Am J Hum Genet 38:848–854, 1986

Relationships between the Human Pepsinogen DNA and Protein Polymorphisms

R. THOMAS TAGGART,[1] I. MICHAEL SAMLOFF,[1] LESLIE J. RAFFEL,[2,3] ANGEL GRAHAM,[1] CYNTHIA CASS,[1] GLORIA M. PETERSEN,[3] JEROME I. ROTTER,[3] MARK H. SKOLNICK,[4] CHARLES E. SCHWARTZ,[4,5] AND GRAEME I. BELL[6]

SUMMARY

Pepsinogens (PGA) are the inactive precursors of pepsin, the major acid protease found in the stomach. The PGA gene family exhibits polymorphic variation in human populations that can either be demonstrated by electrophoretic analysis of the proteins or by analysis of the respective genes with cDNA probes. Here, we describe the interrelationships between the most common pepsinogen protein phenotypes and the corresponding pepsinogen haplotypes (*A*, *B*, and *C*) containing different combinations of the *PGA3*, *PGA4*, and PGA5 genes. We propose that this unusual genetic variation involving haplotypes that contain three, two, and one genes, respectively, is the result of molecular evolution by gene duplication.

INTRODUCTION

At the protein level, human pepsinogens (PGA) exhibit polymorphic variation in the pepsin activities derived from the electrophoretically distinguishable isozymogens Pg 2, Pg 3, Pg 4, and Pg 5 [1–4]. Recently, a multigene complex

Received October 25, 1985.

This work was supported by a grant from the Veterans Administration and grants AM32015 and AM3329 from the National Institutes of Health.

[1] Human Genetics Research (151M), Veterans Administration Medical Center, Sepulveda, CA 91343, and Department of Medicine, University of California, Los Angeles, CA 90073.

[2] Division of Medical Genetics, Harbor General Hospital, University of California Los Angeles, 1000 West Carson Street, Torrance, CA 90509.

[3] Division of Medical Genetics, University of Maryland School of Medicine, 665 W. Baltimore Street, Baltimore, MD 21201.

[4] Department of Human Genetics, University of Utah Medical Center, Salt Lake City, UT 84132.

[5] Greenwood Genetic Center, Greenwood SC 29646.

[6] Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608.

containing variable numbers of pepsinogen genes was localized to the centromeric region of human chromosome 11 using PGA cDNA probes [5, 6]. Among individuals exhibiting the BB phenotype characterized by the absence of the Pg 5 isozymogen, Taggart et al. [5] reported the concurrent absence of specific *Eco*RI PGA fragments detected with a human PGA cDNA probe. These studies suggested that haplotypes on chromosome 11 containing different combinations of the individual PGA genes (*PGA3, PGA4,* and *PGA5*) determined the previously described protein polymorphism.

To examine the molecular organization and structure of these haplotypes and thus the basis for the common polymorphic variation of pepsinogen, we examined informative families that were segregating for the hypothesized *A, B,* and *C* haplotypes. The pepsinogen protein phenotypes were determined by polyacrylamide gel electrophoresis (PAGE) and compared to the corresponding restriction fragment length polymorphism detected by hybridization of *Eco*RI-digested DNA with PGA gene probes. The results of these studies are consistent with the hypothesis that three PGA haplotypes (*A, B,* and *C*) contain three genes (*PGA5-PGA4-PGA3*), two genes (*PGA4-PGA3*), and one gene (*PGA4*), respectively. Close linkage of PGA with distal 11 short arm markers was excluded, a finding consistent with the centromeric location (p11-q13) of the PGA complex as indicated by somatic cell hybrid studies [5, 6]. The human pepsinogen gene family is highly polymorphic, both at the protein and DNA level, and will be an invaluable marker for linkage analysis of chromosome 11.

MATERIALS AND METHODS

Pepsinogen Protein Phenotypes

Urine samples were stored unfrozen at 4°C until analysis by polyacrylamide gel electrophoresis [7]. The pepsinogen phenotypes were classified according to the criteria described by Taggart et al. [2]. Additional rare phenotypes beyond those studied in the present report are suspected to result from additional haplotypes (containing different PGA gene combinations) and are not described here.

