Proc. Natl. Acad. Sci. USA
Vol. 75, No. 2, pp. 966–969, February 1978
Immunology

Complete amino acid sequence of the α2 heavy chain of a human IgA2 immunoglobulin of the A2m(2) allotype

(protein structure/genetic markers/isoallotype/evolution)

ALFREDO TORAÑO* AND FRANK W. PUTNAM†

Department of Biology, Indiana University, Bloomington, Indiana 47401

Contributed by Frank W. Putnam, December 5, 1977

ABSTRACT **The complete amino acid sequence of the α2 heavy chain of a human IgA2 immunoglobulin of the A2m(2) allotype has been determined and is compared to the sequence of the α1 chain of the human IgA1 subclass. The characteristic differences between the α1 and α2 chains are greatest in the hinge region and in the location and number of the oligosaccharides. Apart from the duplication in the hinge region of α1 and the deletion in α2, there are 23 amino acid exchanges in the constant (C) regions of the two chains. Accepted mutations are related to the surface accessibility of the residues and the proximity of carbohydrate. The results indicate that human IgA and IgG subclasses arose late in evolution and reflect similar mutational pressures.**

IgA, the characteristic antibody of secretory fluids and the second most abundant of the five classes of human immunoglobulins, normally exists as two well-defined subclasses, IgA1 and IgA2, that can be differentiated both chemically and antigenically (1, 2). By serological methods, a genetically controlled polymorphism (allotypy) has been identified in the IgA2 subclass. The two allotypes of IgA2 have been given various confusing designations but the recommended notation is now A2m(1) and A2m(2). These terms correspond to the notation $Am_2(+)$ and $Am_2(-)$ used earlier by us (3) and by others. Recently, an isoallotype of human IgA proteins, designated nA2m(2), was identified serologically by van Loghem *et al.* (4). Our objectives have been to determine the primary structure of human IgA1 and IgA2 proteins, including the complete amino acid sequence of their heavy chains (α1 and α2, respectively), to establish the location and kinds of oligosaccharides, to identify the genetically determined differences in structure of the allotypes, and to relate these to the biological function, three-dimensional structure, and evolution of the IgA class of molecules. We have previously reported the complete covalent structure of an IgA1 myeloma protein (designated Bur) (5), the comparison of the Fc structure of human IgA, IgG, IgM, and IgE (6), and the location and nature of the oligosaccharides in human α1 and α2 heavy chains as well as some amino acid substitutions characteristic of the subclasses and allotype of α chains (7).

We report here the complete amino acid sequence of the α2 heavy chain of a human A2m(2) myeloma protein (designated But). The frequency of this allotype is highest in Africa, is less in Australia and Asia, and is lowest in Europe (2). Subclasses and allotypes are also known for human IgG, but only partial sequences of the γ chains of the four human subclasses of IgG have been published; they have been summarized by Beale and Feinstein (8). Overall, the α1 chain and the α2 chain of the A2m(2) allotype are remarkably similar; the major differences are: (*i*) the partial duplication in the α1 hinge and the deletion in the α2 hinge already described (7, 9); (*ii*) the number and location of the oligosaccharides (7); and (*iii*) 23 amino acid substitutions outside the hinge region that are reported here.‡

MATERIALS AND METHODS

Human IgA2 was obtained from the serum of a patient of the black race (But) affected by multiple myeloma. Assignment of the immunoglobulin to the A2m(2) allotype was confirmed independently in three laboratories by serological assay. The IgA2 was isolated from the serum by ammonium sulfate precipitation, ion exchange chromatography with DEAE-Sephadex A-50, and gel filtration with Sephadex G-200 (3). The course of purification was monitored by immunoelectrophoresis with specific antisera, and the final purity was determined by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. Purified IgA2 was completely reduced with 0.01 M dithiothreitol followed by alkylation with iodoacetic acid or ethylenimine, and the light and heavy chains were isolated by gel filtration on Sephadex G-200 in 6 M urea/0.02 M formate buffer. The alkylated α2 heavy chain was digested with trypsin, chymotrypsin, or staphylococcal protease. The resulting peptides were purified by gel filtration on Sephadex G-50, ion exchange chromatography, and paper electrophoresis.

