Sequence investigation of the major gastrointestinal tumor-associated antigen gene family, GA733

(oligonucleotide/intronless gene/exon shuffling/monoclonal antibody)

ALBAN J. LINNENBACH*, JACEK WOJCIEROWSKI†, SHUANG WU, JANINA J. PYRC‡, ALONZO H. ROSS§, BERNARD DIETZSCHOLD, DAVID SPEICHER, AND HILARY KOPROWSKI

The Wistar Institute of Anatomy and Biology, 3601 Spruce Street, Philadelphia, PA 19104

Contributed by Hilary Koprowski, September 19, 1988

ABSTRACT   The monoclonal antibody-defined, tumor-associated antigen GA733 was purified from the SW948 human colorectal carcinoma cell line and its partial amino acid sequence was determined. By using a synthetic oligonucleotide probe, two recombinants were isolated from a total human genomic library. We prove the existence of a family of GA733 genes. One of the genomic isolates is demonstrated to be an intronless gene, which is transcribed in pancreatic carcinoma cell lines and in placenta. The GA733 proteins were observed to contain sequences homologous to a repeat unit occurring 10 times in thyroglobulin and once in the HLA-DR-associated invariant chain. A more evolutionarily distant relationship was found with the α chain of the interleukin 2 growth factor receptor.

Monoclonal antibodies (mAbs) CO17-1A and the related GA733 (1) have been extensively evaluated for the diagnosis and therapy of human gastrointestinal tumors. These mAbs were derived from the immunization of mice with a human colorectal adenocarcinoma cell line and a stomach adeno-carcinoma cell line, respectively. CO17-A and GA733 mAbs bind to tumors of the gastrointestinal tract and also bind in varying degrees to normal epithelial tissues. GA733, unlike CO17-1A, has an extended specificity to carcinomas of other origins, such as the cervix, bladder, and lung (1). These independently derived mAbs immunoprecipitate a 40-kDa cell-surface glycoprotein and cross-inhibit each other's binding to antigen purified by GA733 mAb-affinity chromatography (2). The two mAbs manifest sequence identities in their heavy chain CDR3 regions (3). These findings indicate that the mAbs may bind to partially overlapping epitopes on the same antigen or, alternatively, the mAbs may recognize cross-reacting epitopes present on related antigens.

Both CO17-1A and GA733 mAbs have been shown to inhibit the growth of tumor xenografts in nude mice (4, 5), pointing out the potential suitability of these mAbs for the therapy of human tumors. When human subjects with gas-trointestinal adenocarcinoma were inoculated with mAb CO17-1A (Ab1), many of the patients produced anti-idiotype antibody (Ab2) that reacted with the binding site of the Ab1 (6, 7). In theory, the administration of internal image Ab2 alone may effect an antitumor response.

By using experimental animals, the GA733 mAb (Ab1) has been used to obtain an Ab2 bearing the internal image of the GA733 tumor antigen; this Ab2 could elicit in two other species of animals an Ab3 that had Ab1-like binding specificity to purified tumor antigen (8). It has been shown in other systems that the immune response to antigen can be aug-mented by injection with antibodies to idiotype before ex-posure to antigen (9). Since purified GA733 antigen is not

available in adequate quantities for immunization, having a molecular clone for the corresponding gene would be desir-able, in that quantities of recombinant protein could then be generated. Recombinant immunogen would then be used for the development of protocols for the production of Ab3 in laboratory animals. Ultimately, the treatment of human tumors with a combined Ab2/recombinant tumor-associated antigen regimen could be realized.

In the present study, we determine the partial amino acid sequence of the GA733 antigen expressed in a colorectal carcinoma cell line and isolate a GA733-related gene that is expressed in pancreatic carcinoma cell lines.¶

MATERIALS AND METHODS

Purification and Sequencing of GA733 Antigen. The GA733 antigen was isolated by immunoaffinity chromatography from detergent extracts of SW948 tumors propagated in nude mice as described (2) except that the detergent was omitted from the basic buffer used to elute the antigen from the GA733 antibody column. The fractions judged by electro-phoretic transfer (Western) blotting to contain GA733 antigen were pooled, dialyzed against 0.05 M NH_4HCO_3, and lyoph-ilized. The protein was reduced, alkylated, and desalted using LH-20-Sephadex (1.4 × 5.5 cm) equilibrated with 88% formic acid/ethanol/water (20:50:30) (10). This material was judged to be pure by $NaDodSO_4$/PAGE and silver staining. Forty-kilodalton and 30-kDa species were observed, the latter probably representing a proteolytic breakdown product of the 40-kDa antigen.

