The EMBO Journal vol.3 no.1 pp.185 – 191, 1984

Evolution of a multigene family of V_χ germ line genes

Rolf Joho*, Howard Gershenfeld[1] and Irving L.Weissman[1]

Institute of Pathology, Division of Experimental Pathology, University Hospital, Zurich, Switzerland, and [1]Laboratory of Experimental Oncology, Department of Pathology, Stanford University, Stanford, CA 94305, USA

*To whom reprint requests should be sent

Communicated by C.Weissmann

We have isolated a series of related V_χ germ line genes from a BALB/c sperm DNA library. DNA sequence analysis of four members of this V_χ24 multigene family implies that three V_χ genes are functional whereas the fourth one (ψV_χ24) is a pseudogene. The prototype gene (V_χ24) encodes the variable region gene segment expressed in an immune response against phosphorylcholine. The other two functional genes (V_χ24A and V_χ24B) may be expressed against streptococcal group A carbohydrate. The time of divergence of the four genes was estimated by the rate of synonymous nucleotide changes. This implies that an ancestral gene has duplicated ~33–35 million years ago and a subsequent gene duplication event has occurred ~23 million years ago.

Key words: evolution/immunoglobulin genes/multigene family

Introduction

Antibody proteins are molecules which, in a single organism, show a high degree of amino acid sequence variability related to their antigen recognition functions (mediated by the variable regions of heavy and light chains), and their effector functions (mediated by the heavy chain constant regions). The immunoglobulin (Ig) variable region can be subdivided into three relatively conserved areas, called framework regions 1–3 (FW 1–3), and three hypervariable regions (HV 1–3) (Wu and Kabat, 1970; Kabat, 1978). The hypervariable regions are believed to form the antigen binding site and mainly determine the antigen binding specificity. The genes encoding the variable and constant region parts of Ig molecules are separated in the germ line and are rearranged into a functional transcription unit during maturation of B lymphocyte precursors into antibody producing B cells (for a review, see Joho *et al.*, 1983). The variable region gene segments are represented many fold in the germ line, and are held intact until their rearrangement and activation in virgin B cells. Although significant further diversification of Ig genes apparently occurs subsequent to antigenic stimulation of virgin B cells, a central question in immunology is the phylogenetic origin of the germ line representation of antibody gene diversity. Here we examine the genome representation of germ line kappa light chain genes in the V_χ24 family.

In the BALB/c mouse the antibody response to phosphorylcholine (PC) is restricted to only a few species of immunoglobulins (Claflin, 1976). Three major groups of PC-specific antibodies can be distinguished. Antibodies of all three groups share the same type of heavy chain variable region (the prototype is found in the myeloma T15) but utilize three different types of kappa light chains (the prototypes are found in the myelomas T15, M603 and M167). A single V_H gene segment in the germ line encodes the heavy chain variable region of the PC specific antibody molecules (Crews *et al.*, 1981). The amino acid sequences of the V_H regions of several different PC-specific monoclonal IgM antibodies derived from hybridomas of primary response spleen cells are identical (Gearhart *et al.*, 1981). This suggests that the V_H gene is mainly expressed in its germ line form, without any further somatic mutations, in primary response hybridoma antibodies of the IgM class. Three different V_χ germ line gene segments encode the three types of kappa light chains found in the PC response. Like the V_H gene segments, the three different V_χ gene segments are frequently expressed in their germ line form in PC-specific IgM molecules (Gearhart *et al.*, 1981). Variable region amino acid sequences of IgG and IgA molecules with specificity for PC (isolated from PC specific hybridomas or myelomas) differ at several positions from the germ line sequence found in IgMs. These amino acid differences are due to corresponding nucleotide substitutions in the expressed V_H and V_χ gene segments (Bothwell *et al.*, 1981; Crews *et al.*, 1981; Gershenfeld *et al.*, 1981; Selsing and Storb, 1981). These somatic mutations probably occur during later stages in B cell differentiation when an IgM-producing B cell switches to the synthesis of IgG, IgA, etc.

We and others have shown that a single V_χ gene segment (V_χ24) present in the germ line of the BALB/c mouse is one of the three V_χ gene segments used in an immune response to PC; it is expressed as a somatic variant in the PC-specific myelomas M167 and M511 (Gershenfeld *et al.*, 1981; Selsing and Storb, 1981; Gearhart and Bogenhagen, 1983). While V_χ24 is the only germ line gene (besides the C_χ gene) revealed by high stringency genomic Southern blot hybridization with the p167kRI cDNA probe (a cDNA plasmid derived from the kappa mRNA of M167), lower stringency washes revealed a series of weaker cross-hybridizing DNA fragments (Joho *et al.*, 1980). In this study we used the p167kRI cDNA plasmid to screen a BALB/c sperm DNA library for genomic clones bearing sequences related to the V_χ24 gene segment, expecting to isolate both the V_χ24 (and C_χ) gene and the several weakly cross-hybridizing genes which belong to this V_χ24 multigene family. We report the isolation, characterization and DNA sequence analysis of four members of this V_χ24 multigene family. Three of the four genes appear to be functional, whereas the fourth one is a pseudogene. An analysis of the DNA sequence of these genes provides the basis for a model of the evolutionary mechanisms that gave rise to these four genes from a putative common ancestor.

