Three transposed elements in the intron of a human $V_K$ immunoglobulin gene

Bernhard Straubinger, Eva Osterholzer and Hans G. Zachau

Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie, Universität München, FRG

Received August 10, 1987; Revised and Accepted October 16, 1987

ABSTRACT
Two gene segments coding for the variable region of human immunoglobulin light chains of the kappa type ($V_K$ genes, ref. 2) were found to have unusual structures. The two genes which are called A6 and A22 are located in duplicated gene clusters. Their restriction maps are very similar. About 4 kb of the A22 gene region were sequenced. It turned out that the intron contains an insert with the characteristics of a transposed element. The inserted DNA of 1.2 kb length contains imperfect direct and inverted repeats at its ends; at the insertion site a duplication of five nucleotides was found. Within the inserted DNA one copy each of an Alu element and of the simple sequence motif (T-G)$_7$ were identified. Also these two repetitive sequences are themselves flanked by short direct repeats. The major inserted DNA has no significant homology to published human nucleic acid sequences. The whole structure is interpreted best by assuming a sequential insertion of the three elements. The coding region of the $V_K$ gene itself has several mutations which by themselves would render it a pseudogene; we assume that the insertion event(s) occurred prior to the mutations. According to mapping and hybridization data A6 is very similar to A22.

INTRODUCTION
The eukaryotic genome is a dynamic system in which numerous rearrangements and transpositions have occurred and still occur. There are site specific rearrangements as those between the $V_K$ and $J_K^C$ gene segments (reviews 3,4). Also random transpositions of DNA sequences without detectable target site preferences are known (reviews 5-9). In mouse $K$ genes the insertion of intracisternal A particles was found (10). In human $V_K$ genes, however, up to now no insertions of foreign DNA have been detected.

Here we report the finding of an inserted element of 1.2 kb length within the intron of the $V_K$III gene A22. This gene and the structurally very similar gene A6 had been subcloned and mapped previously (11). A6 and A22 are members of two large regions of transposed elements in the intron of a human $V_K$ immunoglobulin gene
the human $V_K$ locus, the Aa and Ab regions which have recently been characterized on cosmid clones. The Aa and Ab regions span 140 and 110 kb, respectively, and each one contains 14 $V_K$ genes of different subgroups. The two A regions are structurally very similar and must have arisen by a duplication event. Therefore most of the 28 $V_K$ genes exist as pairs of very similar genes (11).

**MATERIALS AND METHODS**

**Enzymes and reagents**

Enzymes and deoxy- and dideoxynucleotides were obtained from Boehringer Mannheim. ($\alpha^{32}$P)dATP and ($\alpha^{32}$P)dCTP came from New England Nuclear, Boston. The organic and inorganic reagents were from Merck, Darmstadt.

**Cosmid library and hybridization**

Construction and screening of the cosmid library used are described in ref. 12. The library had been prepared from the placenta DNA St. The transfer of digested DNA was according to the procedure of Southern (13). The following $V_K$III gene probes were used: $V_K$III, m21-4 (14) and m41V-7 (15). Other $V_K$ gene probes and the hybridization conditions were as in ref. 11.

**Sequence analysis and computer programs**

Fragments isolated from cosmids and subclones were ligated into M13mp8, mp9, mp10, mp11 (16) and M13mp18, mp19 (17) and cloned by the transformation of JM103 cells using the CaCl$_2$ procedure (18). Recombinant phages were grown in JM103. Single-stranded phage DNA was isolated according to published methods (16). The sequences were determined by the dideoxynucleotide chain termination method (19). A 17-mer universal primer (Pharmacia) was used. Computer programs of R. Staden (20) and P.S. Neumaier (21) were applied on a PDP-11/23 processor under the RT-11 operating system (Digital Equipment Corporation). The sequence comparisons and searches were done on a VAXII/GPX computer using the UWGCG program package and the GenBank library (release 44.0; 1986). For the search of sequences homologous to the major inserted sequence (Fig. 2, positions -1262 to -31 without the Alu sequence and the simple sequence) the programs "WordSearch" (options: word size: 6; diagonals: 25; additional diagonals: 3) and "BestFit" were used.
RESULTS

When the A6 and A22 gene regions were probed with clones of the different V<sub>K</sub> subgroups only subgroup III probes gave signals; but the signals were weak even under relaxed conditions (1 x SSC, 68°C for the last washing step). Surprisingly hybridization experiments with radioactively labeled placenta DNA showed that repetitive sequences must be located within the introns of A6 and A22 (data not shown).