Several amino-terminal sequence runs were performed (100–500 pmol of carboxymethylated 30-kDa GA733) on a gas-phase sequencer (Applied Biosystems) with on-line phenylthiohydantoin. Standard programs and reagents were used, except the reverse-phase column for phenylthiohydan-toin amino acid analysis was a 5-μm, 2.1 × 250 mm, LC-18-DB column (Supelco, Bellefonte, PA).

Human Genomic Library Screening. The amino-terminal 18 residues of the 30-kDa fragment were used for the design of a 54-base oligonucleotide probe, based on preferred codon usage in humans (11). The DNA probe had a 70% G + C content and included a 10-base palindromic structure. The oligomer 5'-GTCGGGGTCGTACAGGCCGTCGTTGTTC-TGCAGGGCGCCCTCGGGCTTGGCCCT-3' was synthe-
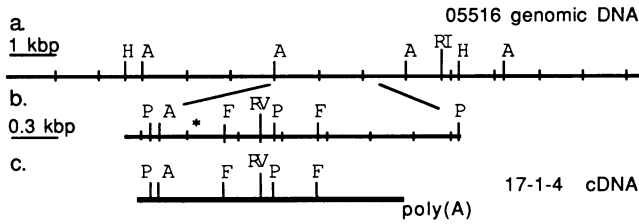

FIG. 1. Relationship between the GA733-1 chromosomal gene and full-length cDNA. (a) Restriction map for the 14.3-kbp genomic insert. Direction of transcription is from left to right. (b) The 2.2-kbp genomic region that was sequenced. The site of hybridization of the 54-base oligonucleotide probe is shown (*). (c) Restriction map of the placental cDNA clone. Enzyme abbreviations are indicated as follows: H, HindIII; A, Apa I; RI, EcoRI; P, Pst I; F, Fsp I; and RV, EcoRV.

sized by automated phosphoramidite chemistry on an Applied Biosystems model 380A DNA synthesizer. Full-length 54-mer was isolated by denaturing PAGE and C$_{18}$ chromatography as described (12).

A total human genomic library constructed by Lawn et al. (13) and obtained through the American Type Culture Collection (ATCC no. 37333) at the third amplification was plated and transferred to nitrocellulose filters in duplicate. One-half microgram of oligomer was 5' phosphorylated in a reaction mixture containing 280 μCi of [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq) and T4 polynucleotide kinase (12). The reaction mixture was adjusted to contain 0.02 M EDTA/0.5% NaDodSO$_4$ and the labeled oligomer was separated on a Sephadex G-25 column. Prehybridization and hybridization conditions for the use of the 54-base oligonucleotide were identical to those described for a 90-base probe (12).

Structural Analysis of the Genomic Isolate. A restriction map for the 14.3-kilobase-pair (kbp) genomic insert was based on electrophoretic analysis of partial digestion products of the Charon 4A recombinant using ^{32}P-end-labeled oligonucleotides complementary to the phage left and right cohesive ends (Collaborative Research, Waltham, MA).

Dideoxynucleotide sequencing was carried out (14) using T7 DNA polymerase (United States Biochemical, Cleveland) by the standard method in parallel with a method that substitutes dITP for dGTP.