Results

Isolation of genomic V_χ24-related clones

We have previously demonstrated that after *Hae*III digestion of BALB/c sperm DNA up to seven DNA fragments can be identified with the p167kRI cDNA probe in a Southern blot analysis (Joho *et al.*, 1980). We isolated clones carrying V_χ24-related sequences on 950-bp, 750-bp and 520-bp *Hae*III restriction fragments that were of identical size as the ones

revealed in whole genomic Southern blot analysis (Gershenfeld *et al.*, 1981). The '950-bp clone' is the prototype of the V_x24 family. This gene is designated V_x24 and is expressed in the myelomas M167 and M511 (Gershenfeld *et al.*, 1981). The '750-bp clones' fell into two classes by further restriction mapping. We named the V_x genes of the other clones V_x24A ('750-bp clone'), V_x24B ('520-bp clone') and ψV_x24 ('750-bp clone'). Isolation of recombinant phages carrying fragments representative of the other, larger *Hae*III fragments (1.2–2.5 kb) has proved to be more difficult, and is the subject of another study. Although 15–18 kb of mouse DNA was inserted into these phage vectors, no two V_x24-related genes were found on the same insert. The restriction map of these four V_x24 like genes (following subcloning in pBR322) is shown in Figure 1. Several common (conserved) restriction sites are identified, including a *Pst*I site at the beginning of HV 1 and an *Alu*I site in FW 2 of all four genes; a *Pst*I site near the beginning of FW 2 in three cases, and an *Xba*I site in HV 3 in three cases.

DNA sequence analysis of the V_x24 multigene family

The DNA sequence of the four genes described was determined, including ~600 bp of the 5′, and ~250 bp of the 3′-flanking regions. The sequences were compared with the prototype V_x24 gene and its flanking sequence. While sequence comparison was trivial for the coding regions, the flanking sequences were matched attempting maximal homology by introducing gaps and/or insertions. Since this comparison was performed by visual inspection it may not represent the best possible match. Figure 2a–c summarize the DNA sequence data.

5′-Flanking and leader region. The coding regions for the signal peptide and the upstream 5′-flanking regions are 70–80% homologous when V_x24, ψV_x24, V_x24A and V_x24B are compared. The position of the coding region for the hydrophobic leader peptide has been determined only for the V_x24 gene segment (Selsing and Storb, 1981). The potential coding region for the V_x24A and the V_x24B gene has been inferred by analogy to V_x24.

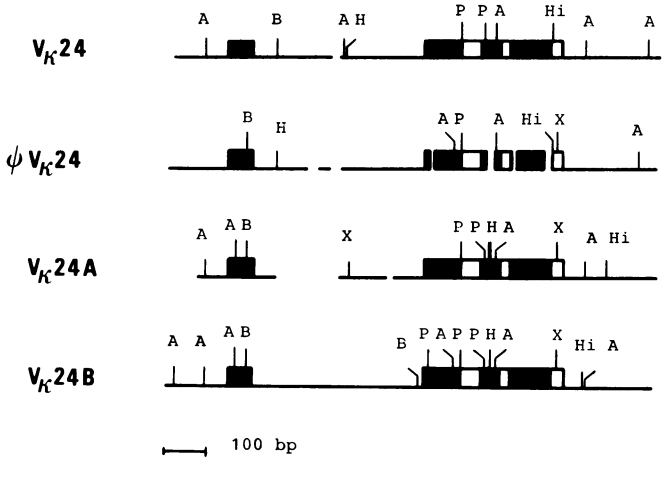

Fig. 1. Restriction site map of four V_x germ line genes of the V_x24 multigene family. Coding sequences are identified by boxes and intervening sequences by a line; large deletions are indicated by gaps. Hypervariable regions are indicated by light boxes in the coding region. A, (*Alu*I); B, (*Bam*HI); H, (*Hae*III); Hi, (*Hinc*II); P, (*Pst*I); X, (*Xba*I).

The size of the intervening sequence between the coding regions for the leader and the mature kappa chain is larger for the V_x24 gene segment (365 bp) and by inference for the V_x24A and V_x24B gene segment than for most other V_x genes (~80–170 bp) (Seidman *et al.*, 1979; Altenburg *et al.*, 1980; Max *et al.*, 1980; Kwan *et al.*, 1981; Pech *et al.*, 1981). The beginning and the end of the intervening sequence of V_x24A and V_x24B are homologous to the corresponding sequences in V_x24. However, deletions and/or insertions are scattered throughout the intervening sequence. In order to maintain sequence homology of V_x24A to V_x24, a deletion of ~135 nucleotides had to be introduced in the central part of the V_x24A sequence.

Coding regions (Figure 2b). The prototype of this multigene family is the one expressed in a response to PC and as a somatic variant in M167 and M511 (Gershenfeld *et al.*, 1981). M167 and M511 kappa light chains belong to kappa subgroup 24; the corresponding gene was therefore designated V_x24. All four genes analyzed are closely homologous. V_x24A and V_x24B are ~82% homologous to V_x24 within the coding region. An open reading frame exists in these two genes, and they can potentially be expressed. The fourth gene analyzed, ψV_x24, appears to be non-functional. At the beginning of FW 1 there is a gap of 11 nucleotides which alters the reading frame of the following coding sequence; this leads to an in-phase TAA termination codon at position 78. A gap of 16 nucleotides has to be introduced in FW 2 to maintain homology to V_x24. Three more deletions have to be introduced into the ψV_x sequence to maintain homology with V_x24. A single nucleotide deletion in HV 2 and gaps of four, and 16 nucleotides occur in FW 3. The ψV_x24 sequence is 86% homologous to V_x24 in the area that can be compared (216 of 250 nucleotide positions are identical).

3′ Flanking region. Figure 2c shows the 3′-flanking sequences. A palindromic heptamer is separated by one or two nucleotides from the 3′ end of the coding region. Both this palindrome and a conserved nonamer, separated by 12 nucleotides from the palindrome, are believed to be of crucial importance for the event of V-J joining (Sakano *et al.*, 1980). These presumptive recognition sequences can be found in the four genes described. The V_x24B gene has a single mutation in the heptamer and another in the nonamer, while in the ψV_x24 sequence three base changes (one in the heptamer and two in the nonamer) have occurred. With the exception of the pseudogene, the V_x24B gene is the only potentially functional sequenced V_x gene segment known to date in which the conserved heptamer CACAGTG has been changed. The significance of this finding will be discussed.