The methods for amino acid and carbohydrate analysis and for sequence determination by the manual Edman and dansylation techniques and by automatic sequence analysis with the Beckman sequencer were the same as those described by Toraño *et al.* (7). Secondary structure predictions were made by the Chou and Fasman method (11). Diagrams by Edmundson *et al.* (12) of the basic immunoglobulin fold have been used for the spatial location of the amino acid exchanges. The β-pleated sheet segments are indicated by the notation 3-1 to 3-3 and 4-1 to 4-4, followed by parentheses containing the

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: Abbreviations for classes, fragments, regions, and domains of immunoglobulins follow official World Health Organization recommendations for human immunoglobulins published in *Biochemistry* (1972) **11**, 3311–3312. For allotypes, the World Health Organization notation is used as published in *J. Immunol.* (1976) **117**, 1056–1058; C region, constant region; V region, variable region.

* Present address: Department of Zoology, University College, London WC1E 6BT, United Kingdom.

† To whom reprint requests should be addressed.

‡ At positions 282–284 there is an apparent reversal of the sequence Pro-Ser-Thr in IgA1 Bur (5) to Thr-Pro-Ser in IgA2 But. Although Liu *et al.* (5) have evidence for Pro-Ser-Thr in the α1 chain Bur, Kratzin *et al.* (10) have reported Thr-Pro-Ser at this position in the human α1 chain Tro. Because this discrepancy may reflect technical problems of analysis, it is not illustrated in Fig. 1. However, α1 Bur and α2 But are identical at five other positions where the sequence given for α1 Tro differs from that for α1 Bur—i.e., Glu-295, Pro-309, Thr-409, Arg-346, and His-350.