The predicted amino acid sequence was evaluated for homology to known protein sequences using the program
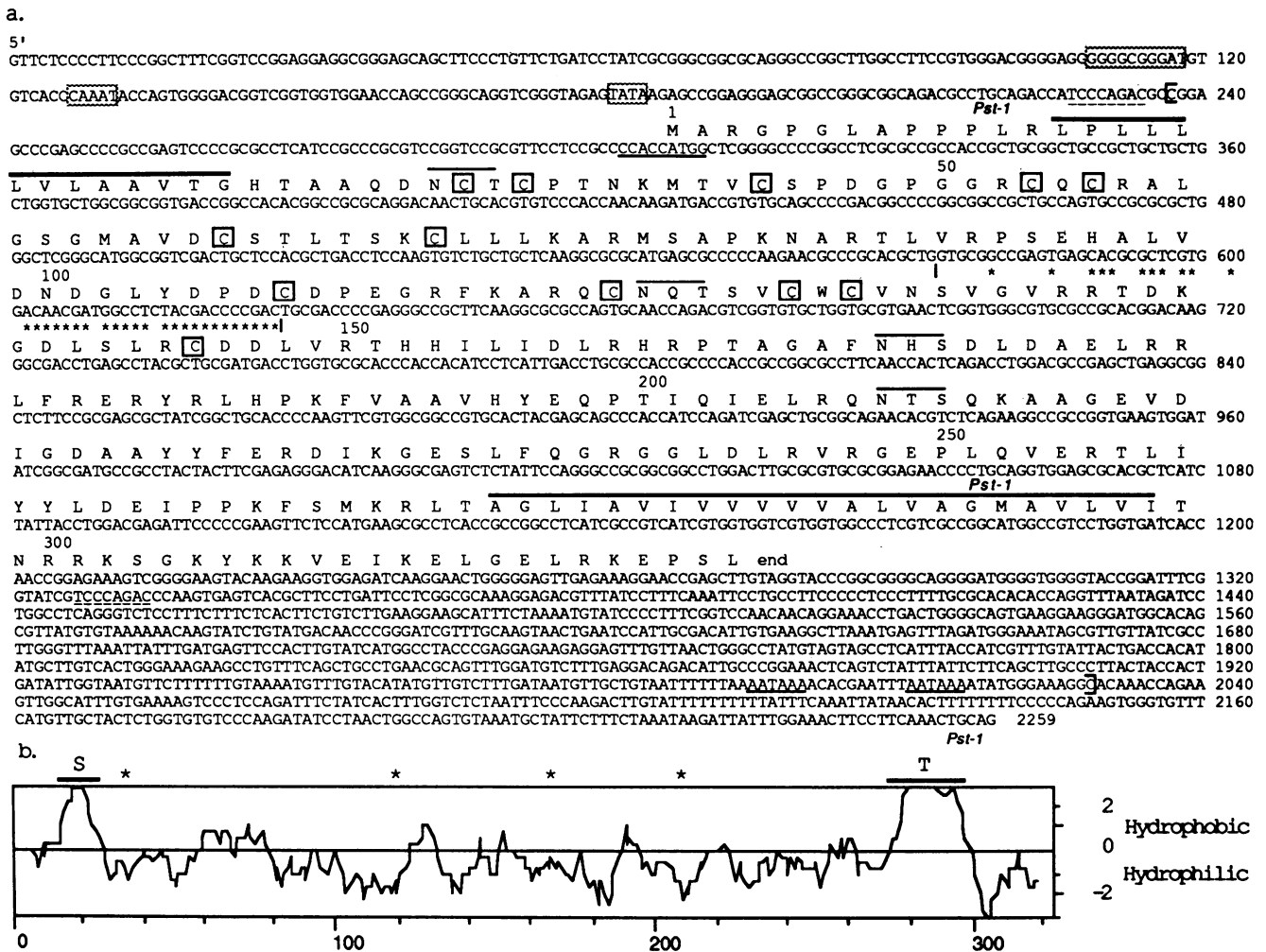