The V_x24 and ψV_x24 sequences are homologous to each other for as far as we have sequenced the 3′-flanking region (~200 nucleotides). A 60-bp repetitive stretch of As and Gs starts at position 110 in the ψV_x24 gene segment (underlined in Figure 2c). In this region of the ψV_x24 gene the sequence AGG is repeated 12 times. Tandem repeats of $(AGG)_9$ and $(AGG)_2$ can be detected. These structures are flanked by triplets that deviate from the basic AGG sequence by not more than one base change (A⟷G). The corresponding region of V_x24 starts with $(AGG)_4(AGA)_2$ followed by a sequence containing a few pyrimidine bases to be followed again by A and G residues. Downstream of the conserved recognition sequences the V_x24A and V_x24B genes exhibit much less homology to V_x24 (42% and 38%, respectively) than ψV_x24 does (70%); however, the former two genes are

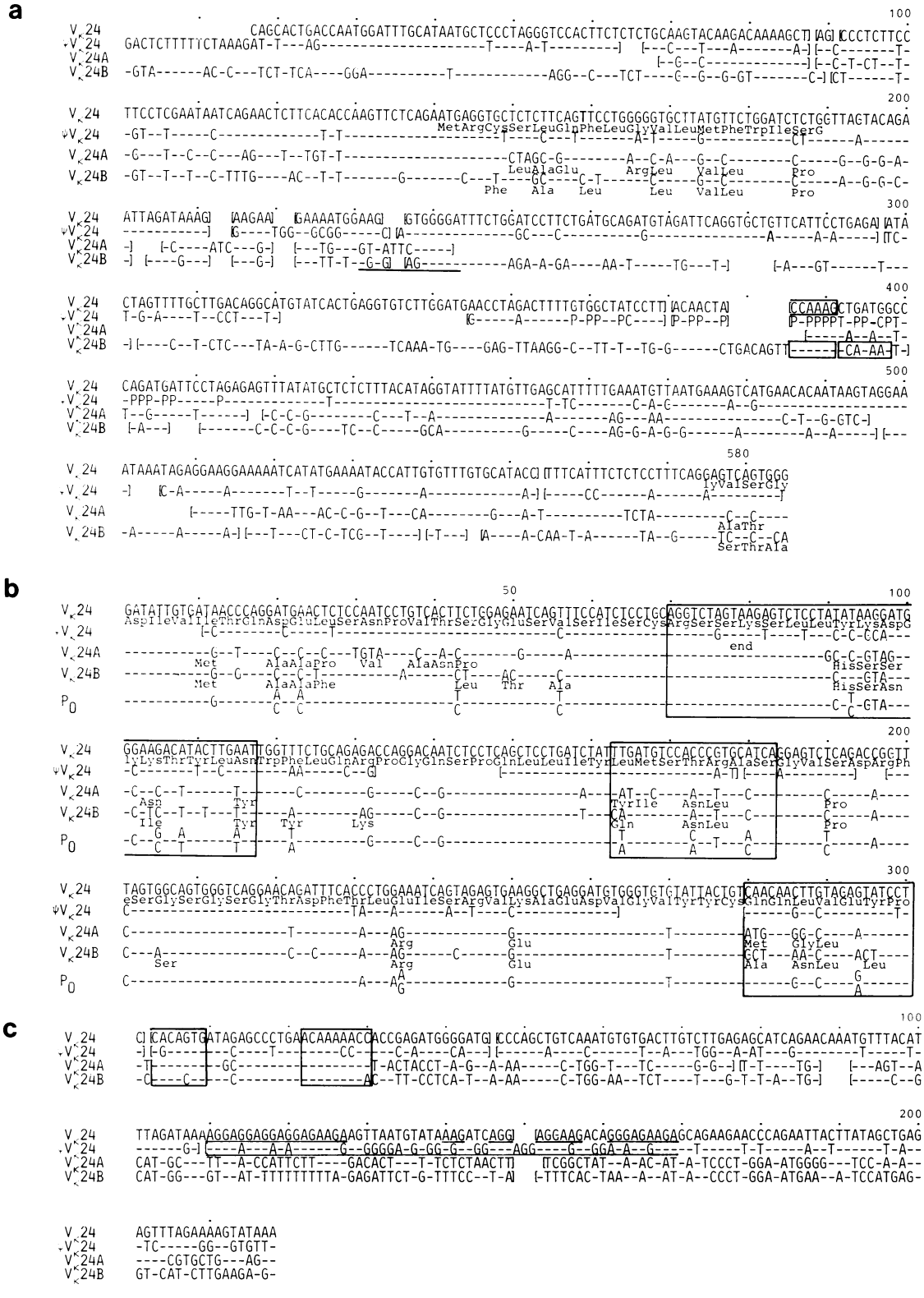

Fig. 2. DNA sequence comparison of four V_κ germ line genes. The prototype sequence $V_\kappa 24$ is on the top. Bases different from the $V_\kappa 24$ sequence are shown by the corresponding letters, identical positions by a dash (P = A or G). The beginning and the end of gaps in the DNA sequence are indicated by brackets. The corresponding three letter code for the amino acids in the protein sequence is shown below the nucleotide sequence for the $V_\kappa 24$ gene. Differences from the $V_\kappa 24$ sequence in the $V_\kappa 24A$ and $V_\kappa 24B$ gene are indicated below the corresponding nucleotide sequence. The 5′-flanking sequence of $V_\kappa 24$ from position 17 to 249 is from Selsing and Storb (1981). The remaining $V_\kappa 24$ sequence is from Gershenfeld *et al.* (1981). **a;** 5′-flanking, hydrophobic leader and intervening sequences. A sequence of eleven nucleotides identical to one found in chicken retroviruses is underlined (Varmus, 1982) and a short repeated sequence is indicated by boxes. **b;** coding region of $V_\kappa 24$, $\psi V_\kappa 24$, $V_\kappa 24A$, $V_\kappa 24B$ and a common precursor gene (P_0). Hypervariable regions are boxed in. **c;** 3′-flanking region. The presumptive recognition sequences for V-J joining are boxed in. The AG-rich region in $V_\kappa 24$ and $\psi V_\kappa 24$ is underlined.

