Genomic Structure of the Human IgX1 Gene Suggests That It May Be Expressed as an IgX14.1-like Protein or as a Canonical B Cell IgX Light Chain: Implications for IgX Gene Evolution

By Robert J. Evans and Gregory F. Hollis

From the Department of Biological Sciences, Monsanto Company, Saint Louis, Missouri 63198

Summary

In pre-B cells, immunoglobulin μ (Igμ) is associated with pre-B cell-specific proteins to form a multimeric complex that is found on the cell surface. One of these proteins is encoded by the three exon IgX-like gene 14.1, whose expression is restricted to pre-B cells and occurs from an unrearranged gene. A comparison of the 14.1 gene structure to the seven-gene human IgX locus revealed that the most 5' gene, IgX1, is organized in a three-exon structure very similar to the 14.1 gene. Transcription and splicing of these three-exon sequences would lead to an mRNA with an open reading frame which could encode a light (L) chain-like protein with a molecular weight of 23,045. Our analysis suggests that two transcripts may be produced from the IgX1 gene that share the same IgX1 constant region-containing third exon. One transcript would include all three 14.1-related exons and be expressed from the germline gene, and the second transcript would be produced after variable joining (VJ) recombination has occurred to IgXJ1 and would encode a classic IgX L chain protein. The conservation of the genomic organization of the human 14.1 and IgX1 genes and the mouse homolog, A5, relative to the classic IgX L chain genes provides insight into the evolution of Ig genes.

The development of B cells from stem cells to mature Ig-secreting B cells can be classified by the rearrangement and expression of Ig H and L chain genes. In this progression, Igμ H chain undergoes VDJ rearrangement, is expressed first, and in part defines the pre-B cell stage of development. Subsequently, the Ig L chain genes, κ and λ, undergo rearrangement and expression. Expression of both Igμ H and L chain proteins leads to the formation of the Ig tetramer μ2λ2 (L: κ or λ L chain) and marks the transition of pre-B cells to the B cell stage of development (1-6).

During our studies of the human IgX locus, we identified and isolated two clones, 14.1 and 16.1, based on their shared homology with the IgX C region (7, 8). More recently, we have shown that at least one of these clones, 14.1, is expressed as a 1-kb transcript exclusively in pre-B cells (9). This transcript contains both IgX J and C region homologies as well as 5' sequence that is not derived from an Ig V region.

Sequence analysis of a cDNA clone, Hom-1, derived from the 14.1 gene revealed a long open reading frame capable of encoding a protein of an unprocessed mol wt of 22,944. Polyclonal antisera, generated to a peptide predicted by the nucleotide sequence, identified a 22-kD protein as the product of the 14.1 gene. Immunoprecipitation experiments using anti-Igμ antisera demonstrated that at least two proteins, of 16 and 22 kD, are associated with Igμ in pre-B cells.

Taken together, the immunoprecipitation and Western analysis suggested that the protein product encoded by the 14.1 gene is complexed to Igμ in pre-B cells (9). Recently, these results have been confirmed and extended by Kerr et al. (10) who have shown that cell surface Igμ in pre-B cells is covalently linked to two protein chains, 16 and 22 kD in size, that crossreact with antisera directed against human Igμ protein.

Similar results have been described for mouse pre-B cells where a surrogate L chain, termed ω, has been shown to be complexed with Igμ to form a μ2ω2 tetramer (11). In separate experiments, a cDNA encoding a pre-B cell-specific transcript, A5, was isolated that contains homology to Ig J and C regions (12-14). The mouse A5 gene contains the putative coding sequence for the mouse ω protein, and based on sequence similarity, is likely to be the mouse homologue of the human 14.1 gene (9, 15).

Analysis of the 14.1 gene indicates that it is expressed without gene rearrangement and is encoded in three exons, in which exons 2 and 3 contain Ig J and C region homology, respectively. Characterization of the genomic structure of the 14.1 gene led to the discovery that the human IgX1 gene, the most 5' gene of the seven gene λ locus, is organized in a similar three-exon structure. This germline IgX1 gene contains appropriate RNA processing signals that could lead to
the production of a transcript with one long open reading frame with an in frame AUG start site. This analysis suggests that the Ig\(\lambda\)1 gene may be expressed in two ways. One transcript would be expressed from the germline gene and include all three 14.1-related exons and the second transcript would be produced from a rearranged Ig\(\lambda\)1 gene after V-J recombination and would encode a classic Ig\(\lambda\) L chain protein.

