Structure, Function, and Evolutionary Relationships of Fc Domains of Human Immunoglobulins A, G, M, and E

Abstract. *Human immunoglobulins A, G, M, and E have strong homology in amino acid sequence (about 30 percent) distributed nonuniformly throughout the Fc region. Immunoglobulins M and A are least alike in the first Fc domain and most alike in the second. Individual domains of heavy chains have evolved with different mutation rates but with conservation of essential structural features. No relation of primary structure to complement binding ability is apparent.*

To ascertain structural, functional, and evolutionary relationships among human immunoglobulins we have determined almost the entire covalent structure of a human IgA1 myeloma globulin designated Bur, including most of the Fd region, all of Fc, which is being reported here, the entire λ light chain, and partial sequences of several other IgA1 and IgA2 proteins (*1*). Extensive but partial sequence data on several human α chains, including Bur, have been reported by us (*2, 3*) and by others (*4, 5*). Our methods combine manual and automatic sequence analysis of some 40 tryptic, 50 chymotryptic, and 70 thermolytic peptides of the whole α chain and its CNBr fragments. The sequence of the Fc region containing 245 amino acid residues was verified by analysis of the Fc fragment prepared by limited cleavage of the whole IgA1 molecule with IgA protease (*6*). We have confirmed much of the sequence for the Fc region of other IgA1 myeloma proteins, and have also determined the chief differences from the Fc region of IgA2 myeloma proteins of $Am_2(+)$ and $Am_2(-)$ allotypes (*7*).

The amino acid sequences of the Fc region of IgA1, IgM (*8*), IgG1 (*9*), and IgE (*10*) are shown in Fig. 1. The first domains (C2γ and C3μ, respectively) of the Fc region of the complement (C1q) binding proteins IgG and IgM have no more sequence homology to each other than they do to the corresponding domains (C2α and C3ε, respectively) of IgA and IgE, which do not bind Clq. However, the last domain of the μ chain (C4μ) and of the α chain (C3α) have much more homology to each other (about 50 percent) than any other pair of domains in human immunoglobulins. Thus, although Clq is bound to the Fc region, the ability to fix complement is not determined by the overall degree of homology in the primary structure of the Fc regions.

Although μ and α are the most homologous pair of chains in the last domain (50 percent), they are the least homologous in the first domain of Fc (19 percent). In contrast, the ε and γ chains are the most homologous in the first domain (30 percent) and are equally homologous in the last domain (32 percent). Thus, the ε and γ chains are more uniformly alike throughout the whole Fc sequence than are the μ and α chains, so that overall there is little difference in the two sets. This suggests that individual domains of heavy chains evolved with different mutation rates or that crossing-over and recombination of segments of genes corresponding to domains may have occurred during evolution.

Rather than being localized in extended sequences, the homologous residues are distributed nonuniformly throughout all four chains, reflecting a framework structure conserved throughout evolution. There is least identity for all four chains in the hinge region at the beginning of Fc and in the COOH-terminal sequence at the end of Fc. In contrast, there is preservation in both Fc domains of the intrachain disulfide bonds and of two tryptophan residues; one is located 14 to 16 residues after the first half-cystine in the loop and the other is located 8 residues before the second half-cystine in the loop. Except for these, no other amino acids in homologous positions are identical in both domains of Fc.

The alignment of Fig. 1 suggests that an interchain bond (see Cys-337μ) in the hinge region of the μ, γ, and ε chains is homologous in the μ and α chains to the penultimate half-cystine (Cys-570μ) which links to the J chain or forms an interchain disulfide bond (*11*). Of the 21 half-cystine residues aligned at the right side of Fig. 1, 12 are in a Thr-Cys sequence. The alignment also juxtaposes Cys-414μ and Cys-312α. The former is thought to form the intersubunit disulfide bond in IgM (*8*); the latter probably forms an interchain or intersubunit disulfide bond in IgA (*3, 4*). The remaining four half-cystines of the Fcα (positions 242, 243, 300, 302) are unique to the α chain. The ones at 243α and 300α are joined in an extra intrachain disulfide bridge (*3, 4*); the others are probably in interchain bridges. In Fig. 1 the chains are aligned to show maximum symmetry of the domains and maximum homology of the sequences, with the implication that homologous residues have similar coordinates in the three-dimensional structure. This appears to be true for the highly conserved intrachain disulfide bond of each domain and the first tryptophan residue within the loop: for these three residues are close together and have homologous coordinates in the crystallographic structures of the V_H and C_H1 domains of Fd and the V_L and C_L domains of the light chain (*12*).