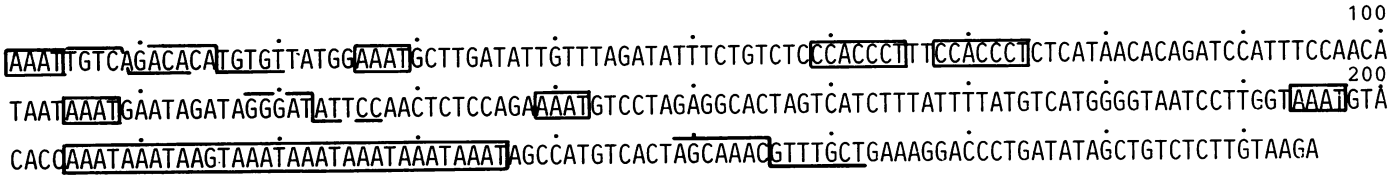

Fig. 3. Repetitive DNA near the V_x24 gene. This sequence starts ~500 bp downstream of the V_x24 gene. Internal repeats are indicated by boxes and palindromic sequences are marked by lines above and below the nucleotide sequence.

more homologous to one another (60%). At many positions V_x24A and V_x24B sequences differ from the V_x24 prototype sequence but have the same nucleotide when compared with each other. This implies that these are two related pairs of V_x genes; the first pair consists of the V_x24 and the ψV_x24 genes, and the second pair is V_x24A and V_x24B (see Discussion).

Repetitive DNA sequence downstream of the V_x24 gene segment. In an attempt to use the genomic clone carrying the V_x24 gene segment and its flanking sequences as a radioactively labeled probe for Southern blot experiments with whole genomic DNA, we discovered that this cloned DNA contained DNA sequences highly repeated in the genome (result not shown). Subsequent analysis showed that a region 500–800 bp downstream of the 3′ end of the V_x24 gene segment contains a stretch of highly repeated DNA which differs from the 31 doublet CA repeat previously reported 3′ of another V_x gene (Nishioka and Leder, 1980). Figure 3 shows the corresponding DNA sequence. Several prominent features can be detected such as a perfect palindrome of 14 bp. Internal repeats occur, e.g., the tetramer AAAT repeated eight times in tandem (with only one mismatch). The function of this stretch of 300 nucleotides is unknown.

Discussion

Evolution of coding regions

We have analyzed four genes belonging to the V_x24 multigene family. Three genes (V_x24, V_x24A and V_x24B) are probably functional whereas the fourth one (ψV_x24) is a pseudogene with several deletions and termination codons. A DNA sequence comparison of the coding regions [from gly(−4) to pro(95)] allows the construction of the nucleotide sequence of a potential ancestral gene. The most frequently occurring base was regarded as having been present in the ancestral sequence. At 15 positions such decisions cannot be made and the sequence remains ambiguous. The V_x24A and V_x24B genes have an identical base at 12 out of 15 positions. This divides this set of four genes into two pairs. V_x24 and ψV_x24 are more closely related to each other than V_x24 is to V_x24A or V_x24B; V_x24A and V_x24B on the other hand form another homologous pair.

We estimated the evolutionary distance between the genes under consideration by determining the rate of synonymous nucleotide substitutions (nucleotide changes that do not alter the encoded amino acid). The rate of synonymous nucleotide changes is approximately constant for different genes and can be used as a clock for relatively short evolutionary times (<100 million years) (Kimura and Ohta, 1972; Miyata and Yasunaga, 1980; Miyata *et al.*, 1980). Table I lists the K_S values (fraction of synonymous changes that have occurred since the two genes compared have separately evolved to the total number of synonymous positions along the two regions compared) and the time of divergence (T_d) for several pairwise comparisons of the V_x24-like genes. The pseudogene

Table I. Evolutionary distance between V_x24 like germ line genes

Comparison	K_S	T_d (years)
V_x24/V_x24A	0.2824	34.7×10^6
V_x24/V_x24B	0.2723	33.2×10^6
V_x24A/V_x24B	0.2015	23.0×10^6

K_S is the number of synonymous substitutions relative to the total number of synonymous sites (Miyata and Yasunaga, 1980) and T_d is the divergence time of the two genes compared (Miyata *et al.*, 1980).

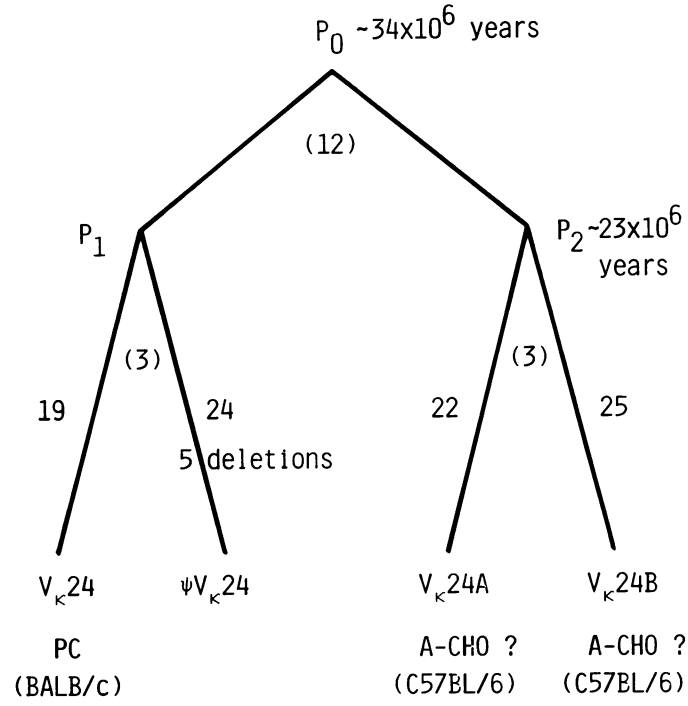

Fig. 4. Evolution of the V_x24 multigene family. P_0, P_1 and P_2 indicate ancestral genes. The numbers mark base changes between branch points and the present genes. Numbers in parentheses indicate the number of base changes that have occurred on either branch. Only the coding sequences from asp(1) to pro(95) have been compared in this analysis. PC (phosphorylcholine) and A-CHO (type A carbohydrate of *Streptococcus*) indicate the antigens against which these genes are expressed.