**Materials and Methods**

**DNA and RNA Analysis.** Plasmid DNAs (1 \(\mu\)g/lane) were digested with the indicated restriction enzyme according to manufacturers recommendations, fractionated on a 0.8% agarose gel, and transferred to nitrocellulose paper (16). RNA extraction and blots were performed as described (9, 17). DNA and RNA blots were hybridized to the following \(^{32}\)P-labeled human DNA probes: (a) 847-bp fragment including all of the 14.1 cDNA clone Hom-1 (9); (b) 300-bp SstI-BamHI fragment from clone 14.1, which contains homology to human IgX C (probe A); (c) a 319-bp fragment containing 14.1 exon 1 sequences generated by PCR amplification using the oligonucleotides 5'-GGCCACATGGACTGGGGTGC-3' and 5'-CCACCGGCTCCTCAGGCTGG-3' (probe B) (as recommended by Perkin-Elmer Corp., Norwalk, CT); (d) 200-bp Smal fragment from clone 16.1, which shares \(>95\%\) sequence similarity with exon 2 of 14.1; (e) a 60-bp fragment from nucleotide position 78-137 of GA1 exon 2 (probe D). DNA and RNA blots were hybridized overnight at 42°C in a 10% dextran sulfate, 4x SSC, 40% formamide, 0.8% Denhardt's Tris buffered solution. After hybridization, filters were washed with 2x SSC, 0.1% SDS three times at room temperature, and with 0.1x SSC, 0.1% SDS two times for 20 min at 55°C before autoradiography.

**Genomic Cloning.** A human placental DNA bacteriophage library (10\(^6\) bacteriophage) or an MboI partial digest of DNA from a human pre-B cell line HPB Null inserted into Stratagene's \(\lambda\) DashII vector (5 \(\times\) 10\(^5\) bacteriophage) were screened with probes from 14.1 exon 1 or exon 3 as described previously (8, 17). Positive bacteriophage were plaque purified, and mini-lyse DNAs were prepared. Insert fragments from these clones were subcloned into Bluescript (Stratagene) for restriction mapping and DNA sequence analysis. Restriction digests were performed with indicated enzymes under conditions recommended by the manufacturer (Bethesda Research Laboratories, Gaithersburg, MD).

**Nucleotide Sequence Analysis.** The DNA sequence analysis was determined by the dideoxy chain termination method using the USB Sequenase DNA sequencing kit. Synthetic oligonucleotides (Research Genetics) were used as sequencing primers. Nucleic acid alignments and translations were done using the University of Wisconsin (Madison, WI) Sequence analysis software package (18).

**Results**

**Genomic Structure of the IgX-like Gene 14.1.** Our original identification and isolation of the pre-B cell-specific 14.1 gene included a partial genomic characterization (7). With the isolation of the 14.1 cDNA, it became apparent that upstream sequences, not found on our original clone, were part of the transcription unit. A human placental DNA library was screened with a Hom-1 cDNA probe and one positive clone

**Figure 1.** Restriction map of human Ig\(\lambda\) locus, G\(\lambda\)1, and 14.1 genes. A partial restriction map of the seven-gene Ig\(\lambda\) C region locus on human chromosome 22 is depicted. J and C region sequences are depicted by solid boxes and numbered. An expanded restriction map of the 14 kb EcoRI fragment containing Ig\(\lambda\) C region 1 is shown. Three coding exons of GA1 gene are depicted by solid boxes and are numbered. E=EcoRI, Bg=BglIII, H=HindIII, S=Stul (note: not all Stul sites are shown). A partial restriction map of the 14.1 gene is shown. The coding exons of the 14.1 gene are depicted by solid boxes and are numbered. E=EcoRI, B=BamHI, H=HindIII, and X=XbaI.
was isolated. A partial restriction map of this clone is shown in Fig. 1. Experiments using the Hom-1 cDNA as a probe identified three areas of hybridization that define the three coding exons of the 14.1 gene. Exon 1 is positioned ~5 kb 5' of exon 2, which is in turn located ~1.2 kb 5' of exon 3. Sequence analysis of the three areas of homology to the Hom-1 cDNA proved that the clone contained the coding sequence for 14.1. These results are consistent with our prior partial characterization of the 14.1 gene as well as the recently described genomic characterization of 14.1 by Schiødt et al. (15).