Restriction Fragment Length Polymorphism

DNA was isolated from whole blood and subjected to restriction enzyme digestion following the manufacturer's protocols, except that the final concentration of bovine serum albumin was 50 μg/ml. The procedures for digestion of DNA, gel electrophoresis, Southern transfer, and detection of PGA DNA fragments by hybridization with nick-translated probes were as described [5, 8]. Agarose gel concentrations varied between 0.5% and 0.8%. The restriction enzymes employed were: *Ava*II, *Bam*HI, *Bgl*II, *Eco*RI, *Hae*III, *Hin*dIII, *Kpn*I, *Mst*I, *Mst*II, *Pst*I, *Pvu*II, *Sma*I, *Taq*I, and *Xma*I. Only *Eco*RI and *Taq*I revealed polymorphic restriction fragment variation. The probes used for detection of insulin (INS), catalase (CAT), and β-hemoglobin (HBBC) have been described [9–11].

Linkage Analysis

Families informative for pepsinogen (PGA), insulin (INS), catalase (CAT), and β-hemoglobin (HBBC) were analyzed for linkage by the likelihood method of Elston and Stewart [12] using the computer program LIPED as described by Ott [13]. The frequencies

of the PGA haplotypes were determined from a previous population survey of the pepsinogen protein polymorphism: $A = .54$, $B = .37$, and $C = .06$ [2].

RESULTS

Comparison of Protein Phenotypes and PGA Restriction Fragment Length Polymorphism

Hybridization analysis of *Eco*RI-digested DNA from the five most common pepsinogen phenotypes (AA, AB, BB, AC, and BC) with the pepsinogen probe specific for exons 3–8 is shown in figure 1. There were five *Eco*RI fragments observed among these phenotypes (22, 17.8, 17.2, and 13.5, and a 3.9-kilobase [kb] doublet). The pepsinogen AA and AB phenotypes contained four of these *Eco*RI restriction fragments (17.8, 17.2, 13.5, and 3.9 kb). In comparison, the BB phenotype did not contain the 13.5-kb fragment and exhibited a reduced autoradiographic intensity of the 3.9-kb fragment. The specific association between the absence of Pg 5 isozymogen in the BB phenotype and the autoradiographic intensity variation of the 13.5- and 3.9-kb fragments when compared to the AA and AB phenotypes has been noted previously and was confirmed in the present study (table 1). In contrast, the AC and BC phenotypes contained an

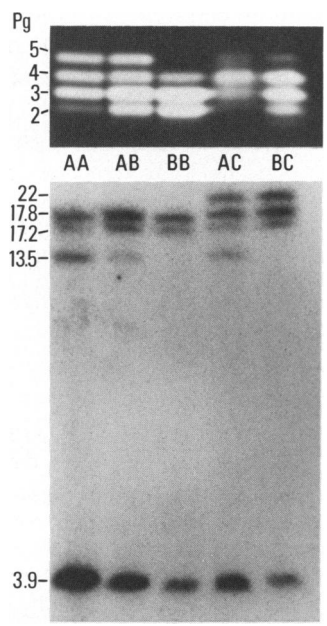

FIG. 1.—The urinary pepsinogen phenotypes AA, AB, BB, AC, and BC (*upper*) and the corresponding *Eco*RI restriction fragments detected with pepsinogen cDNA probes (*lower*). Urine samples were separated by polyacrylamide gel electrophoresis and stained for pepsin-derived activity as described in MATERIALS AND METHODS. The individual Pg 2-Pg 5 isozymogens are indicated. DNA from each phenotype was isolated from whole blood, digested with *Eco*RI, subjected to agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to ^{32}P-labeled PGA cDNA probe as described in MATERIALS AND METHODS. The location of the 22-, 17.8-, 17.2-, 13.5-, and 3.9-kb fragments are indicated.