Carbohydrate groups in Fc appear to be in exposed locations and often close to disulfide bonds, but only some are in homologous positions. The five galactosamine-containing oligosaccharides unique to IgA1 are in the α1 hinge (*13*). However, the glucosamine oligosaccharide that is homologous in the μ, γ, and ε chains (Asn-402μ) is replaced at 300 to 302α by the sequence Cys-Gly-Cys. The homologous position in the second domain of Fc favors a bend or turn in the polypeptide chain, that is, it is Pro-Gly in μ and α (see 512μ), and has gaps and Gly-Ser in the alignment of γ and ε. Model-building experiments (*14*) suggest that in all eight cases this portion of the molecule will be on the surface, probably forming a crevice, and close to the interdomain segment of the Fc region.

Although the differences in biological functions of the Fc regions of IgM, IgG, IgA, and IgE must lie in their molecular structure, the structure-function relationships are not evident from the amino acid sequence. Cytotropic functions have been localized to the Fc regions and have been assigned to the last domain in IgG (*15*) and IgE (*10*). This is the most homologous domain for the four classes, but the cytotropic properties of IgG and IgE differ greatly. The complement binding site has been assigned to the first domain (Cγ2) of the Fc region of mouse and rabbit IgG (*16, 17*) but has been ascribed (*18*) to a 24-residue sequence in the second (Cμ4) domain of human IgM (residues 468 to 491μ). However, in this region the α chain (364 to 387α), which does not bind C1q, is significantly more homologous to the μ chain than is the γ chain, which does bind C1q. It is difficult to understand how the six-stranded structure of C1q, which has a molecular weight of 400,000, could bind to such a small sequence, the more so when physiologically one IgM pentamer or several IgG molecules are required for optimum fixation of C1q.

On the other hand, the inability of IgA or IgE to bind with C1q may be attributed to the conformational hindrance of Fab to the binding site as in the case of IgG4 (*16*), and not due to the degree of differences in the primary structure at this region.

All of the chains have low homology in the C_H1 region and least homology in the interdomain sequences such as the hinge region. Because the C regions of the four