FIG. 2. (a) Complete DNA sequence of the GA733-1 chromosomal gene. The region of hybridization of the 54-base oligonucleotide probe (vertical lines) and the positions of base pairing (*) are indicated. Consensus DNA sequences include a putative Sp-1 binding site, a CAAT box, and a TATA box in the promoter region (boxes); a sequence for the initiation of translation (underline); and two possible poly(A) signals (underline). An 8-bp direct repeat flanks the coding region (dashed lines). The predicted protein is characterized by a signal sequence (overline), 12 cysteine residues (boxes), 4 potential N-linked glycosylation sites (overline), and a single 23-residue transmembrane domain (overline). Positions corresponding to the 5' and 3' ends of a full-length GA733-1 cDNA clone are marked (brackets). (b) Kyte–Doolittle plot of the 323-amino acid sequence of GA733-1. Structural features illustrated are the putative hydrophobic signal peptide (S), the potential N-linked glycosylation sites (*), and the hydrophobic transmembrane domain (T).
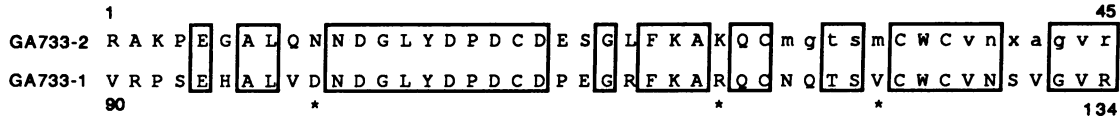
FIG. 3.   Sequence comparison of the two GA733 proteins. Lowercase letters indicate provisional assignments; position 41 is undetermined (x). Sequence identities (boxes) and chemically conserved substitutions are illustrated (*).

FASTP (15) to search release 15.0 of the NBRF protein database. Sequences with the highest score were further evaluated with the program ALIGN, using the mutation data matrix [250 point accepted mutations (PAMs)] (16). Unless otherwise indicated, the pairwise alignments were done using a gap penalty of 20. Alignment scores are expressed as standard deviation (SD) units above the mean score of 100 random runs. A score between 3 and 8 is indicative of a possible relationship; scores >8 are considered highly significant.

RNA Preparation and Hybridization Analysis. Cytoplasmic poly(A)$^+$ mRNA was purified as described (17) from the human colorectal carcinoma cell lines SW948 and SW707 and from the pancreatic carcinoma cell lines BXPC-3 and Capan-2 (ATCC). mRNAs were denatured and electrophoresed by the method of Lehrach (18), transferred to nitrocellulose filters, and hybridized to a gel-purified (12), nick-translated 0.85-kbp *Pst* I genomic fragment derived from GA733-1.

RESULTS

The sequence of the amino-terminal 45 residues of the 30-kDa fragment of GA733 was established. The oligonucleotide probe identified two different recombinants from the human total genomic library, as determined by restriction enzyme analysis and by intensity of hybridization signals. Here we analyze in detail the sequence of the initial genomic isolate, 05516, and designate this gene GA733-1; the antigen isolated from the SW948 cell line will be referred to as GA733-2.

Identification of a GA733-Related Gene. To target the initial DNA sequence characterization, 05516 DNA (Fig. 1a) was analyzed by Southern blotting and found to contain a 0.85-kbp *Pst* I restriction fragment that hybridized to the oligonucleotide probe (Fig. 1b). Analysis with the program BEST-FIT (19) indicated that the oligomer hybridized with the *Pst* I fragment at 35 (65%) of 54 positions, with a distribution of base pairing indicative of a related sequence (Fig. 2a).

An extended DNA sequence analysis of flanking restriction fragments (Fig. 1b) identified a (G + C)-rich promoter region, including three GGGCGG hexanucleotides (Fig. 2a). In the context of a decanucleotide, one GC box is identical to that of the simian virus 40 GC box IV, which has been characterized as a medium affinity site for the transcription factor Sp-1 (20), although direct experiments have not been carried out to determine if the GA733-1 promoter is in fact Sp-1 responsive. The promoter also has an atypical CAAT box (21) and a canonical TATA box.

A 323-amino acid protein is predicted with a molecular mass of 35,710 Da, which is consistent with it being a member of a family of 40-kDa glycoproteins. The probable relation-

ship between the 40-kDa and 30-kDa forms of GA733-2 that were observed during the antigen purification may be explained by the finding that the amino-terminal 45 residues of the 30-kDa form of GA733-2 correspond to GA733-1 sequences located 90 residues from the proposed amino terminus. The 40-kDa form probably contains the amino terminus and during isolation, the amino-terminal 90 residues ($\approx$10 kDa) are cleaved, giving rise to the 30-kDa breakdown product. This interpretation is consistent with the finding of 40 kDa being blocked, whereas 30 kDa was not.