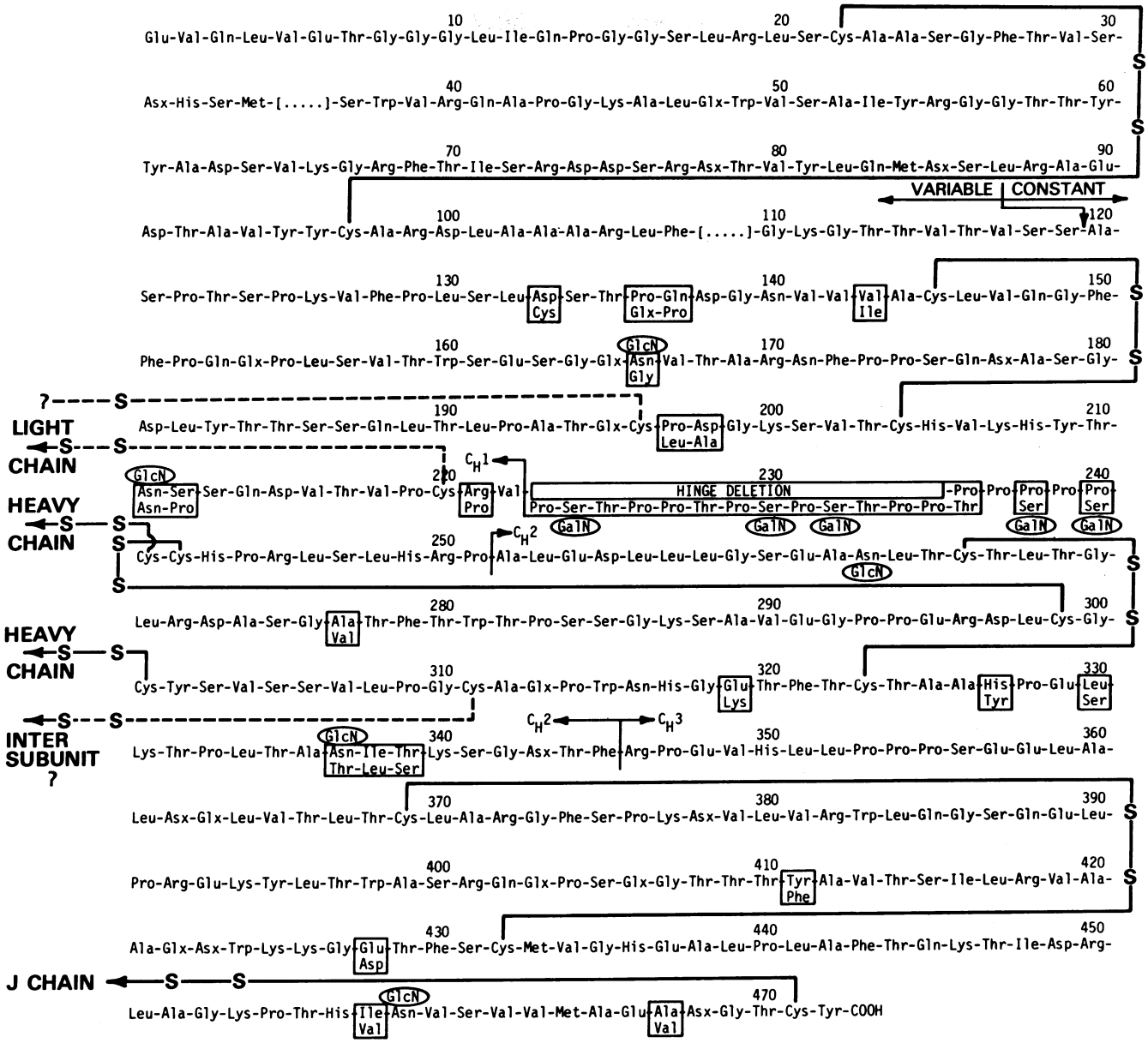

FIG. 1. Amino acid sequence of the $\alpha 2$ heavy chain of a human IgA2 protein (But) of the A2m(2) allotype. The $\alpha 2$ chain is given as a continuous sequence; amino acids that differ in the C region of the $\alpha 1$ chain of human IgA1 Bur (5) are shown in boxes below the $\alpha 2$ sequence.‡ To facilitate comparison, the same numbering system is used for the two α chain subclasses; this requires omission of numbers 223–235 in $\alpha 2$ because of the deletion in the hinge region (6, 10). Oligosaccharides containing glucosamine are identified by GlcN in ovals and those containing galactosamine, by GalN in ovals. The intrachain disulfide bridges characteristic of the domains are at the right, and the interchain bridges and other intrachain bridges are at the left. As indicated by the dotted lines, some of the disulfide bridge assignments are tentative.

residue numbers of the amino acids included in the segment. The structure presented is based on sequence analysis of 62 tryptic peptides, 42 chymotryptic peptides, and 14 staphylococcal protease peptides, supplemented by partial sequence analysis of the whole $\alpha 2$ chain and of CNBr fragments and by other ancillary data.

RESULTS AND DISCUSSION

The complete amino acid sequence of the A2m(2) heavy chain But is given in Fig. 1. To facilitate comparison of the two subclasses, the numbering system used for the amino acid residues in the constant (C) region is the same as that used for the $\alpha 1$ chain (5); this requires no adjustment of the two sequences except for omission of numbers 223–235 owing to the deletion of the hinge region in the $\alpha 2$ chain. The variable (V) regions of heavy chains of all classes vary in length, especially in the third hypervariable region just before the V/C switch region juncture at position 120 (1). In numbering the sequence of this $\alpha 2$ chain, two gaps of two residues each have been inserted at positions 35 and 36 and positions 108 and 109. Thus, the V region actually has 115 residues, and the C region has 340 residues for a total length of 455 residues instead of the 472 in $\alpha 1$ Bur and given in the numbering system.

The V Region. The V region sequence of this $\alpha 2$ chain belongs to the V_{HIII} subgroup and has no remarkable features. However, three highly conserved residues characteristic of the V_{HIII} subgroup have been substituted—i.e., Thr-7 in place of the hitherto invariant Ser-7, Ile-12 for the almost invariant Val-12, and Ala-46 for the almost invariant Gly-46. Also, the Gly-Lys-Gly sequence at positions 110–112 just before the V/C

switch replaces the Gly-Gln-Gly sequence previously invariant in all human V_{HIII} heavy chains.

The V/C Switch Region. The V/C switch region is the point of division between the V and C regions of an immunoglobulin chain. Because V regions are shared among the different classes of heavy chains (13), the V/C switch region has intrinsic interest and potential genetic significance in that it represents the putative point of union of V_H and C_H genes. As shown in Fig. 1, the C region of both α1 and α2 chains begins with the sequence Ala-Ser-Pro-Thr-Ser corresponding to Ala-120 in the α chains. The tripeptide Val-Ser-Ser, which ends the V region of α2 But, at first had appeared to be a recognition sequence common to the COOH terminus of the V_H region of all human heavy chains. Indeed, it is the COOH-terminal sequence of the V_H region of 13 of the 17 μ, γ, and α chains for which complete V_H sequences have been published (1). There is better than 90% homology among these heavy chains in the switch region; however, at least four substitutions have been identified. Other residues just before Val-Ser-Ser are also strongly conserved: for example, the invariant glycine five residues before it, and the dipeptide Val-Thr just before it. It is significant that the pentapeptide sequence Val-Thr-Val-Ser-Ser not only is present at the end of the V_H region of most human heavy chains but also appears in a comparable location in a number of animal heavy chains. By appropriate choice of codons, Fougereau *et al.* (14) deduced a DNA stretch containing rotational symmetry that is centered on the first serine in the sequence Val-Ser-Ser that precedes the switch point in many heavy chains. They proposed that this provides a recognition symbol for restriction-like enzymes, such as those that operate in prokaryotes. A recent report relating the pentapeptide Val-Thr-Val-Ser-Ser to a palindromic nucleotide sequence seems to support a similar type of recognition mechanism (15).