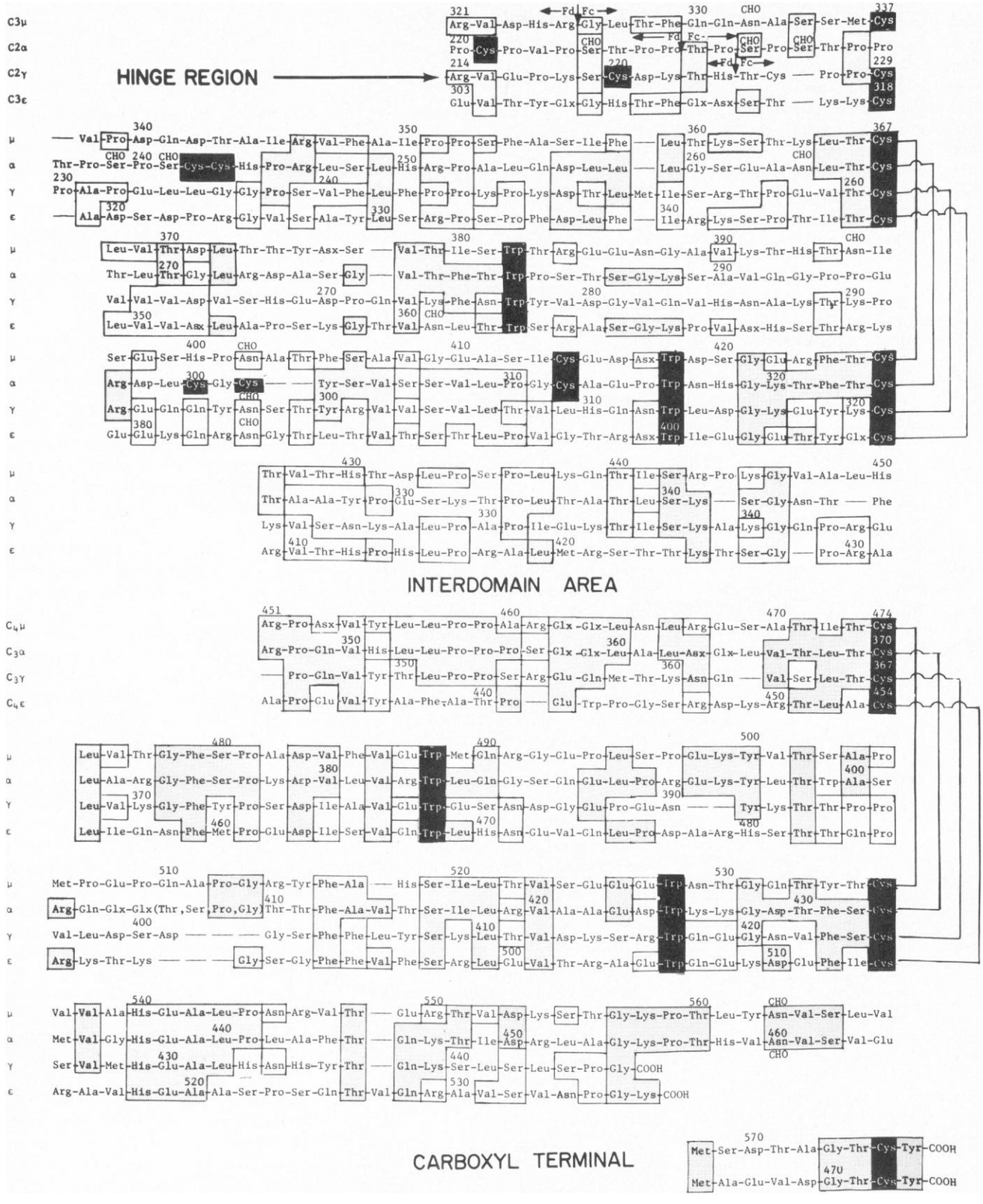

Fig. 1. Comparison of the amino acid sequence of the Fc regions of human IgM, IgA1, IgG1, and IgE, each denoted by its heavy chain (μ, α, γ, and ϵ, respectively). In the upper half the sequence is given for the hinge region and the first domain of Fc (C3μ, C2α, C2γ, and C3ϵ) and in the lower half for the second domain (C4μ, C3α, C3γ, and C4ϵ). The sequence of each chain is given continuously from left to right moving down the figure, but the chains are aligned to show maximum symmetry of the domains and maximum homology of the sequences. Residues identical in two or more chains are boxed and are shaded if identical to the α chain. Only intrachain disulfide bridges characteristic of domains are shown. The sequences are aligned to put this disulfide bridge and the two invariant tryptophan residues in register. These tryptophan residues and all half-cystine residues are shown as a negative image. A few gaps have been inserted to maximize the homology. CHO indicates carbohydrate. In the text, a position is cited both by the number and the chain; for example, the first amino acid shown for the μ chain is Arg-321 μ. Sources of sequence data are μ (*8*), γ1 (*9*), and ϵ (*10*). The sequence and tentative numbering system for the α1 chain is from the present work.

classes of heavy chains differ in length and are not uniform in homology from domain to domain, phylogenetic trees cannot be based on a single small segment such as the hinge region or the COOH-terminus (*19*). The individual domains and not the whole chains should be compared. The nonuniformity in homology of corresponding C region domains of human μ, α, γ, and ϵ chains suggests that the four classes evolved by recombination of primordial tandem genes coding for single domains.