ψV_x24 could not be used in this analysis because synonymous changes and amino acid changes cannot be distinguished. Table I shows that V_x24A and V_x24B diverged from each other ~23 million years ago, whereas the divergence time of V_x24A and V_x24B from V_x24 is ~35 and 33 million years, respectively. The calculation of these divergence times assumes that no gene rectifying events, like gene conversion, occurred after the gene had duplicated and had separately evolved. These T_d values may therefore represent minimal time estimates.

Figure 4 summarizes a possible evolutionary relationship of

Table II. Evolution of V_χ24-related germ line genes

	Substitutions		Total number		Frequency (%)	
	nucl.	a.a	nucl.	a.a.	nucl.	a.a.
FW 1	21	12	3 x 69	3 x 23	10.1	17.4
HV 1	13	7	3 x 48	3 x 16	9.0	14.6
FW 2	8	2	3 x 45	3 x 15	5.9	4.4
HV 2	7	5	3 x 21	3 x 7	11.1	23.8
FW 3	13	4	3 x 96	3 x 32	4.5	4.2
HV 3	14	7	3 x 21	3 x 7	22.2	33.3
V_χ gene	76	37	3 x 300	3 x 100	8.4	12.3

Frequencies of nucleotide (nucl.) and amino acid (a.a.) substitutions were calculated for different parts of the V_χ gene. V_χ24, V_χ24A and V_χ24B nucleotide and amino acid sequences were compared with the potential ancestral sequence and the smallest number of substitutions leading to the genes described was used for calculation.

the coding sequences of the four genes of the V_χ24 multigene family. An ancestral gene (whose potential sequence is shown in Figure 2b) presumably duplicated ~33–35 million years ago. About 23 million years ago these two precursor genes may have simultaneously duplicated again and evolved into the four genes we observe today. This mechanism implies that the events leading to the pseudogene occurred after this second gene duplication. Alternatively only the precursor gene to V_χ24A and V_χ24B duplicated ~23 million years ago and the precursor gene to V_χ24 and ψV_χ24 may have duplicated independently before or after 23 million years. This latter possibility requires three gene duplication events to account for the four genes described.

Since immunoglobulin variable regions can be subdivided into hypervariable and framework regions, these subgenic areas could probably have evolved at different rates. A comparison of the potential ancestral gene sequence with V_χ24, V_χ24A and V_χ24B sequences (at the level of nucleotide and amino acid sequences) gives the approximate substitution frequencies for hypervariable and framework regions (Table II). Amino acid substitutions are not distributed uniformly along the V_χ region. FW 2 and FW 3 have very few amino acid exchanges (4.4% and 4.2%) contrary to HV 2 and HV 3 which have evolved much more rapidly (23.8% and 33.3%). FW 1 and HV 1 show intermediate substitution frequencies (17.4% and 14.6%) closer to the average value of the V region as a whole (12.3%). These findings imply that FW 2 and FW 3 have been conserved during recent evolution probably for reasons of structural integrity of the V_χ domain. However, HV 2 and HV 3 have diverged substantially and may have formed new antigen binding sites. The intermediate substitution frequencies of FW 1 and HV 1 are more difficult to interpret, especially the fact that FW 1 has diverged more rapidly than HV 1. However, the relatively high variability in FW 1 reflects the general variability in FW 1 seen in all sequenced kappa light chains (Kabat *et al.*, 1979). This may indicate that the sequence of FW 1 is less important in maintaining structural integrity of the variable region than the sequence of FW 2 and FW 3.

Evolution of the different mouse species has occurred fairly recently (1–2 million years ago) (Yonekawa *et al.*, 1980). Speciation of mouse and rat occurred ~10 million years ago. No substantial changes like gene duplications seem to have taken place for more than 10 million years in the V_χ24 multigene family of the analyzed genes. Therefore the gene family described was established before speciation of the mouse and rat (~10 million years ago) and should be present in other recent rodent species (unless the gene family was lost due to gene contraction events that occurred after speciation).

Evolution of flanking regions

Immediately adjacent to the coding regions are the presumptive recognition sequences thought to be involved in the process of immunoglobulin gene rearrangement (V-J joining). The pseudogene sequence shows three differences from the canonical recognition sequence, one change in the palindromic heptamer, and two changes in the conserved nonamer. These differences may be explained by the assumption that in non-functional genes evolutionary constraints upon the recognition sequences have been relieved. The V_χ24B gene is the only sequenced V_χ gene known to date with an altered palindromic heptamer. The sequence CACAGTG has been changed into CACACTG. The corresponding heptamer preceding the J_χ gene segments is the palindrome CACTGTG with the exception of the J_χ2 gene segment that is preceded by CAGTGTG (Sakano *et al.*, 1980). The heptamer preceding the J_χ2 gene segment could form a perfect stem structure with the heptamer following the V_χ24B gene segment. Base pairing with the heptamer of the other three J_χ gene segments would lead to a single base pair mismatch. Whether these differences are sufficient to lead to preferential V-J joining of certain pairs of V and J gene segments is unclear. Significantly, the V_χ24B gene segment is expressed in conjunction with the J_χ2 gene segment (Herbst *et al.*, 1982; Chang *et al.*, 1983). However, the J_χ2 gene segment is expressed in as many myelomas as the J_χ1 gene segment and about twice as often as J_χ4 or J_χ5 gene segments (Kabat *et al.*, 1979). Therefore, there seems to be no bias against J_χ2 usage. It will be interesting to analyze the palindromic recognition sequences of those germ line genes that are rearranged to, and expressed with, J_χ2. If they, too, contain the CACACTG heptamer an important element of nonrandom association of V_χ genes with J_χ gene segments will be indicated. This could reduce the estimate of antibody diversity generated by combinatorial joining.