Disulfide Bridges. The interchain and intrachain disulfide bridges for this α2 chain are given tentatively in Fig. 1 and are placed largely by homology to previous reports for the bridging of α1 and α2 chains (5, 9, 16) because only preliminary experiments were done to confirm the arrangement. The greatest uncertainty is in the placement of the bridge between the heavy and light chains (H-L bridge) at Cys-220. Quite unexpectedly, we found aspartic acid rather than cysteine at position 133; Cys-133 is the site of the H-L bridge in α1 chains and has been reported to be the site in the IgA2 A2m(2) protein Rou for which partial sequences around some of the disulfide bridges have been reported (16). However, the latter protein was typed shortly after the discovery of the A2m allotypes when the antisera available were perhaps not as specific as they are now. [IgA2 was submitted as a coded sample to three independent laboratories and has been unambiguously typed as A2m(2) by all three.]

Unlike the A2m(1) allotype, in the A2m(2) allotype the H and L chains are covalently linked by a disulfide bond; hence, we have tentatively proposed Cys-220 as the site of linkage because this cysteine is exposed in the Fd portion just before the hinge region and is homologous to Cys-220, the site of the H-L bridge in human IgG1. However, this leaves no function for Cys-196, which is thought to form an intrachain bridge with Cys-220 in α1 and A2m(1) chains (5, 16). An alternative model would place the H-L bridge at Cys-196. In addition, it is noteworthy that A2m(1) and A2m(2) heavy chains do not differ in their cysteine content (16 residues) but do differ in the pattern of disulfide bond distribution. Because the sites for secretory piece and the intersubunit bridges also are not well substantiated, further work on the disulfide bridging pattern of α2 chains is needed.

The Hinge Region. The most notable difference between the α1 chain and the α2 chain of both A2m allotypes is in the hinge region. At positions 223–238 in the α1 hinge, there is a duplication of the octapeptide sequence Pro-Ser-Thr-Pro-Pro-Thr-Pro-Ser. This sequence is deleted in the α2 hinge of both A2m allotypes and is replaced by a remarkable pentaproline sequence that must greatly affect the conformation of the molecule. Streptococcal IgA protease cleaves IgA1 after the first Thr-Pro-Pro sequence in the hinge, and gonococcal IgA protease cleaves after the second Thr-Pro-Pro sequence, but neither cleaves IgA2 (17). Cleavage *in vivo* probably leads to more rapid metabolic degradation; so susceptibility to cleavage is disadvantageous in mucosal secretions such as intestinal fluids where IgA is the predominant antibody class. We do not know whether IgA2 evolved as an escape from bacterial proteases or whether the bacteria adapted to IgA1, which is resistant to cleavage by intestinal proteases such as trypsin and pepsin and also to papain.

We propose that characteristic properties of the α1 and α2 hinge regions may contribute to the inability of IgA to bind complement by blocking the allosteric signal that is thought to induce the correct contact of the C_H1 and C_H2 domains in immunoglobulins such as IgG that do bind Clq (18, 19).

Besides the remarkable content of proline that must certainly affect C_H1–C_H2 interactions, the unique features of the IgA hinge are the five GalN oligosaccharides in α1 and the deletion in α2 of 13 residues that include the segment containing GalN. Both features may contribute to the inability of intact IgA to bind complement. The deletion in IgA2 with substitution of a pentaproline sequence may interfere with transmission of the allosteric signal, thereby preventing C_H1–C_H2 contact. Similarly, the signal transfer could be blocked by the five GalN oligosaccharides in the IgA1 hinge. In both cases the retraction of the hinge peptide to permit the correct C_H1–C_H2 contact for binding of complement might be impaired. It has been shown in the Kol IgG protein that a single spatially fixed GlcN prevents lateral contact of the C_H2 domains (18). Similarly, the five GalN oligosaccharides in IgA1 may prevent the C_H1–C_H2 longitudinal interaction. Suggestive evidence for this is found in the claim that Fc fragments of IgA can bind complement via the classical pathway (20). Our hypothesis could be tested experimentally by stripping IgA1 of its five hinge GalN oligosaccharides and assessing its ability to bind complement.