**Genomic Southern Analysis with the 14.1 Exon 1 Probe and Isolation of a 14.1-related Gene.** Hybridization experiments using an exon 1 probe from 14.1 on a BamHI and SstI genomic Southern blot were performed using the 14.1 exon 1 probe. This analysis identified two dominant bands at 7.6 and 3.0 kb in the BamHI digest and a 1.3- and a 1.1-kb band in the SstI digest (data not shown). Restriction mapping indicates that exon 1 of 14.1 is present on a 7.6-kb BamHI fragment and a 1.3-kb SstI fragment (Fig. 1). To clone the exon 1 sequence contained on the 3.0-kb BamHI fragment, an MboI partial genomic DNA library was made from the human pre-B cell line HPB Null. Genomic Southern blot analysis showed that this gene was identical in structure in both pre-B cell and nonhematopoietic DNA (data not shown). The primary screen of the library was done with the 14.1 exon 1 probe and positive clones were counter screened with an exon 3 probe. Three clones hybridized with both probes and were plaque-purified. One of these clones contained a 7.6-kb BamHI...
fragment diagnostic of 14.1 and was not characterized further. The remaining two clones contained a 3.0-kb BamHI fragment that hybridized to the 14.1 exon 1 probe and restriction map analysis indicated that the two clones shared common sequence.

One of these clones, GA1, was analyzed in more detail and shown to contain a 14-kb EcoRI fragment. A partial restriction map of this clone was determined (Fig. 1). Comparison of the GA1 restriction map to clones containing the seven human Igλ C region genes indicated that the GA1 clone was similar to the most 5' gene Igλ1 (8, 19, 20). Sequence of the region homologous to IgλC of clone GA1 was identical to the previously determined sequence for Igλ1 C region confirming the nature of the clone (Fig. 2). Sequence of the exon 1 homologous region identified a potential coding exon similar in structure to exon 1 of 14.1. This exon had one long open reading frame beginning with a start ATG capable of encoding 68 complete amino acids. This exon has a splice donor site splitting amino acid 69 that is identical to the splice donor site found at the same position of 14.1 and is a good match for a consensus splice donor site (21). This exon shows 81% and 71% homology with 14.1 exon 1 at the nucleotide and amino acid sequence level, respectively. This high degree of similarity to 14.1 exon 1 explains its identification on Southern blot and its subsequent cloning using a 14.1 exon 1 probe. Analysis of the sequence upstream of the cDNA start failed to identify a characteristic TATA or CAT box. Search of the 5' sequence failed to identify the octamer motif ATGCAAAT that has been implicated in tissue-specific expression of Ig genes (22–27).

To determine if the GA1 clone contained a sequence analogous to exon 2 of 14.1, we sequenced the area surrounding the Igλ1 J region. This analysis revealed an open reading frame capable of encoding 40 amino acids, one amino acid longer than exon 2 of 14.1, with a consensus splice acceptor site at its 5' end and a consensus splice donor site at its 3' end (Fig. 3). This sequence includes the heptamer and nonamer recombination signal sequences found upstream of all J regions and in this exon would be used as coding sequence. This putative exon shares 72% and 55% similarity with the 14.1 exon 2 sequence at the nucleotide and protein level, respectively. The structure of this putative exon is such that it can join this sequence to the exon 1 and 3 sequences in an open translational reading frame.

This newly identified gene, GA1 (for germline Igλ1), is capable of encoding a protein of 214 amino acids with a mol

| Table 1. Nucleotide and Amino Acid Comparison of 14.1, GA1, and λ5 |
|-------------------|--------|--------|--------|
| Exon   | Amino acid | 14.1   | GA1    | λ5     |
| 1      | 14.1      | 81     | 61     |
|        | GA1       | 71     | 58     |
|        | λ5        | 50     |        |
| 2      | 14.1      | 72     | 66     |
|        | GA1       | 55     | 75     |
|        | λ5        | 52     |        |
| 3      | 14.1      | 89     | 72     |
|        | GA1       | 84     | 77     |
|        | λ5        | 61     | 62     |

Comparison of nucleotide and amino acid sequence similarity of 14.1, GA1, and Mouse λ5; Comparison of nucleotide sequence and amino acid sequence is shown in percent identity for 14.1, GA1, and mouse λ5.