Downstream of the conserved recognition sequences, homology of V_χ24A and V_χ24B sequences to V_χ24 sequences appears to be much less than the homology of ψV_χ24 to V_χ24. One can speculate that the break point during the first gene duplication occurred fairly close to the 3′ end of V_χ gene segment. Such an event would have abolished homologies in the 3′-flanking regions of the two precursor genes to V_χ24 and ψV_χ24, and to V_χ24A and V_χ24B.

A stretch of 60 nucleotides of A and G residues (consisting of tandem repeats of a basic structure whose common sequence is AGG) occurs at position 110 from the 3′ end of the ψV_χ24 gene segment. A similar although less perfect repeat is present in the V_χ24 sequence but not in V_χ24A and V_χ24B. The function of this repetitive DNA sequence is not known. Another type of repetitive DNA sequence can be found between 500 and 800 bp downstream of the V_χ24 gene. Sequences occurring in this latter stretch of DNA are highly repeated in the mouse genome (unpublished observation). The tandemly repeated $(AAAT)_8$ structure should have a very low melting point and may have the potential to form single strands.

The V_χ24, ψV_χ24, V_χ24A and V_χ24B genes are homologous in their 5′-flanking regions, in the coding region for the hydrophobic leader peptide and in the terminal parts of the intervening sequences. It is known that non-coding

regions in the DNA diverge more rapidly than coding regions. This rapid evolution involves base substitutions, deletions and/or insertions. To align the 5'-flanking sequences of the four members of the V_x24 multigene family for substantial stretches of homology we had to allow deletions (insertions) to occur in our sequence comparison. A particularly large deletion of ~135 bp occurs in the intervening sequence of V_x24A. Junction areas between coding and intervening sequences appear to be fairly conserved probably due to evolutionary pressure maintaining correct RNA splice signals. Many insertions and deletions have taken place since the first gene duplication (~34 million years ago). This implies that the genome is fairly dynamic as long as it does not code for any protein sequences. The fact that in a multigene family three of four related genes have remained intact, although a lot of 'DNA damage' (insertions/deletions) has occurred in the flanking sequences, implies that these genes are essential.

A very remarkable short sequence occurs at position 231 in the V_x24B sequence (underlined in Figure 2a). A stretch of 11 nucleotides is identical to a sequence in the genome of Rous sarcoma virus and an endogenous chicken provirus. This sequence is close to the 3' end of the viral genome, adjacent to the U3 sequences of viral LTR's (Varmus, 1982). All four V_x24-related genes show evidence of deletion and/or insertion at this sequence (e.g., the beginning of the large deletion in V_x24A). Perhaps this region of the common precursor gene contained a retroviral-like transposable element, and the modern-day V_x24 gene family shows the traces of its variable fate. If such an element was present, it might have provided one of the regions involved in gene duplication itself. At position 386 the sequence CCAAAG can be found in V_x24 and V_x24B to be followed by CCAAAAG at position 392 in the V_x24B gene (drawn as boxes in Figure 2a). This position corresponds to the end of the large deletion in the V_x24A gene. Therefore, the DNA that was deleted in the ancestral V_x24A gene may have been flanked by a sequence of 11 bp, that can also be found towards the 3' end of Rous sarcoma virus, and by a short repeat of CCA_3GCCA_4G. The significance of these sequences in deletion and/or insertion of DNA segments is not clear.

Expression of the V_x24 multigene family

Knowing the coding sequence of V_x24A and V_x24B, we can deduce the corresponding amino acid sequences and compare them with all sequenced kappa light chains. The predicted V_x24A amino acid sequence is with a single exception identical to the amino acid sequence of A5A idiotype-positive antibodies of A/J mice elicited in response to streptococcal group A carbohydrate (as far as the amino acid sequence of the variable region part has been established) (Kabat *et al.*, 1979). The deduced V_x24B amino acid sequence with a single exception is identical to the kappa light chain of the myeloma SAMM368 which has an as yet unknown antigen binding specificity and was previously placed into kappa subgroup 25 (Kabat *et al.*, 1979). Herbst *et al.* (1982) and Chang *et al.* (1983) prepared hybridomas from BALB/c x C57BL/6 mice (CxB) immunized with streptococcal group A carbohydrate. Amino acid sequence analyses of the variable regions of heavy and light chains of hybridoma antibody molecules with specificity for *Streptococcus* group A implies that as few as three V_x germ line genes are expressed in an immune response to streptococcal group A carbohydrate. The described BALB/c germ line genes V_x24A and V_x24B are very similar to the light chain genes expressed in these hybridomas. Figure 5 summarizes the amino acid sequences encoded by the V_x24, V_x24A and V_x24B genes and the amino acid sequences found in anti-streptococcal group A carbohydrate-specific hybridoma antibodies. Differences from the BALB/c germ line genes may be due to somatic mutations, or to allelic differences between BALB/c and C57BL/6 (if the C57BL/6 allele is expressed in these hybridomas). It is also possible that very closely related members of the V_x24 multigene family (that have not yet been cloned and sequenced) are expressed in these hybridomas. The antibodies analyzed have isotypes different from IgM, and it is therefore possible that they have accumulated several somatic mutations similar to the antibodies of M167 and M511 (Gershenfeld *et al.*, 1981; Selsing and Storb, 1981; Gearhart and Bogenhagen, 1983) or the hybridoma antibodies raised in an immune response against PC (Gearhart *et al.*, 1981). The findings summarized in Figure 5 strongly suggest that two members of the V_x24 multigene family (V_x24A and V_x24B) are used in an immune response against the carbohydrate moiety of *Streptococcus* Group A, while the V_x24 gene is expressed in response to phosphorylcholine, a cell wall constituent of *Streptococcus pneumoniae* (strain R36A). It is formally possible that only the V_x24A and V_x24B alleles of C57BL/6 are used against streptococcal group A carbohydrate whereas the BALB/c alleles are expressed in an immune response against a different antigen. If this were the case, different alleles of the same genetic loci (V_x24A and V_x24B) would have different functions (antigen recognition). Study of the immune