The Oligosaccharides. We have previously reported the location and nature of the GlcN oligosaccharides in human α1 and α2 chains (7). The A2m(2) allotype chain But has five GlcN oligosaccharides in the C region. Of these, two are shared by α1 and α2 chains (those at Asn-263 and Asn-459), two are present only in the α2 chains of both allotypes (those at Asn-166 and Asn-337), and one is unique to the A2m(2) allotype (the one at Asn-211 in IgA2 But) (Fig. 1). Although the α1 chain and the A2m(1) allotype of the α2 chain have a triplet acceptor sequence for GlcN at the homologous position, the sequence is Asn-Pro-Ser instead of Asn-Ser-Ser as in the A2m(2) allotype. The presence of proline might distort the acceptor sequence, probably lessening its recognition by the transglycosidases because the Asn-Pro-Ser triplet is rarely glycosidated in glycoproteins (21). Conversely, the existence of the proposed intrachain disulfide bridge between Cys-196 and Cys-220 in α1 and A2m(1) chains, which is absent in the A2m(2) allotype, could impose restrictions on the glycosylation process.

Distribution of the Amino Acid Substitutions in α1 and α2 Chains and Its Meaning for the Domain Structure and Antigenic Determinants. Excluding the hinge deletion and one possible inversion of sequence‡ at positions 282–284, the

A2m(2) allotype protein But has 20 amino acid substitutions in the C region compared to IgA1 Bur. The substitutions are distributed throughout the C region, but almost half (nine) are concentrated in the first C domain (C_H1). Comparative analysis of α1 and A2m(2) sequences with the regions already known for the A2m(1) allotype (7) shows that substitutions Ser-212 and Arg-221, which are located on either side of the β-pleated segment 3-3 (214–218) in the C_H1 domain, have an allotypic character. Both amino acids are on the domain surface. Asn-166, which has a GlcN-containing oligosaccharide attached to it, is a subclass-specific substitution at the edge of β-pleated segment 4-4 (167–172). Presumably, the replacement of Cys-133 in the α1 subclass for Asp-133 in the A2m(2) allotype is subclass-specific as well. However, direct proof for the occurrence of aspartic acid at position 133 in the A2m(1) allotype is lacking at present. The five remaining substitutions, at positions 136, 137, 143, 197, and 198, cannot be characterized for the same reason. Incidentially, it is noteworthy that all the amino acid exchanges occurring in the C_H1 domain, with the exception of Val-143, are outside of the β-pleated sheet segments. Val-143 in the β-pleated segment 4-2 (141–151) is at the C_H1–C_L interface.

At the hinge region, the stretch of 13 residues deleted and the three proline substitutions at positions 236, 238, and 240 are shared by α2 subclass heavy chains and absent in α1 proteins. Therefore, they have a subclass-specific character.

Three amino acid substitutions are found in the C_H2 domain. Ala-277 is at the beginning of the β-sheet segment 3-1, presumably in an outer position. Three other substitutions, Glu-319, His-327, and Leu-330, are located just outside or between β-pleated sheet segments 3-2 (320–326) and 3-3 (334–340), on the surface of the domain. Again, the lack of information about the homologous residues in A2m(1) heavy chains precludes further analysis of their nature. The three remaining amino acid exchanges clustered at positions 337 (Asn), 338 (Ile), and 339 (Thr) are subclass-specific in character. In addition, Asn-337 has an attached carbohydrate moiety that is also absent in the α1 subclass. These three substitutions are located in the β-pleated sheet segment 3-3 (334–340) on the surface of the domain and near the C_H2–C_H3 interdomain region.

Four amino acid exchanges have been found in the C_H3 domain. Three of them, Glu-428, Ile-458, and Ala-467, are located on the surface, whereas Tyr-411 is in the β-segment 4-3 (408–419) at the C_H3–C_H3 interface. Our analysis indicates that Ile-458 and Ala-467 (and probably Glu-428 as well) are involved in the A2m(2) allotypic determinant antithetic to the isoallotypic [nA2m(2)] marker identified serologically by van Loghem *et al.* (4). From these facts, we conclude that most of the mutations have occurred outside the β-pleated sheets and in segments having β-turns or possibly helical structure. Furthermore, when an amino acid change has occurred it is almost always conservative in nature, thus avoiding a major perturbation of the conformation. In fact, 18 of the 20 amino acid exchanges outside the hinge conform to single base changes in the codons. The basic immunoglobulin fold (the sandwich of β-sheet layers) seems to have been strictly conserved.