Figure 3. Northern analysis of pre-B cell RNA with GA1 specific probe. (A) 1 μg of plasmid DNA containing 14.1, 16.1, and GA1 gene sequences were digested with BamHI and EcoRI (14.1 and 16.1) or EcoRI and HindIII (GA1) and Southern blotted. (B) Total RNA from five pre-B cell lines (Reh, RS4;11, Nalm1, Nalm6, and HPB Null), two B cell lines (BL33 and DHL6), and HeLa cells were analyzed by Northern analysis. DNA and RNA blots were hybridized with the GA1 exon 2–specific probe D.
region sequences, the components found in an Ig L chain
plex, the VpreB and 14.1 gene products together encode V+C
sequences an IgV gene-related sequence (12, 29, 30). In this com-
plex appears to include the product of the 14.1 gene and
part of a protein complex on the cell surface (10, 11). This
complex is being expanded to examine a larger number of cell
lines and to look for GA1 protein.

The Genomic Organization of the Murine Immunoglobulin
Related Gene, λ5, Is Similar To 14.1 and GA1. Like 14.1 and
GA1, mouse λ5 is encoded by three exons in which exon 3
contains Ig C region homology, exon 2 contains Ig J region
homology as well as 79 nucleotides of 5′ sequence, and exon
1 contains sequence unrelated to Ig V regions (28). Sequence
comparisons at both the nucleotide and amino acid level re-
veal striking homology between 14.1, GA1, and mouse λ5
through all three exons (Table 1). For instance, exons 1, 2,
and 3 of mouse λ5 are 58, 75, and 77% similar, respectively,
to GA1 in their nucleotide sequence. We and others have
identified mouse λ5 as the murine homolog of 14.1 (9, 15).

Based on genomic structure and a high degree of sequence
similarity it appears that mouse λ5, human 14.1, and GA1
are all derived from a common ancestor.

Discussion

Specific DNA rearrangement that juxtaposes V and J regions
(and in the case of Ig H chain, diversity regions) is a prereq-
uisite for expression of functional Ig H and L chain proteins.
These events, which are readily detected by Southern anal-
ysis on clonal B cell populations, have been used with great
success to mark B cell development (1–6). These studies have
shown that Igµ rearrangement and expression occurs at the
pre-B cell stage and precedes Ig L chain rearrangement.
Recently, the pre-B cell Igµ has been found to be expressed as
part of a protein complex on the cell surface (10, 11). This
complex appears to include the product of the 14.1 gene and
another pre-B cell–specific gene product, VpreB, which en-
codes an Ig V gene–related sequence (12, 29, 30). In this com-
plex, the VpreB and 14.1 gene products together encode V-J-C
region sequences, the components found in an Ig L chain
protein. Their presence on the cell surface suggests that this
Igµ–surrogate L chain complex may mark pre-B cells that
have undergone functional IgH gene rearrangement. These
cells may then proceed on to rearrange and express Ig L chains.
The signals required for Ig L chain gene rearrangement are
unknown, but it will be interesting to determine if the for-
maton of an Igµ–surrogate L chain complex on the pre-B
cell plasma membrane is a necessary step for B cell differen-
tiation.

The human 14.1 gene is encoded by 3 exons of which exons
2 and 3 share a high degree of similarity with Ig J and C
regions, respectively. Unlike classical Ig L chains, the 14.1
gene is expressed from a germline gene. The high degree of
conservation of the Ig C region domain in the third exon
of the 14.1 gene presumably affords the necessary contacts
for IgH-14.1 interactions. The NH2-terminal half of the 14.1
protein, encoded by exon 1 and the 5′ half of exon 2, does
not share homology with Ig V region genes (7, 9, 15).

Surprisingly, like 14.1, the most 5′ gene of the human λ
locus, Igλ1, was shown to be organized in a similar three-
exon structure in which exon 3 corresponds to the Igλ C
region and exon 2 includes Igλ J region sequence. While
this gene could be an evolutionary remnant, analysis of these
three exons indicates that this germline gene, termed GA1,
contains several hallmarks of a functional gene. First, all three
exons are open reading frames. Second, an initiating ATG
is found in frame in the putative exon 1 sequence. Third,
consensus splice donor and acceptor sites are found at the
intron-exon junctions. Finally, a poly A addition site is found
110 bp downstream of exon 3. Transcription and processing
of these three exons would produce a transcript containing
an open reading frame capable of encoding a protein with a
mol wt of 23,045. This protein would use germline se-
quences found upstream of exon 2 as coding sequence and the
transcript would include the recombining nonamer and
heptamer signal sequences that are normally deleted by a V-J
rearrangement event. These results suggest that the Igλ1 gene
can be expressed in two different manners. First, as we had
previously described, as a classic Igλ L chain after Ig V-J gene
rearrangement (8). Second, the Igλ1 gene may be expressed
in an unrearranged form (GA1) utilizing the three exons related
to the 14.1 gene. Because the exon 1 and 5′ portion of exon
2 are in the region deleted by V-J rearrangement, this gene
must be expressed before Igλ V-J joining in B cell develop-
ment and/or be expressed in other cell types.