TERESA L. K. LOW
YU-SHENG VICTOR LIU
FRANK W. PUTNAM

Department of Zoology, Indiana University, Bloomington 47401

References and Notes

1. For the nomenclature of immunoglobulins and schematic diagrams of their polypeptide chain structure and of the Fab and Fc pieces of immunoglobulins A, M, and G (IgA, IgM, and IgG), see F. W. Putnam (*2*). Abbreviations of the amino acid residues are as follows: Lys, lysine; His, histidine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Asx, aspartic acid or asparagine, identity not established; Thr, threonine; Ser, serine; Glu, glutamic acid; Gln, glutamine; Glx, glutamic acid or glutamine, identity not established; Pro, proline; Gly, glycine; Ala, alanine; Val, valine; Ile, isoleucine; Leu, leucine; Tyr, tyrosine; Phe, phenylalanine; Trp, tryptophan.
2. F. W. Putnam, in *Progress in Immunology II*, L. Brent and J. Holborow, Eds. (North-Holland, Amsterdam, 1974), vol. 1, pp. 25–37; *Science* **163**, 633 (1969); *J. Hum. Evol.* **1**, 591 (1972).
3. V. Moore and F. W. Putnam, *Biochemistry* **12**, 2361 (1973); F. W. Putnam, T. Low, V. Liu, H. Huser, E. Raff, F. C. Wong, J. R. Clamp, *The Immunoglobulin A System*, J. Mestecky and A. R. Lawton, Eds. (Plenum, New York, 1974), pp. 177–189.
4. B. Frangione and C. Wolfenstein-Todel, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3673 (1972); E. Mendez, B. Frangione, E. C. Franklin, *Biochemistry* **12**, 5186 (1973); C. Wolfenstein-Todel, F. Prelli, B. Frangione, E. C. Franklin, *ibid.*, p. 5195.
5. C.-Y. Chuang, J. D. Capra, J. M. Kehoe, *Nature (London)* **244**, 158 (1973).
6. A. G. Plaut, R. Wistar, Jr., J. D. Capra, *J. Clin. Invest.* **54**, 1295 (1974).
7. Y. Tsuzukida and F. W. Putnam, IgA2κ, Am$_2$(+) allotype, unpublished results; A. Toraño and F. W. Putnam, IgA2λ, Am$_2$(−) allotype, unpublished results.
8. F. W. Putnam, G. Florent, C. Paul, T. Shinoda, A. Shimizu, *Science* **182**, 287 (1973).
9. G. M. Edelman, B. A. Cunningham, W. E. Gall, P. O. Gottlieb, U. Rutishauser, M. J. Waxdal, *Proc. Natl. Acad. Sci. U.S.A.* **63**, 78 (1969); E. M. Press and N. M. Hogg, *Biochem. J.* **117**, 641 (1970).
10. H. Bennich and H. von Bahr-Lindström, in *Progress in Immunology II*, L. Brent and J. Holborow, Eds. (North-Holland, Amsterdam, 1974), vol. 1, pp. 49–58.
11. M. S. Halpern and M. E. Koshland, *Nature (London)* **228**, 1276 (1970); J. Mestecky, J. Zikan, W. T. Butler, *Science* **171**, 1163 (1971); J. Mestecky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 544 (1974).
12. D. R. Davies, E. A. Padlan, D. M. Segal, *Annu. Rev. Biochem.* **44**, 639 (1975); R. J. Poljak, *Adv. Immunol.* **21**, 1 (1975); A. B. Edmundson, K. R. Ely, E. E. Abola, M. Schiffer, N. Panagiotopoulos, *Biochemistry* **14**, 3953 (1975).
13. J. Baenziger and S. Kornfeld, *J. Biol. Chem.* **249**, 7270 (1974).
14. Model prediction for IgA Fc was done by Margaret H. Garner using x-ray coordinates for Fab structures and the tables of P. Y. Chou and G. Fasman [*Biochemistry* **13**, 211 (1974)].
15. K. J. Dorrington and R. H. Painter, in *Progress in Immunology II*, L. Brent and J. Holborow, Eds. (North-Holland, Amsterdam, 1974), vol. 1, pp. 75–84.
16. D. E. Isenman, K. J. Dorrington, R. H. Painter, *J. Immunol.* **114**, 1726 (1975).
17. J. M. Kehoe, A. Bourgois, J. D. Capra, M. Fougereau, *Biochemistry* **13**, 2499 (1974); M. Colomb and R. R. Porter, *Biochem. J.* **145**, 177 (1975).
18. M. M. Hurst, J. E. Volankis, R. B. Hester, R. M. Stroud, J. C. Bennett, *J. Exp. Med.* **140** 1117 (1974).
19. H. L. Spiegelberg, *Nature (London)* **254**, 723 (1975).
20. We thank J. Madison, G. East, E. Dodge, D. Radjeski, J. Novotny, and J. Dwulet for technical assistance; Dr. Elliot Ossermann, Institute for Cancer Research, Columbia University, New York, for plasma from patient Bur; and Dr. A. G. Plaut, Tufts University School of Medicine, for IgA protease. The sequence of the λ light chain of Bur IgA was determined by A. Infante. Supported by grants CA-08497 from NIH and IM-2C from the American Cancer Society. Contribution No. 1014 of the Zoology Department, Indiana University.