The A22 gene region was mapped in detail (Fig. 1). The sequence of the region spanning about 3.9 kb revealed a highly diverged V<sub>K</sub> pseudogene. The gene is more homologous to genes of subgroup III than of other subgroups but even for subgroup III 16 of 41 codons for invariant amino acids are changed or deleted (Fig. 2). Within the coding region of the gene three deletions and insertions are found: the codon for amino acid 30 and the...
eight codons for amino acids 57-64 are deleted whereas two codons are inserted between the amino acid positions 81 and 82 (Fig. 2). The leader sequence consists of 21 codons, that is one more than in other V_KIII genes (14,22). Also the decanucleotide promoter (23) and the heptanucleotide signal sequence (3,4) are mutated.

The most unusual feature of the pseudogene is the insertion of a DNA element into its intron. The element which is 1,232 bp long has characteristics of a transposed element. It is flanked by directly repeated pentanucleotides (GTTTT). The intron of the A22 pseudogene outside of the inserted element is rather homologous to the introns of other V_KIII genes as Vg and Vh (14) which contain one copy of the GTTTT sequence at the respective positions. This sequence was apparently duplicated upon insertion of the element. Analysis of the terminal sequences of the inserted element reveals overlapping inverted and direct repeat structures: a 28 bp inverted repeat with 68% homology between the two sequences (Fig. 2) and a direct repeat of about 160 bp with about 50% homology. The latter repeat is at the border of signi-

Figure 2. Sequence of the pseudogene A22 including the sequences of the inserted transposed elements. The sequence was determined according to the strategy shown in Fig. 1. A formal translation product is given which includes a stop codon at the nucleotide positions 95-97 (indicated by an asterisk). In order to stay within the V_KIII specific reading frame a one nucleotide gap was introduced at position 82. To optimize the homology to V_KIII protein sequences a deletion of one codon at the amino acid position 30 and a deletion of eight triplets at the amino acid positions 57 to 64 were assumed and an insertion of two amino acids at positions 81A and 81B had to be introduced. Invariant amino acids (95-100% invariant according to Kabat et al., 1987; for the leader sequence comparisons of 6 leader sequences by Ch. Huber, unpublished) are indicated by lines underneath the amino acid symbols. Lines above amino acid symbols mean that at these sites invariant amino acids are replaced by others. The following sequence motives are indicated by underlining: imperfect pentanucleo- deca and decanucleotides (23), TATA (27) and V-J joining signal boxes (3,4). The direct repeats (GTTTT) bordering the major insert are indicated by filled arrows. The open arrows mark the direct repeats (GAATTTCTT(G)TTT) flanking the Alu sequence and the waved arrows show the imperfect direct repeats (A ATT ATG) at the borders of the (T-G)17 sequence. The inverted repeat within the major insert (positions -1260 to -1233 and -68 to -40) is marked by a dashed line while the overlapping direct repeat (positions -1262 to -1092 and -190 to -31) is not shown. An Alu sequence downstream of the gene is indicated.
Figure 3. Schematic model illustrating the insertions of transposed sequences into the intron of the V\textsubscript{III} gene A22. The symbols indicating different sequences and sequence motives are the same as the ones used in Fig. 2. The intron-exon structure is indicated by filled boxes. One copy each of the three different direct repeats is shown (not on scale). It is unknown whether the Alu and the \((T-G)_{17}\) sequences were inserted into the major insert before or after it was itself transposed into the intron of the A22 gene.

Significance since in aligning the sequences numerous gaps have to be introduced.

Within the inserted DNA which we call the major inserted element two repetitive sequences were detected, an Alu element and a so-called simple sequence. The Alu element is highly homologous to a recently compiled consensus sequence (24) and is flanked by two nearly perfect direct repeats of 15 bp (Fig. 2). The simple sequence \((T-G)_{17}\) is located immediately downstream of the Alu element and is flanked by 9 and 10 bp direct repeats (Fig. 3). The short direct repeats indicate that the Alu element as well as the simple sequence \((T-G)_{17}\) have been inserted by transposition events into the major inserted element.