TABLE 1

RELATIONSHIPS BETWEEN PEPSINOGEN PHENOTYPES AND PGA RESTRICTION FRAGMENTS

PROTEIN PHENOTYPE	INDIVIDUALS EXAMINED	PGA RESTRICTION FRAGMENTS (kb)				
		22	17.8	17.2	13.5	3.9
AA	11		+	+	+	+
AB	17		+	+	+	+
BB	6		+	+		+
AC	4	+	+	+	+	+
BC	3	+	+	+		+

NOTE: + refers to the presence of the particular fragment. The 3.9-kb fragment is a doublet resulting from two distinct PGA genes.

additional 22-kb *Eco*RI fragment to that observed in the corresponding homozygous AA (17.8, 17.2, 13.5, and 3.9 kb) and BB (17.8, 17.2, and 3.9 kb) fragment patterns (fig. 1). This additional 22-kb fragment was observed only among the AC and BC individuals (table 1).

Restriction Analysis of PGA Eco*RI Fragments*

To determine the composition of the distinctive *Eco*RI restriction fragment patterns observed among the *A*, *B*, and *C* haplotype containing phenotypes, we performed hybridization analysis with exon specific probes (fig. 2). This analysis provided for the alignment of fragments with regard to a PGA genomic clone [14]. The 22- and 17.8-kb fragments hybridized to both the exon 3–5 and exon 6–8 probes, which suggested their size heterogeneity resulted from variation at the 5′ region of two distinct genes. Conversely, the 17.2- and 13.5-kb fragments hybridized only with the exon 3–5 probe and the 3.9-kb doublet hybridized only with the exon 6–8 probe. These studies indicated the compositions of the *A*, *B*, and *C* haplotypes as diagrammed in figure 2 (lower). The *C* haplotype contains a single *PGA4* gene (22 kb), the *B* haplotype contains a *PGA4* and *PGA3* gene (17.8, 17.2, and 3.9 kb), whereas the *A* haplotype contains a *PGA5*, a *PGA4*, and a *PGA3* gene (17.8, 17.2, 13.5, and 3.9 kb). The distribution of observed restriction fragment patterns among the AA, AB, BB, AC, and BC individuals examined in the present study (table 1) was completely consistent with previous predictions [5].

Linkage Analysis of PGA with Chromosome 11 Markers

Linkage studies were conducted with the chromosome 11p markers INS, HBBC, and CAT (table 2) since initial studies suggested a chromosome 11 location. In order of most distal to proximal location on the short arm of chromosome 11, linkage of PGA was excluded with: HBBC at $\theta < .24$, INS at $\theta < .12$, and CAT at $\theta = .03$. The findings of these studies were consistent with the localization of the PGA family on the proximal short arm or long arm location as previously suggested by studies of somatic cell hybrids.

