The EMBO Journal vol.4 no.13A pp.3439–3445, 1985

Light chain germ-line genes and the immune response to 2-phenyloxazolone

Jos Even[1], Gillian M.Griffiths[2], Claudia Berek and Cesar Milstein

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[1]Present address: INSERM U255, Institut Curie, 26 rue d'Ulm, 75231 Paris, Cedex 05, France
[2]Present address: Department of Pathology, Stanford University Medical Center, Stanford, CA 94305, USA

Communicated by C.Milstein

Direct sequencing of mRNA has shown that the early primary response of the BALB/c mouse to the hapten 2-phenyloxazolone is dominated by antibodies with a particular light chain, V_χ-Ox1. Although the V_χ-Ox1 sequence is still commonly expressed later in the response it now includes a number of nucleotide changes. From two independent BALB/c germ-line DNA libraries 13 different genes hybridizing to a V_χ-Ox1 probe were isolated and characterized. Two are identical to mRNA sequences found in the early primary response, one of which is the V_χ-Ox1 sequence. None of the germ-line clones show the characteristic nucleotide changes contained in the late anti-phenyloxazolone light chain mRNAs. These results demonstrate that the V_χ-Ox1 sequence used in the early primary response is entirely encoded by the germ-line and further substantiate the importance of somatic mutations in the maturation of the anti-phenyloxazolone response. The statistical analysis of the data shows that the V_χ-Ox1 related germ-line gene family contains >20 and probably <50 genes.

Key words: mouse antibody diversity/immunoglobulin light chain genes/nucleotide sequences/oxazolone response

Introduction

The diversity of antibodies is generated at various levels (Tonegawa, 1983). To start with, there are a considerable number of the different gene segments V_HD and J_H for the heavy and $V(_\chi$ or $_\lambda)$ and $J(_\chi$ or $_\lambda)$ for the light chains contained in the germ-line DNA (germ-line diversity) that can be joined to generate a functional variable region (combinatorial diversity). The joining process is imprecise and adds to the diversity (junctional diversity). In addition there is ample evidence that somatic point mutations can increase the diversity particularly during the maturation of the immune response (mutational diversity) (Griffiths *et al.*, 1984; Tonegawa, 1983). So far most of the information has been obtained by comparing the nucleotide sequences of germ-line elements with proteins or DNA (and more recently mRNA) of myelomas and hybridomas. But for few examples the majority of the structural data obtained in the past were rather static and gave little information about the development of the diversity during the maturation of the immune response.

In our laboratory we have undertaken a systematic study of the structural changes of the antibodies at different stages of the response to the hapten 2-phenyloxazolone (phOx) by direct sequencing of mRNAs from hybridomas derived at different times after immunisation of BALB/c mice (Berek *et al.*, 1985; Griffiths *et al.*, 1984; Kaartinen *et al.*, 1983a, 1983b). These studies have defined two variable region sequences, V_H-Ox1 in the heavy chain and V_χ-Ox1 in the light chain, that correlate with the best affinity antibodies and with idiotypic specificities which dominate the anti-phOx primary response (Kaartinen *et al.*, 1983b). Those studies further suggested that somatic mutations play a major role in the maturation of the response. The importance of V_χ-Ox1 and closely related light chains was true not only during the early and late primary responses but even in the secondary response, when there is a noticeable shift to other germ-line genes in the generation of high affinity antibodies (Berek *et al.*, 1985; Griffiths *et al.*, 1984).

In this paper we have analysed the complexity of V_χ-Ox1 related germ-line genes by isolating 16 independent clones from two BALB/c germ-line libraries. Using a V_χ-Ox1 cDNA probe we have characterized 13 germ-line genes. Two of them participate in the response and one is the predicted V_χ-Ox1 germ-line gene. None of them showed the pattern of changes characteristic for high affinity antibodies found in the late response to phOx. This further substantiates the importance of somatic mutation in the maturation of an immune response.

