Florent, G., Paul, C., Shinoda, T. & Shimizu, A. (1973) *Science* **182**, 287–291.
3. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 78–85.
4. Bennich, H. & Bahr-Lindström, H. von (1974) in *Progress in Immunology II*, eds. Brent, L. & Holborow, J. (North Holland Publishing Co., Amsterdam), Vol. 1, pp. 49–58.
5. Baenziger, J. & Kornfeld, S. (1974) *J. Biol. Chem.* **249**, 7260–7269.

6. Baenziger, J. & Kornfeld, S. (1974) *J. Biol. Chem.* **249,** 7270–7281.
7. Liu, Y.-S. V., Low, T. L. K., Infante, A. & Putnam, F. W. (1976) *Science* **193,** 1017–1020.
8. Marshall, R. D. (1972) *Annu. Rev. Biochem.* **41,** 673–702.
9. Putnam, F. W., Low, T., Liu, V., Huser, H., Raff, E., Wong, F. C. & Clamp, J. R. (1974) in *The Immunoglobulin A System,* eds. Mestecky, J. & Lawton, A. R. (Plenum Press, New York), pp. 177–189.
10. Liu, Y.-S. V. (1975) Doctoral dissertation, Indiana University.
11. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28,** 350–356.
12. Liu, T.-Y. & Chang, Y. H. (1971) *J. Biol. Chem.* **246,** 2842–2848.
13. Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89,** 379–380.
14. Pisano, J. J., Bronzert, T. J. & Brewer, H. B., Jr. (1972) *Anal. Biochem.* **45,** 43–59.
15. Mendez, E. & Lai, C. Y. (1975) *Anal. Biochem.* **68,** 47–53.
16. Inagami, T. & Murakami, K. (1972) *Anal. Biochem.* **47,** 501–504.
17. Moore, V. & Putnam, F. W. (1973) *Biochemistry* **12,** 2361–2370.
18. Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3506–3509.
19. Despont, J.-P. & Abel, C. A. (1974) *J. Immunol.* **112,** 1623–1627.
20. Loghem, E. van, Lange, G. de & Koistinen, J. (1976) *Scand. J. Immunol.* **5,** 161–164.
21. Clamp, J. R. (1975) in *The Plasma Proteins,* ed. Putnam, F. W. (Academic Press, New York), Vol. II, 2nd. ed., pp. 163–211.
22. Low, T. L. K., Liu, Y.-S. V. & Putnam, F. W. (1976) *Science* **191,** 390–392.
23. Hill, R. L., Delaney, R., Fellows, R. E., Jr. & Lebovitz, H. E. (1966) *Proc. Natl. Acad. Sci. USA* **56,** 1762–1769.
24. Tracey, D. E. & Cebra, J. J. (1974) *Biochemistry* **13,** 4796–4803.
25. Fougereau, M., Bourgois, A., Preval, C. de, Rocca-Serra, J. & Schiff, C. (1976) *Ann. Immunol. (Paris)* **127C,** 607–631.
26. Deisenhofer, J., Colman, P. M., Epp, O. & Huber, R. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357,** 1421–1434.
27. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–221.
28. Putnam, F. W. (1977) in *The Plasma Proteins,* ed. Putnam, F. W. (Academic Press, New York), Vol. III, 2nd. ed., in press.