The copy number of the major inserted element in the human genome (without the Alu and the simple sequences) was not determined by hybridization experiments. In sequence comparisons the major inserted element showed no significant homology to the
published human and viral nucleic acid sequences (search conditions in Materials and Methods), but this is at best an indication that the element is not highly repetitive. No long open reading frame was found in the major inserted element.

The A6 gene region was not sequenced but its restriction map turned out to be very similar to the one of the A22 region (11). Hybridization with radioactive placenta DNA showed a repetitive sequence within the intron of A6. The hybridization with the V_{KIII} probes gave as weak signals with A6 as with A22. The similarity between the A6 and A22 regions is not surprising since the genomic regions Aa including the gene A6 and Ab containing the gene A22 are highly homologous for more than 100 kb and have most probably arisen by a duplication event. The fact that both A22 and A6 possess the described unusual features indicates that the insertion event(s) must have occurred before the duplication of the A region.

DISCUSSION

The numerous amino acid exchanges and the deletions and insertions in the V_{K} gene A22 caused problems in assigning this gene unambiguously to one of the known V_{K} subgroups. The first indication that an assignment to subgroup III is appropriate came from hybridization experiments: both the leader and the gene region hybridized better with subgroup III probes than with probes of other subgroups. Because of the distance of 1.5 kb between the fragments hybridizing with the leader and gene probes, respectively, we thought at first that we had detected two unrelated genes or parts of genes. Only the subsequent sequence analysis clarified the situation showing that the distance between leader and gene segments is due to the insertion of transposed elements. The formal translation product of the nucleotide sequence also indicated that the A22 gene was probably derived from a V_{KIII} gene; at least according to its invariant amino acids A22 is more similar to the subgroup III genes than to those of other subgroups. Another indication that A22 belongs to subgroup III is the extensive homology between its intron sequences (outside of the transposed elements) and the ones of other V_{KIII} genes as Vg, Vh (14) and O7 (K. Schäble, unpublished-
ed); also the 300 bp downstream of these \( V_k \)III genes are highly homologous to the ones of A22.

The imprecise inverted repeats at the ends of the major inserted element are reminiscent of the repeats in numerous transposed elements. Such repeats may have played a role in transposition processes (5,9,25). Similar imperfect inverted repeats had been found upstream and downstream of several human \( V_k \)I genes; this gave rise to the hypothesis that in the evolution of the kappa locus transposition events had taken place and the inverted terminal repeats had played a role in them (21).

The short direct repeats flanking all three transposed elements are clear footprints of insertion. The joint occurrence of Alu and simple sequences was observed before and was explained by model considerations (7). Multiple insertions of other repeated elements had also been seen previously, e.g. in the so-called R sequences (26) which are part of the LINE 1 family (6). We do not know whether the Alu and the \((T-G)_{17}\) sequences were inserted into the major insert before or after it was itself transposed into the intron of the A22 gene. The insertion(s) may have occurred prior to the numerous mutations within the leader and gene regions of A22 since one may assume that the inserted elements inhibit the surveillance of the sequences by gene conversion-like processes. Such processes are believed to contribute both to the homogeneity and the diversity of sequences within the multigene family. What seems clear is the fact that the insertion(s) happened in evolution before the duplication of the \( V_k \) gene regions. This conclusion rests on the above mentioned close similarity of the duplicated gene regions A6 and A22 (ref. 11), both of which contain the inserts described in the present paper.

ACKNOWLEDGMENT
The work was supported by Bundesministerium für Forschung und Technologie and Fonds der Chemischen Industrie.

REFERENCES
1. Present address: Department of Plant Pathology, Cornell University, Ithaca, New York 14853, USA.
2. The following abbreviations and nomenclature are used in this paper: L, L', part of a leader gene segment; \( V \), variable gene segment; FR, CDR, framework and complement-
arity determining regions; V. I to V. IV, variable gene segments of immunoglobulin light chains of the K type belonging to subgroups I to IV; for reasons of simplicity V. gene segments are called mostly V. genes; major insert, positions -1262 to -31 excluding positions -753 to -373 (the Alu sequence and the simple sequence); m163-4, the fourth subclone prepared in M13 from the cosmid clone cos 163; bp, base pair; kb, 10^3 base pairs;