A Kyte–Doolittle plot (22) of the predicted protein suggests the features of an integral membrane protein (Fig. 2b). A classic signal sequence (23) is predicted with charged residues in the precore sequence, a 13-residue hydrophobic core, and a postcore region containing amino acids with small, uncharged side chains as candidate signal peptidase cleavage sites (Fig. 2a). Assuming that the signal peptidase recognition site is T-A-A↓, where cleavage (↓) would be located after the fourth amino acid following the core sequence, a 244-amino acid extracellular domain is predicted. A clustering of 12 cysteine residues and 4 potential N-linked glycosylation sites are present in the extracellular domain. A single 23-residue transmembrane domain is followed by a 26-residue cytoplasmic domain, 9 of which are positively charged.

cDNA clones 1.8 kbp in length have been isolated; these clones are probably full-length, as their length correlated with the results of Northern blot experiments (see below). Based on restriction analysis (Fig. 1c) and preliminary DNA sequence (S.W., K. Huebner, J.J.P., H.K., and A.J.L., unpublished data), it has been determined that GA733-1 is an intronless gene. The 5' end residue of the full-length cDNA corresponds to a position in the gene sequence that is 53 bases from the TATA box (Fig. 2a), although the actual RNA start site has not been ascertained by a primer-extension experiment. The 3' end of the cDNA is 13 residues after one of two possible poly(A) signals.

Examination of the DNA sequence presented in Fig. 2a using the program REPEAT (19) detected several 8-base direct repeats. One in particular—TCCCAGAC—occurs directly before the probable RNA start site and again in the 3' untranslated region before the poly(A) addition site.

Homology of the Two GA733 Proteins. The region of 05516 flanking the site of hybridization was evaluated by comparing the open reading frame defined by the oligonucleotide with extended amino acid sequence data derived from native protein. The GA733-1 amino acid sequence was shown to be identical to that of native GA733 antigen sequence at 19 of the first 30 positions, with conserved substitutions occurring at 3 other positions (Fig. 3). In addition, a rare C-W-C amino acid sequence at positions 36–38 of the native protein is also
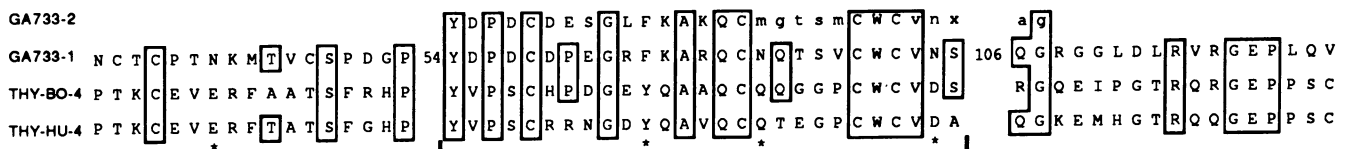

FIG. 4.   Homology of thyroglobulin and the two GA733 proteins. GA733 sequences are aligned with the fourth type I repeat unit of bovine and human thyroglobulins (THY-BO-4; THY-HU-4), respectively. Only identities (boxes) and chemically conserved substitutions (*) between GA733 and thyroglobulin sequences are shown. Vertical bars indicate 27 of the 29-residue overlap detected by FASTP. The gapped version was generated by the comparison of GA733-1 sequences with those encoded by exon 8 of the human thyroglobulin gene using the program ALIGN; the other sequences were aligned manually.
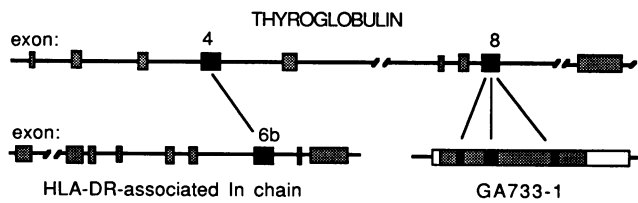
FIG. 5. Relationship between the present-day structures of the human genes for thyroglobulin, the HLA-DR-associated invariant (In) chain, and GA733-1. The exon/intron organization of part of the thyroglobulin gene and all of the HLA-DR-associated In chain gene are illustrated. Exon homologies among genes are indicated (dark boxes). The GA733-1 coding sequence region (filled boxes) and 5' and 3' untranslated sequences (open boxes) are shown.