In our attempts to predict secondary structure by the Chou and Fasman method (11), we had greatest difficulty in trying to fit the predicted structure for the C_H2 domain into the model of the basic immunoglobulin fold, corroborating the observations of Huber *et al.* (18). The analysis for the C_H3 domain was in general agreement with the basic pattern. However, the COOH-terminus of the α chain has an 18-residue extension compared to the γ chain, and this "tail" appears to be an extra β-pleated sheet, although sharing a strong α-potential. This suggests that the last domain of the α chain may be made up of two layers, each containing four β-pleated sheets. These extra residue "tails" are only present in polymeric immunoglobulins IgM and IgA, where they are the locus for attachment of the J chain (22, 23). We suggest that the potential for the $\beta \rightarrow \alpha$ transition found in the extra β-segment might be related to the J chain role. Further study, particularly crystallographic analysis, will be required to define the role of the carbohydrate moieties and the amino acid substitutions in the A2m(2) allotype.

We thank J. Dwulet, E. Dodge, J. Madison, and S. Dorwin for Technical assistance, Dr. G. M. Bernier for plasma from patient But, and Drs. H. G. Kunkel, M. S. Schanfield, and E. van Loghem for serological typing of the IgA proteins. This work was supported by Grants CA08497 from the National Cancer Institute and IM-2D from the American Cancer Society and by fellowships to A.T. from the Fundación del Instituto Nacional de Industria, Madrid, Spain, and the Ministerio de Educación y Ciencia Español.

1. Putnam, F. W. (1977) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic Press, New York), 2nd Ed., Vol. III, pp. 1–153.
2. Heremans, J. F. (1974) in *The Antigens*, ed. Sela, M. (Academic Press, New York), Vol. II, pp. 365–522.
3. 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.
4. van Loghem, E., de Lange, G. & Koistinen, J. (1976) *Scand J. Immunol.* **5**, 161–164.
5. Liu, Y.-S., Low, T. L. K., Infante, A. & Putnam, F. W. (1976) *Science* **193**, 1017–1020.
6. Low, T. L. K., Liu, Y.-S. V. & Putnam, F. W. (1976) *Science* **191**, 390–392.
7. Toraño, A., Tsuzukida, Y., Liu, Y.-S. V. & Putnam, F. W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2301–2305.
8. Beale, D. & Feinstein, A. (1976) *Q. Rev. Biophys.* **9**, 135–180.
9. Wolfenstein-Todel, C., Frangione, B. & Franklin, E. C. (1972) *Biochemistry* **11**, 3971–3975.
10. Kratzin, H., Altevogt, P., Ruban, E., Kortt, A., Staroscik, K. & Hilschmann, N. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1337–1342.
11. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–221.
12. Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M. & Panagiotopoulos, N. (1975) *Biochemistry* **14**, 3953–3961.
13. Köhler, H., Shimizu, A., Paul, C., Moore, V. & Putnam, F. W. (1970) *Nature* **227**, 1318–1320.
14. Fougereau, M., Bourgois, A., de Preval, C., Rocca-Serra, J. & Schiff, C. (1976) *Ann. Immunol. (Paris)* **127C**, 607–631.
15. Wuilmart, C., Urbain, J. & Givol, D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2526–2530.
16. Wolfenstein-Todel, C., Frangione, B. & Franklin, E. C. (1975) *Biochim. Biophys. Acta* **379**, 627–637.
17. Plaut, A. G., Gilbert, J. V., Artenstein, M. S. & Capra, J. D. (1975) *Science* **190**, 1103–1105.
18. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M. & Palm, W. (1976) *Nature* **264**, 415–420.
19. Pecht, I. (1976) in *The Immune System*, eds. Melchers, R. & Rajewski, K. (Springer-Verlag, New York), pp. 41–54.
20. Burritt, M. F., Calvanico, N. J., Mehta, S. & Tomasi, T. B., Jr. (1977) *J. Immunol.* **118**, 723–725.
21. Clamp, J. R. (1975) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic Press, New York), 2nd Ed., Vol. II, pp. 163–211.
22. Mendez, E., Prelli, F., Frangione, B. & Franklin, E. C. (1973) *Biochem. Biophys. Res. Commun.* **55**, 1291–1297.
23. Mestecky, J., Schrohenloher, R. E., Kulhavy, R., Wright, G. P. & Tomana, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 544–548.