26 September 1975; revised 3 November 1975

Genetic Characteristics of the HeLa Cell

Abstract. *The genotype of the patient Henrietta Lacks from whose cervical carcinoma the HeLa cell was derived was deduced from the phenotypes of her husband and children, and from studies of the HeLa cell. Hemizygous expression of glucose-6-phosphate dehydrogenase in HeLa, together with the deduced heterozygosity of Mrs. Lacks, is consistent with clonal origin of her neoplasm.*

HeLa, the first established human cell line, originated from a biopsy of a cervical carcinoma of a black patient, Henrietta Lacks. The biopsy was taken at the Johns Hopkins Hospital on 8 February 1951 and the cell line was established by G. O. Gey. In 1971 Jones *et al.* (*1*) reported that review of the histologic characteristics of the original biopsy suggests that Mrs. Lacks' cancer was an adenoepidermoid tumor, an unusual type for the cervical site. This may account for the unusual gross appearance, particularly malignant clinical course, and unusual in vitro characteristics of Mrs. Lacks' neoplasm.

Over the past quarter century, the HeLa cell has made major contributions to cell biology. Gartler (*2*) reported that 20 human heteroploid cell lines had identical glucose-6-phosphate dehydrogenase (G6PD) and phosphoglucomutase (PGM) phenotypes: G6PD(A) and PGM(1-1). Although G6PD(A) is found mainly in blacks, some of the heteroploid cell lines were known to have been derived from whites. By examining the G6PD and PGM phenotypes in a variety of transformed cell lines, Gartler found that in no case were the phenotypes of G6PD and PGM different or modified from the normal tissue, which suggests that the phenotypes found were not an artifact of establishment in culture per se. He concluded that contamination with HeLa cells was the most likely explanation for G6PD(A) and PGM(1-1) phenotypes found in all heteroploid human cell lines. The problem of possible contamination of other long-term cultured tumor cell lines (*3*) with HeLa cells not only caused an international embarrassment, but also raised the concern of misattributing a specific property to another cell line, for example, a virus or a tumor-specific marker, which actually belongs to HeLa.

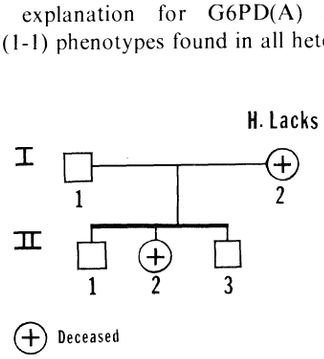

Fig. 1. The pedigree of the Lacks family.

With the continued and growing use of tissue culture in biochemical research, intra- and interspecific contamination becomes a significant risk. The determination of stable genetic markers on cultured cells is a powerful tool for monitoring such contamination. Recent experiments in which cultured cells and innumerable clones of somatic cell hybrids have been used for genetic analysis have shown that, with the proper use of polymorphic markers to characterize the cells, the possibility of undetected cross contamination of cultures is no longer the problem it once may have been (*4*). Therefore, in an effort to clarify the characteristics of the HeLa cell and establish its probable genotype for better-known polymorphisms, we studied HLA and other markers in the surviving husband and children of Henrietta Lacks.

The pedigree of the Lacks family is given in Fig. 1. The genetic markers are shown in Table 1. Henrietta Lacks' ABO and Rh types were known from her medical records—type O, Rh positive. Most of her other red cell antigens were deduced from the genotyping of her children and from data published by Kelus *et al.* (*5*), who used the method of mixed agglutination to demonstrate the M, N, S, s, and Tj^a antigens on HeLa cells.

In Table 1, the genetic markers indicated by a section symbol (§) could not be deduced for Mrs. Lacks with certainty. These included the red cell antigens systems, Jk^a, Fy^a, Fy^b, P$_1$, Le^a, Le^b, and Se, and the serum protein, haptoglobin. There is no doubt that she possessed Gm(1,5), but whether she also possessed GM(3,4),