present at corresponding positions in the GA733-1 sequence. Using ALIGN, the alignment score of the first 30 residues of GA733-2 and the GA733-1 sequence was 17 SD units. Thus, our initial genomic isolate together with protein sequence data have proven that GA733 is a member of a family of at least two closely related genes.

Homology of GA733-1 to Other Sequences. Both the GA733-1 and GA733-2 sequences were discovered to be homologous to the human and bovine thyroglobulin type I repeat unit (Fig. 4). The alignment scores of GA733-1 with a 29-residue segment of human and bovine type I repeats are 11 and 14 SD units, respectively, thus establishing the high statistical significance (16) of this homology.

Sequences encoded by each of the 16 known 5' end exons of the human thyroglobulin gene (24) were aligned to the entire 323-amino acid sequence of GA733-1. Exon 8, which encodes the fourth repeat unit in its entirety, gave the highest alignment score. Two gaps were incorporated into the 61-residue, exon 8-encoded sequence. Even taking these two gaps into account with the high gap penalty of 20, the alignment score was 9.6 SD units.

The GA733 genes are related to the HLA-DR-associated invariant chain gene (25, 26), in that exon 6b of this gene (27) is homologous to the second type I repeat of thyroglobulin, and the second and fourth type I repeats derive from an ancestral 60-amino acid unit (24). The relationships among these three genes are depicted in Fig. 5.

The database search also showed evidence of homology between GA733 sequences and the α subunit of the interleukin 2 (IL-2) growth factor receptor (28). The boundary of the homology coincides with the beginning of exon 2 of the human gene and extends over 45 residues (Fig. 6). The corresponding murine sequences were taken from the cDNA sequence (29). When the homology with the human exon 2 was analyzed by ALIGN, the highest alignment score attainable was 7.5 SD units, using a gap penalty of 40 and matrix bias of +3. This indicates a possible relationship. The GA733 sequences were more distantly related to exon 4, which has common ancestry with exon 2 (28), as 9 of 18 positions identical or chemically similar among the GA733 genes and human exon 2 remain so in exon 4 (Fig. 6).

FIG. 7. Northern blot analysis of GA733-1 mRNA in gastrointestinal tumor cell lines. (*Upper*) The nitrocellulose filter contained 2.5 μg each of poly(A)$^+$ mRNAs from the colorectal carcinoma cell line SW948 (lane 1) and the pancreatic carcinoma cell lines BXPC-3 and Capan-2 (lanes 2 and 3, respectively). Capan-2 mRNA was diluted 1:10, 1:50, and 1:100 (lanes 4–6). (*Lower*) The filter was dehybridized and then hybridized with a control enolase probe.

Expression of GA733-1 in Gastrointestinal Tumor Cell Lines. Two pancreatic carcinoma cell lines were observed by Northern blot analysis to express larger amounts of a 1.8-kb mRNA species (Fig. 7 *Upper*, lanes 2 and 3) relative to the colorectal carcinoma cell line SW948 (Fig. 7 *Upper*, lane 1). When the Capan-2 pancreatic carcinoma mRNA was diluted 1:100 (lane 7), the hybridization signal was still more intense compared to that of SW948. This apparent difference in the level of GA733-1 mRNA observed in these two cell types was normalized to enolase mRNA levels, which were observed to be constant in both types (Fig. 7 *Lower*). However, taking into account the relatedness of the GA733 genes, this experiment may not distinguish between transcription of only cross-hybridizing GA733-2 mRNA in SW948 or transcription of GA733-2 in addition to a low level GA733-1 mRNA in these cells. In similarly controlled experiments, GA733-1 mRNA was present in placenta and was not detected in the SW707 rectal carcinoma cell line or in the SK-mel-37 melanoma cell line.