Results

V_χ-Ox1 cDNA synthesis and Southern blotting

To make possible the analysis of the V_χ-Ox1 encoding and related germ-line V-regions we synthesized cDNA from mRNAs isolated from the hybridoma NQ2.48.2.2 (Kaartinen *et al.*, 1983b) and cloned it into M13 (Messing, 1983) and pAT153 (Twigg and Sherratt, 1980). This cDNA containing 199 bases of the 3′ end of the prototype V_χ-Ox1-region and the 36 adjacent bases of the $J_{\chi 5}$ segment, was used as a probe on Southern blots. The results in Figure 1b show that BALB/c germ-line DNA contains at least 10 different bands when digested with *Hind*III or with *Eco*RI. Some of these bands are broader and more intense than average suggesting the presence of more than one V_χ-Ox1 related restriction fragment per band. *Hind*III-digested germ-line DNAs (Figure 1a) from BALB/c (lane 2) and A/J (lane 3) yield an identical pattern of bands different from that of AKR DNA (lane 4) although some of the bands seem to be conserved between the three mouse strains. DNA from the V_χ-Ox1 light chain-producing hybridoma NQ5.96.2 (Kaartinen *et al.*, 1983b) (lane 1) shows an additional band of ~3.0 kb, that is neither present in BALB/c germ-line DNA (lane 2) nor in DNA from the cell-line NSO (Galfre and Milstein, 1981), the fusion partner in the hybridoma (not shown). This band probably corresponds to the rearranged V_χ-Ox1 gene because a J_χ probe detects the 3.0-kb band only in NQ5.96.2 DNA and not in germ-line nor in NSO DNAs (not shown). The faint 2.8-kb band in Figure 1a probably contains the $J_{\chi 5}$ sequences contained in our probe since re-hybridization of the same blot with a J_χ-specific probe detects the same 2.8-kb band in all the lanes (not shown).