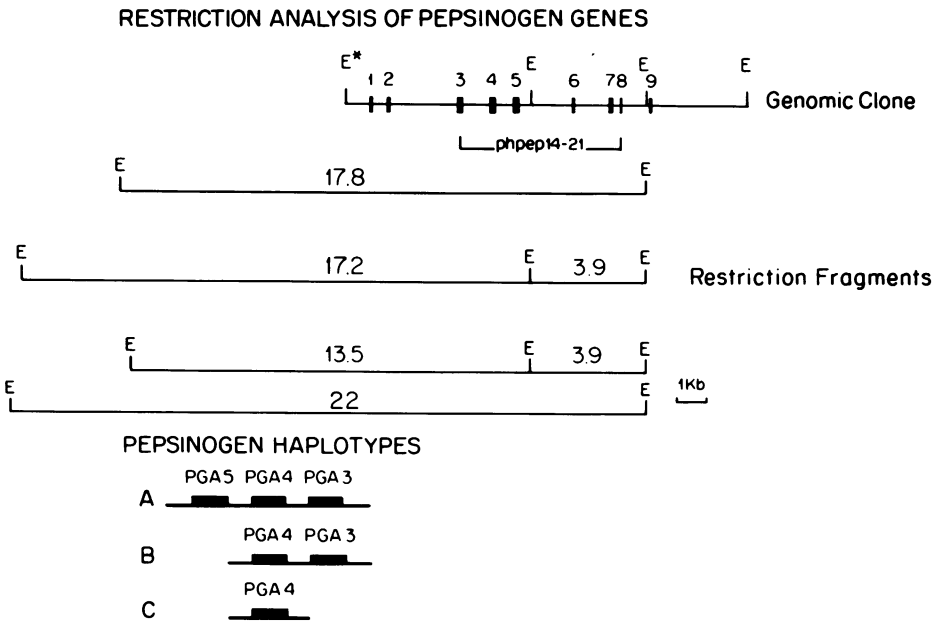

FIG. 2.—Analysis of pepsinogen restriction fragments with exon-specific probes (*upper*) and the PGA haplotypes (*lower*). The restriction fragments identified in the pepsinogen phenotypes (fig. 2) were hybridized with probes specific for exons 3–5 and 6–8 as described [5]. The fragments were aligned relative to the published pepsinogen genomic clone [14]. The *Eco*RI sites on the fragments are indicated (*E*). The 5′ *Eco*RI site in the PGA genomic clone is an artifactual result of the cloning procedure used for construction of the library. The predicted composition of the pepsinogen *A*, *B*, and *C* haplotypes with regard to the *PGA3*, *PGA4*, and *PGA5* genes are indicated.

DISCUSSION

Pepsinogen is an unusual protein polymorphism in that the observed variation is primarily the result of variation in the copy number of three genes, *PGA3*, *PGA4*, and *PGA5*, each of which produces an electrophoretically distinguishable protein. Studies of human × mouse somatic cell hybrids, informative for each human chromosome, with PGA cDNA probes demonstrated that a complex containing variable numbers of PGA genes was located in the centromeric region of human chromosome 11 [5, 6]. Preliminary comparisons between the *Eco*RI restriction fragment patterns detected by hybridization with

TABLE 2

LINKAGE ANALYSIS OF PGA WITH CHROMOSOME 11p MARKERS

TEST LOCUS	$\theta_{male} = \theta_{female}$ .05	.10	.15	.20	.25	.30	FAMILIES	INDIVIDUALS
HBBC ..	−*	−*	−*	−*	−1.75	−1.14	7	75
INS	−*	−2.57	−1.25	−.57	−.23	−.07	7	79
CAT ...	−1.20	.09	.61	.78	.77	.65	7	79

* Refers to a lod score that excluded linkage ($\ll -2$).

PGA cDNA probes from AA, AB, and BB pepsinogen phenotypes suggested that variation in the number of PGA genes was associated with the polymorphic protein variation in a specific fashion [5]. The intensity variation of one isozymogen, Pg 5, as associated with the absence of a 13.5-kb fragment and the reduced autoradiographic intensity of a 3.9-kb fragment. The studies reported here confirm and expand these initial suggestive studies of the AA, AB, and BB phenotypes to include the AC and BC phenotypes, respectively. We have presented evidence that these phenotypes are the result of polymorphic variation in the copy number of genes encoding the pepsinogen Pg 3, Pg 4, and Pg 5 isozymogens. Specifically, the *A* haplotype (*PGA3, PGA4,* and *PGA5*), the *B* haplotype (*PGA3* and *PGA4*), and the *C* haplotype (*PGA4*) contain three, two, and one genes, respectively.

Our current understanding of the pepsinogen protein polymorphism is that each PGA gene encodes a specific protein: Pg 3, Pg 4, or Pg 5, respectively [15, 16]. Intracellular posttranslational modification (probably deamidation of glutamine or asparagine) of these primary gene products gives rise to a minor, more anodally migrating, electrophoretic fraction. Thus, it is not surprising that such complex protein patterns can result from the interaction of the genetic haplotype variation with the posttranslational modification(s). Immunochemical differences between these proteins have been identified using isozymogen specific monoclonal antibodies [15]. These unique antibodies have provided for the division of the previously immunochemically indistinguishable proteins into two subclasses and have led to the purification and characterization of the Pg 3, Pg 4, and Pg 5 isozymogens [16]. Net ionic charge differences between isozymogens have been identified in the activation and pepsin regions of the molecules. The functional differences between the proteins have not yet been extensively examined; however, preliminary studies have indicated differences between the activation of Pg 5 and Pg 3 to their corresponding pepsins [15]. It is also interesting to note the large differences between the frequency of the individual *A, B,* and *C* haplotypes in the U.S., Oriental, and Northern European populations [1–4, 17]. The absence of the BB phenotype has been reported in the Oriental population in Hawaii, and two surveys of Northern European populations have reported an increased frequency of phenotypes containing the *C* haplotype. Future investigations of the structural variation of the PGA gene family with PGA cDNA gene probes will provide for the detailed characterization of the structural heterogeneity and lead to identification of additional pepsinogen haplotypes containing different combinations of PGA genes than those described here. The ability to examine the PGA variation at both the protein and gene level increases the potential for assessment of their function in the digestion of dietary protein and their role in ulcer disease.