DISCUSSION

A chromosomal gene encoding an integral membrane glycoprotein, termed GA733-1, was isolated and proven to be closely related to the gastrointestinal tumor-associated antigen GA733-2. Paradoxically, although the oligonucleotide probe was designed from the amino acid sequence of purified GA733-2 protein, the probe hybridized most strongly to the related GA733-1 sequence. A possible explanation for this is that the 18 codons used to design the oligomer may be interrupted by an intron in the GA733-2 gene.

FIG. 6. Homology of the GA733 proteins to the α subunit of the IL-2 receptor. GA733 sequences are aligned with 45 of the 63 residues of the second exon of the human (HU-2) IL-2 receptor gene and with corresponding murine sequences (MO-2). A single-residue gap is inserted in the mouse sequence (.). Sequences encoded by the fourth exon of the human (HU-4) IL-2 receptor gene and the corresponding murine sequences (MO-4) are shown below.
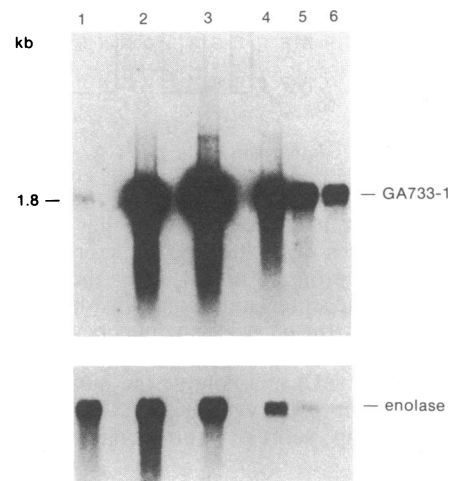
That GA733-1 is a functional intronless gene is rather uncommon. In contrast, inactive retropseudogenes are a relatively frequent occurrence. Although no evidence for the remnants of a poly(A) tail were observed in the genomic sequence, a flanking direct repeat was observed. Thus, retrotransposition (30) is a possible mechanism for the gene duplication in this family, with truncation occurring in the 3' untranslated sequence of the precursor cDNA. This can be evaluated more definitively when the organization of the genes for the other members of this family are determined.

The phenomenon of exon shuffling (31) has probably played a role in the evolution of the GA733 genes, as portions of these genes were found to be homologous to exon 8 of thyroglobulin. The exon shuffling event must have preceded the gene duplication event(s) within the GA733 gene family. After the acquisition of sequences encoding the 60-amino acid domain, insertional events may have ensued, resulting in the segmentation of the domain into three parts. Although statistically significant, the functional significance of this relationship is unknown, as the function of the type 1 repeats in the thyroglobulin molecule is not understood. A comparative analysis of the present-day sequences of the thyroglobulin, HLA-DR-associated invariant chain, and the GA733 genes suggests that two exon shuffling events have occurred.

The GA733 sequences with homology to the IL-2 receptor overlap with those that are related to thyroglobulin. Although exon 2 of the IL-2 receptor and exon 8 of thyroglobulin both derive from ancestral 60-amino acid repeating units (24, 32), no statistically significant relationship could be demonstrated between them. A possible relationship between these two ancestral units is recognizable only in the context of their relationship to GA733 sequences and may be limited to a structural motif present within the two ancestral units.

The homology of GA733-1 to the IL-2 receptor is intriguing, in that it coincides with exon 2 of the IL-2 receptor, which encodes sequences involved in growth factor binding (33). Although conclusions cannot be drawn as to the biological significance of this sequence relationship, it raises the possibility that the GA733 tumor-associated antigen gene family may function as growth factor receptors, perhaps binding growth factors related in structure to IL-2. This seems plausible in light of the finding that the human genome contains sequences homologous to the gene for IL-2 (34).