Based on the high degree of similarity between the pre-B
cell–specific gene 14.1 and the GA1 clone, we examined the
expression of GA1 in human pre-B cells. Our Northern anal-
ysis indicated that unlike 14.1, the GA1 clone was not ex-
pressed in pre-B cells. Of the cells examined, only an
Igκ–producing B cell, DHL6, showed any GM transcript.

This transcript was ~800 bp in size and therefore slightly
smaller than the 14.1 transcripts seen in pre-B cells. The bi-
ological role of this germline transcript is unknown at this
time. Transcription of germline Ig genes is well documented
in both IgH and κ (for review see reference 31). Studies of these
germline RNAs have shown that they can undergo RNA
processing (32). While most of these transcripts are believed to be “sterile” (incapable of producing a functional protein), at least in one case a truncated Ig H chain protein is produced. Several studies have demonstrated that germline transcription of these genes precedes gene rearrangement and may be necessary to open the chromatin structure to the recombination machinery. The expression of GX1 indicates that transcription can occur from the germline human IgA locus. This germline Igα transcript could be a forerunner of V-J gene rearrangement. Alternatively, because we have shown that there are three exons that can be spliced to produce an open reading frame, this transcript may lead to the expression of the GX1 sequences at the level of protein. A more extensive analysis of GX1 expression is being pursued to determine if transcription of the GX1 gene leads to production of Igα protein.

Gene duplication from a single λ J and C region segments has given rise to the tandemly repeated J-C region structure of the present day seven-gene human IgA locus. This gene duplication appears to have occurred after mouse and man diverged (7, 8, 19, 20). The shared three-exon structure of the 14.1 and GA1 genes indicates that they arose by gene duplication from a common ancestor. Sequence comparisons of the 14.1, GA1, and the mouse λ genes indicate that 14.1 and GA1 are more similar to each other (exon 3: >89%) than to mouse λ genes (exon 3: 74% 14.1 and 75% GA1 to mouse λ1). This suggests that the gene duplication that gave rise to 14.1 and GA1 followed divergence of mouse and man. Similarly, GA1 is more similar to the remaining genes of the seven-gene λ locus than to 14.1 (exon 3: >95% vs. 89%), suggesting that gene duplication of the seven-gene cluster occurred after GA1 and 14.1 duplicated. This comparison suggests that GA1 may have been the original human Igλ gene whose exon 2 and 3 sequences, which contain J and C regions, were duplicated to give rise to the present day human λ locus.

The conserved three-exon motif of 14.1 and GA1 is unlikely to have arisen recently, because the murine homologue of 14.1, mouse λ5, is organized in a similar fashion (28). This suggests that the three-exon structure found in 14.1, GA1, and λ5 existed in an ancestral gene before mouse-man divergence. Like man, the mouse has multiple copies of λ genes that appear to have arisen by gene duplication. A comparison of 14.1 to the four mouse λ genes and λ5 reveals that 14.1 is as similar to mouse λ 1–4 (exon 3: 72–74%) as it is to mouse λ5 (exon 3: 72%) (34). This suggests that the duplication that produced mouse λ genes 1–5 occurred from a single gene after mouse and man diverged. Based on these comparisons, we feel that the ancestral gene to 14.1/GA1/λ5 and the present day Ig λ genes of mouse and man is likely to have been organized in the three-exon structure seen in 14.1, GA1, and mouse λ5. This primordial Igλ gene may have been expressed as an Igλ 14.1-like protein from the germline gene or, after VJ recombination, as a canonical Igλ L chain. The results described above lead to the question, is the Vpre-B-14.1 surrogate L chain complex found associated with Igμ in pre-B cells adapted from the V-J recombining Igλ genes or does it represent a primordial two-peptide Ig L chain that was later adapted to produce the single-peptide Igλ L chains? Examination of this system in more primitive immune systems may help answer this question.

We thank J. Stafford-Hollis, I. R. Kirsch, and W. M. Kuehl for critically reading the manuscript and S. Sammons for help with computer analysis.

Received for publication 24 August 1990 and in revised form 19 October 1990.

References