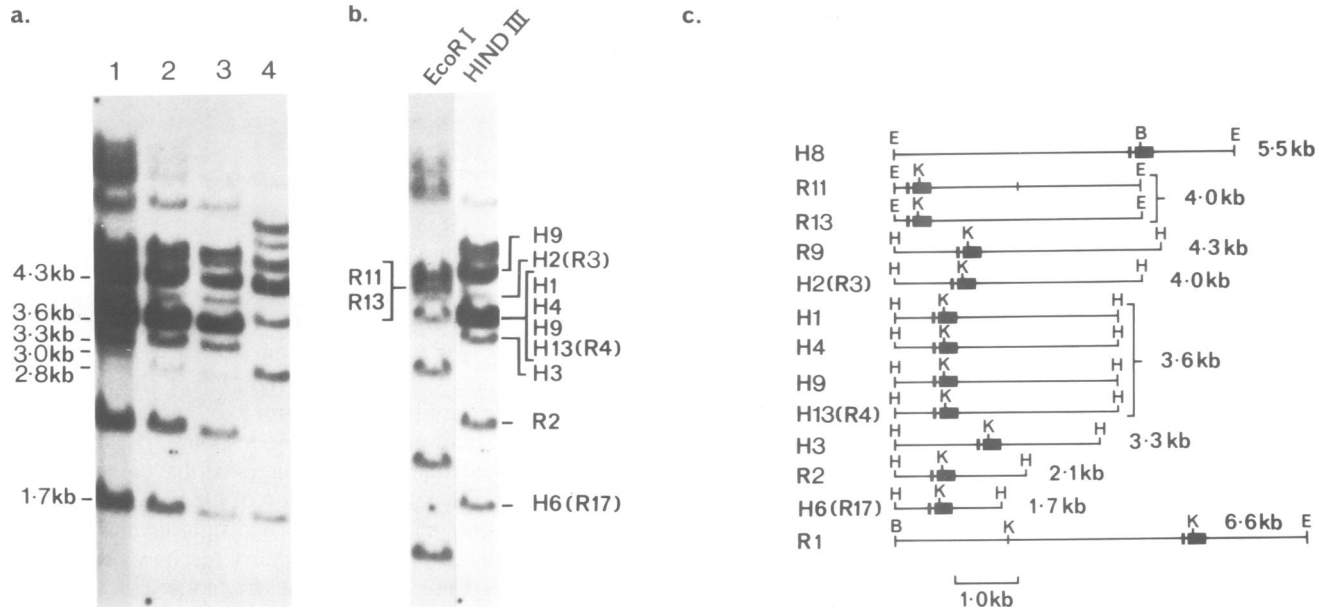

Fig. 1. Southern blot analysis of mouse DNAs with a V_x-Ox1 cDNA probe. 10 – 15 μg of DNA was digested with *Eco*RI and *Hind*III and separated on a 0.7% agarose gel, transferred to nitrocellulose paper and hybridized to nick-translated V_x-Ox1 cDNA insert. (**a**) All DNAs were digested with *Hind*III, **lane 1** contains DNA from hybridoma NQ5.96.2 (Kaartinen *et al.*, 1983b) producing a V_x-Ox1 mRNA and **lanes 2 – 4** contain mouse germ-line DNA from BALB/c (**lane 2**), A.J (**lane 3**) and AKR (**lane 4**). (**b**) Both lanes contain BALB/c liver DNA restricted with *Eco*RI or *Hind*III and the names of the different clones are assigned to bands on the blot and their schematic maps are represented in (**c**) the black boxes (▬) correspond to the leader and V coding regions. The external restriction sites B (*Bam*HI), E (*Eco*RI) and H (*Hind*III) correspond to the plasmid cloning sites. The internal restriction sites K (*Kpn*I) and B (*Bam*HI) are the additional sites used to clone fragments into M13mp18 and mp19 (Yanisch-Perron *et al.*, 1985) which were then used for sequencing.

Isolation and characterisation of V_x germ-line clones

About 5 × 10^5 plaques from two different BALB/c germ-line lambda libraries (Early *et al.*, 1979; Ollo *et al.*, 1981) were screened with the V_x-Ox1 cDNA probe and a total of 30 hybridizing phages were identified and purified. Reprobing the clones with the J_x probe revealed two clones that were not analysed any further. DNA was extracted from 16 different clones [eight from each library referred to as H (Early *et al.*, 1979) and R (Ollo *et al.*, 1981)] and analysed by agarose gel electrophoresis after digestion with *Bam*HI, *Eco*RI and *Hind*III and after double digestion with *Bam*HI-*Eco*RI, *Bam*HI-*Hind*III and *Eco*RI-*Hind*III restriction endonucleases. Ethidium bromide staining of the gels revealed a unique band pattern for each phage, proving that they had been the result of 16 independent cloning events. Southern blots hybridized with the V_x-Ox1 probe identified the DNA fragments containing the homologous V region sequences to be subcloned in plasmids. The sequences of the V_x-Ox1 related RNAs contain *Kpn*I and/or *Bam*HI restriction site(s) within the coding regions (Kaartinen *et al.*, 1983a; Berek *et al.*, 1985), therefore the subcloned DNAs were digested with *Kpn*I or *Bam*HI and the restriction enzymes used in the plasmid cloning step. The fragments could thus be inserted into M13 vectors in such a way that the coding regions could be directly sequenced by the dideoxy chain termination method. The comparison of their nucleotide sequences (Figure 2) revealed that the 16 clones contained 13 different germ-line V-region genes. Figure 1c shows the restriction map of the different clones. Some clones yield restriction fragments of identical size (Figure 1b). Thus the 3.6-kb *Hind*III band contains a minimum of four and the 4.0-kb *Eco*RI band contains at least two different V regions. Clones H2 and H8 are pseudogenes that each contain one stop codon in the coding region as the only detectable alteration.

V_x-Ox1 related germ-line sequences

A comparison of the nucleotide sequences with V_x-Ox1 showed various degrees of homology. Clone H3 matched exactly the mRNA V_x-Ox1 sequence (Kaartinen *et al.*, 1983a) demonstrating that V_x-Ox1 is entirely encoded by the germ-line and that the germ-line DNA can give rise to a functional anti-phOx light chain upon rearrangement without undergoing somatic mutations. H2, the most distantly related gene still shows 88% homology with V_x-Ox1. The sequences were also compared with published nucleotide and protein sequences. The nucleotide sequence of clone R1 is identical to that of the S107B gene (Kwan *et al.*, 1981) and the sequence of clone H6 translates into a protein identical to anti-galactan light chain V-regions (Pawlita *et al.*, 1982) suggesting that clone H6 corresponds to the germ-line gene used in the anti-galactan response. With the exception of clone R1 all the other 12 clones contain nucleotide sequences never described before although H3 was predicted as the germ-line V_x-Ox1 gene. To our list of V_x-Ox1 related germ-line genes, we have added the sequence of the germ-line gene L8 (91% homology with V_x-Ox1) (Höchtl *et al.*, 1982).