The isolation of other members of the GA733 gene family will allow a determination of their tissue specificity and their binding reactivities to a group of related mAbs. The generation of additional mAbs of different specificities will also be possible. The availability of recombinant protein or of synthetic peptides for B- and T-cell epitopes will provide immunogen for possible therapeutic applications. The relatedness of GA733-1 to the gastrointestinal-associated antigen GA733-2 and the observed expression of the GA733-1 antigen in pancreatic carcinoma cell lines identifies this gene product as a potential target for immunotherapeutic approaches to this cancer.

1. Herlyn, M., Steplewski, Z., Herlyn, D. & Koprowski, H. (1986) *Hybridoma* **5**, S3–S8.
2. Ross, A. H., Herlyn, D., Iliopoulos, D. & Koprowski, H. (1986) *Biochem. Biophys. Res. Commun.* **135**, 297–303.
3. Caton, A. J. (1986) *Hybridoma* **5**, S11–S15.
4. Herlyn, D., Steplewski, Z., Herlyn, M. & Koprowski, H. (1980) *Cancer Res.* **40**, 717–721.
5. Herlyn, D., Herlyn, M., Ross, A. H., Ernst, C., Atkinson, B. & Koprowski, H. (1984) *J. Immunol. Methods* **73**, 157–167.
6. Koprowski, H., Herlyn, D., Lubeck, M., DeFreitas, E. & Sears, H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 216–219.
7. Herlyn, D., Lubeck, M., Sears, H. & Koprowski, H. (1985) *J. Immunol. Methods* **85**, 27–38.
8. Herlyn, D., Ross, A. H. & Koprowski, H. (1986) *Science* **232**, 100–102.
9. Kennedy, R. C., Adler-Storthz, K., Henkel, R. D., Sanchez, Y., Melnick, J. L. & Dreesman, G. R. (1983) *Science* **221**, 853–855.
10. Marano, N., Deitzschold, B., Earley, J. J., Schatteman, G., Thompson, S., Grob, P., Ross, A. H., Bothwell, M., Atkinson, B. F. & Koprowski, H. (1987) *J. Neurochem.* **48**, 225–232.
11. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9**, r43–r74.
12. Linnenbach, A. J., Speicher, D. S., Marchesi, V. T. & Forget, B. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2397–2401.
13. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
16. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
17. Linnenbach, A. J., Huebner, K., Reddy, E. P., Herlyn, M., Parmiter, A. H., Nowell, P. C. & Koprowski, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 74–78.
18. Lehrach, H., Diamond, D., Woznez, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.
19. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
20. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
21. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connel, C., Spritz, R. A., DeReil, J. K., Forget, B. F., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
22. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
23. Perlman, D. & Halvorson, H. O. (1983) *J. Mol. Biol.* **167**, 391–409.
24. Parma, J., Christophe, D., Phol, V. & Vassart, G. (1987) *J. Mol. Biol.* **196**, 769–779.
25. Kudo, J., Choa, L.-Y., Narni, F. & Saunders, G. F. (1985) *Nucleic Acids Res.* **13**, 8827–8841.
26. Strubin, M., Berte, C. & Mach, B. (1986) *EMBO J.* **5**, 3483–3488.
27. Koch, N., Lauer, W., Habicht, J. & Dobberstein, B. (1987) *EMBO J.* **6**, 1677–1683.
28. Leonard, W. J., Depper, J. M., Kanehisa, M., Kronke, M., Peffer, N. J. Svetlik, P. B., Sullivan, M. & Greene, W. C. (1985) *Science* **230**, 633–639.
29. Miller, J., Malek, T. R., Leonard, W. J., Greene, W. C., Shevach, E. M. & Germain, R. N. (1985) *J. Immunol.* **134**, 4212–4217.
30. Weiner, A., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
31. Gilbert, W. (1985) *Science* **228**, 823–824.
32. Kristensen, T., D'Eustachio, P., Ogata, R. T., Chung, L. P., Reid, K. B. M. & Tack, B. F. (1987) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **46**, 2463–2469.
33. Neeper, M. P., Kuo, L.-M., Kiefer, M. C. & Robb, R. J. (1987) *J. Immunol.* **138**, 3532–3538.
34. Mita, S., Meada, S. & Shimada, K. (1986) *Biochem. Biophys. Res. Commun.* **138**, 966–973.