ACKNOWLEDGMENTS

We are grateful to the individuals who participated in this study for their generous cooperation. The technical assistance of Farzin Boudi and Carol Seehof is appreciated. The *HBBC* probe was provided by Corinne Boehm, and the *CAT* probes were provided by Robert Korneluk and Gail Burns.

REFERENCES

1. SAMLOFF IM, TOWNES PL: Pepsinogens: genetic polymorphism in man. *Science* 168:144–145, 1970
2. TAGGART RT, KARN RC, MERRITT AD, YU PL, CONNEALLY PM: Urinary pepsinogen isozymes: a highly polymorphic locus in man. *Hum Genet* 52:227–238, 1979
3. KORSNES L, GEDDE-DAHL T: Genetics of pepsinogen I. *Ann Hum Genet* 43:199–212, 1980
4. FRANTS RR, PRONK JC, PALS G, ET AL.: Genetics of urinary pepsinogen: a new hypothesis. *Hum Genet* 65:385–390, 1984
5. TAGGART RT, MOHANDAS TK, SHOWS TB, BELL GI: Variable numbers of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high frequency electrophoretic polymorphism. *Proc Natl Acad Sci USA* 82: 6240–6244, 1985
6. ZELLE B, VAN KESSEL AG, DEWIT J, ET AL.: Assignment of human pepsinogen A locus to the q12-pter region of chromosome 11. *Hum Genet* 70:337–340, 1985
7. TAGGART RT, MILLER RB, KARN RC, ET AL.: Vertical thin layer slab polyacrylamide gel electrophoresis of selected human polymorphic proteins, in *Electrophoresis '78,* New York, Elsevier-North-Holland, 1978, pp 231–242
8. BELL GI, KARAM JH, RUTTER WJ: A polymorphic DNA region adjacent to the 5′ end of the human insulin gene. *Proc Natl Acad Sci USA* 78:5759–5763, 1981
9. BELL GI, HORITA S, KARAM JH: A polymorphic locus near the human insulin gene is associated with insulin dependent diabetes mellitus. *Diabetes* 33:176–183, 1984
10. KORNELUK RG, QUAN F, LEWIS WH, ET AL.: Isolation of human fibroblast catalase cDNA clones. *J Biol Chem* 259:13819–13823, 1984
11. WILSON JT, WILSON LB, DERIEL JK, ET AL.: Insertion of synthetic copies of human globin genes into bacterial plasmids. *Nucleic Acids Res* 5:563–581, 1978
12. ELSTON RC, STEWART J: A general model for the genetic analysis of pedigree data. *Hum Hered* 21:523–542, 1971
13. OTT J: Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588–597, 1974
14. SOGAWA K, FUJII-KURIYAMA Y, MIZUKAMI Y, ICHIHARA Y, TAKAHASHI K: Primary structure of human pepsinogen gene. *J Biol Chem* 258:5306–5311, 1983
15. TAGGART RT, SAMLOFF IM: Immunochemical, electrophoretic and genetic heterogeneity of pepsinogen I: characterization with monoclonal antibodies. *Gastroenterology*. In press, 1986
16. SAMLOFF IM, TAGGART RT, HENGELS KJ: Molecular variants of human aspartic proteinases, in *Aspartic Proteinases and Their Inhibitors,* New York, Walter de Gruyter, 1985, pp 79–95
17. SAMLOFF IM, LIEBMAN WM, GLOBER GA, MOORE JO, INDRA D: Population studies of pepsinogen polymorphism. *Am J Hum Genet* 25:178–180, 1973