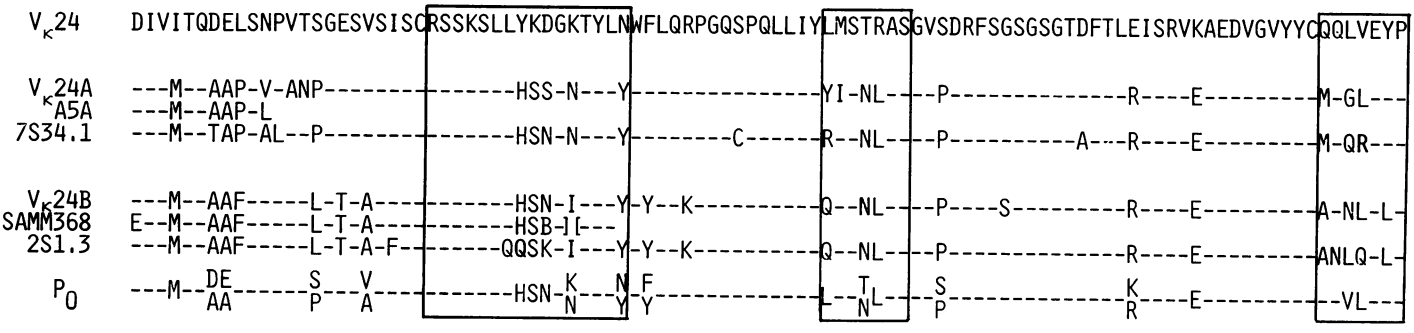

Fig. 5. Amino acid sequences encoded by genes of the V_x24 multigene family. The single letter code for amino acids is used. A dash denotes a position identical to the prototype sequence of V_x24. A5A is a partial amino acid sequence of serum antibody kappa light chain against Streptococcus group A; SAMM368 is a partially sequenced IgA producing myeloma (Kabat *et al.*, 1979). 7S34.1 and 2S1.3 are kappa chain sequences of hybridoma antibodies directed against streptococcal group A carbohydrate (Herbst *et al.*, 1982; Chang *et al.*, 1983). P_0 denotes the potential amino acid sequence of an ancestral gene. Hypervariable regions are indicated by boxes.

response against streptococcal group A carbohydrate in BALB/c should allow resolution of this question.

Coupled evolution of Ig genes and their 'cognate antigens'?

As already meantioned, the immune response to phosphorylcholine (*S. pneumoniae* strain R36A) uses three different types of kappa light chains encoded by three different V_κ germ line genes, one of which is $V_\kappa 24$. The single V_H gene segment expressed belongs to the $V_H III$ subgroup (Crews *et al.*, 1981; Gearhart *et al.*, 1981). It is interesting that the streptococcal group A carbohydrate-specific antibodies express a V_H gene segment, which is similar, but not identical, to the V_H gene segment used in an anti-*S. pneumoniae* response (Herbst, 1982). Some of the differences may be due to different alleles in BALB/c and C57BL/6, or to somatic mutations, as discussed earlier. This finding together with the analysis of the light chain system raises an interesting possibility regarding the evolution of immunoglobulin genes and the immune response. The fact that members of the same multigene family are used in a response to apparently unrelated antigens raises the question as to whether there was a coupled evolution of immunoglobulin genes (and the immune response due to their expression) and their 'cognate antigens'. The precursor gene that gave rise to the $V_\kappa 24$ multigene family may have been expressed in conjunction with an ancestral $V_H III$ gene. This ancestral antibody could have been directed against a 'precursor antigen', i.e., against a common ancestor of *S. pneumoniae* and of type A streptococci *(S. pyogenes)*; or it could have been directed against similar (cross-reactive) immunogenic determinants on the surface of two different bacterial precursors that subsequently evolved into *S. pneumoniae* and *S. pyogenes*, respectively. If duplication of Ig genes had occurred prior to the 'evolution' of the cognate antigen, then the immune system would have been able to adapt its response to the 'evolving antigen' (i.e., bacteria) by mutation of some of its germ line genes and selection of those genes that were best adapted to cope with the newly arising antigens. The $V_\kappa 24$ multigene family as we observe it today may represent such an example.

Materials and methods

Materials

Restriction endonucleases were from Bethesda Research Laboratories; calf alkaline phosphatase from Boehringer, Mannheim; and T4 polynucleotide kinase from PL biochemicals. Low gel temperature agarose from BIORAD Laboratories was used for preparative gel electrophoresis. [γ-^{32}P]ATP was received from the Institute of Molecular Biology I, University of Zurich.

V_κ gene cloning and characterization

Construction of the BALB/c mouse sperm DNA library in phage Charon 4A (Davis *et al.*, 1980) and isolation of phage clones bearing V_κ-like sequences has been described (Gershenfeld *et al.*, 1981). DNA of such phage isolates was purified and digested with restriction endonucleases *Bam*HI and/or *Eco*RI. Restriction fragments carrying V_κ gene segments were subcloned into pBR322. Restriction maps of the V_κ genes and their flanking sequences were established by single or double digests of cloned DNA with several restriction endonucleases and the resulting fragments were analyzed by electrophoresis through 0.7–2.0% agarose in 40 mM Tris-acetate pH 8.3, 2.0 mM EDTA.

DNA sequencing

DNA fragments to be sequenced were electroeluted from low gel temperature agarose (BIORAD), extracted once with buffer-saturated phenol and precipitated with ethanol. Labeling and sequencing of DNA fragments was performed as described (Maxam and GIlbert, 1980).

Alignment of nucleotide sequences and calculation of evolutionary distance

Nucleotide sequences were aligned by visual inspection attempting maximal homology. Deletions were introduced when they were necessary to maintain stretches of homology. The boundaries of deletions cannot always be determined precisely and may be shifted in some instances by a few nucleotides to the left or right. The calculation of the evolutionary distance of the coding region was done as described (Kimura and Ohta, 1972; Miyata and Yasanuga, 1980; Miyata *et al.*, 1980). The evolutionary distance (T_d) of two coding region sequences compared relates to the synonymous nucleotide difference (K_s) by the equation $T_d = -(3/4)\ln(1-(4/3)K_s)/(2 \times 5.1 \times 10^{-9})$ years.

Acknowledgements

The authors wish to thank several members of the Institute for Molecular Biology I, University of Zurich, for the generous supply of [γ-^{32}P]ATP, and Drs. M. Billeter, J.Lindenmann and F.Rupp for critically reviewing the manuscript and M.Bucher for expert secretarial help. We thank Drs. R. Aebersold, D.Braun, J.Y.Chang and D.Herbst for making their results available to us prior to publication. This work was supported by grant 3.765-0.80 of the Schweizerischer Nationalfonds to R.J., and USPHS grant AI 18162 to I.L.W. H.G. was a predoctoral fellow of the Stanford Cancer Biology Program.

References

Altenburg,W., Steinmetz,M. and Zachau,H.G. (1980) *Nature,* **287**, 603-607.
Bothwell,A.L.M., Paskind,M., Reth,M., Imanishi-Kari,T., Rajewsky,K. and Baltimore,D. (1981) *Cell,* **24**, 625-637.
Chang,J.Y., Herbst,H., Aebersold,R. and Braun,D.G. (1983) *Biochem. J.,* in press.
Claflin,J.L. (1976) *Eur. J. Immunol.,* **6**, 669-674.
Crews,S., Griffin,J., Huang,H., Calame,K. and Hood,L. (1981) *Cell,* **25**, 59-66.
Davis,M.M., Calame,K., Early,P.W., Livant,D.L., Joho,R., Weissman,I.L. and Hood,L. (1980) *Nature,* **283**, 733-739.
Gearhart,P.J., Johnson,N.D., Douglas,R. and Hood,L. (1981) *Nature,* **291**, 29-34.
Gearhart,P.J. and Bogenhagen,D.F. (1983) *Proc. Natl. Acad. Sci. USA,* in press.
Gershenfeld,H.L., Tsukamoto,A., Weissman,I.L. and Joho,R. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 7674-7678.
Herbst,H. (1982) Thesis, Johann Wolfgang Goethe Universität, Frankfurt am Main.
Herbst,H., Chang,J.Y., Aebersold,R. and Braun,D.G. (1982) *Hoppe-Seyler's Z. Physiol. Chem.,* **363**, 1069-1076.
Joho,R., Weissman,I.L., Early,P., Cole,J. and Hood,L. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1106-1110.
Joho,R., Nottenburg,C., Coffman,R.L. and Weissman,I.L. (1983) *Curr. Top. Dev. Biol.,* **18**, 15-58.
Kabat,E.A. (1978) *Adv. Protein Chem.,* **32**, 1-75.
Kabat,E.A., Wu,T.T. and Bilofsky,H. (1979)
Kabat,E.A., Wu,T.T. and Bilofsky,H. (1979) in Sequences of Immunoglobulin Chains.
Kimura,M. and Ohta,T. (1972) *J. Mol. Evol.,* **2**, 87-90.
Kwan,S., Max,E.E., Seidman,J.G., Leder,P. and Scharff,M.D. (1981) *Cell,* **26**, 57-66.
Max,E.E., Seidman,J.G., Miller,H. and Leder,P. (1980) *Cell,* **21**, 793-799.
Maxam,A.M. and Gilbert (1980) *Methods Enzymol,* **65**, 499-560.
Miyata,T. and Yasunaga,T. (1980) *J. Mol. Evol.,* **16**, 23-36.
Miyata,T., Yasunaga,T. and Nishida,T. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 7328-7332.
Nishioka,Y. and Leder,P. (1980) *J. Biol. Chem.,* **255**, 3691-3694.
Pech,M., Höchtl,J. and Zachau,H.G. (1981) *Nature,* **291**, 668-670.
Sakano,H., Maki,R., Kurosawa,Y., Roeder,W. and Tonegawa,S. (1980) *Nature,* **286**, 676-683.
Seidman,J.G., Max,E.E. and Leder,P. (1979) *Nature,* **280**, 370-375.
Selsing,E. and Storb,U. (1981) *Cell,* **25**, 47-58.
Varmus,H.E. (1982) *Science (Wash.),* **216**, 812-820.
Wu,T.T. and Kabat,E.A. (1970) *J. Exp. Med.,* **132**, 211-250.
Yonekawa,H., Moriwaki,K., Gotoh,O., Watanabe,J., Hayashi,J., Miyashita,N., Petras,M.L. and Tagashira,Y. (1980) *Jap. J. Genet.,* **55**, 289-296.

Received on 5 October 1983