The sequences were grouped according to their relatedness (Figure 2) as described in this section. Several clones contain two additional codons in CDR1 and a gap was introduced to minimize base pair mismatching. This gap clearly falls into positions 29 and 30 in the Kabat nomenclature (Kabat *et al.*, 1983). The need to put gaps in positions 29 – 32 has been noticed by Potter *et al.* (1982) but the gap position assigned by us is different from theirs. To establish the relationship between the different V-coding regions we first did quantitative pair-wise comparisons of all the different genes (Figure 3). The results are plotted in several ways in Figure 4. Figure 4a shows the distance from H3-V_x-Ox1 to all the other genes. A plot of all the pair-

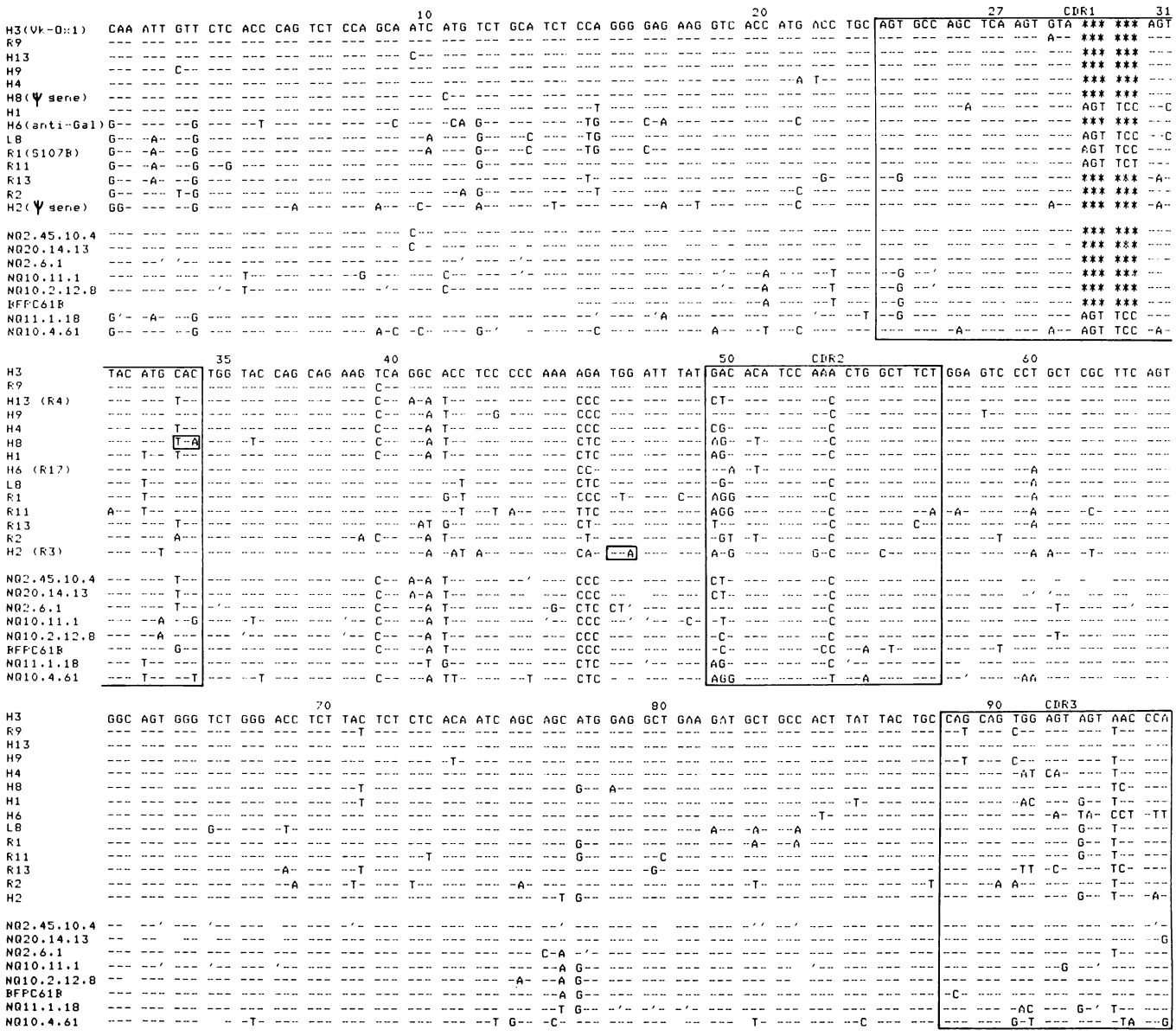

Fig. 2. Nucleotide sequences of V_x-Ox1 related germ-line genes and mRNAs. The sequence of germ-line gene H3, identical to the major light chain V region in the anti-oxazolone response V_x-Ox1 (Kaartinen *et al.*, 1983a), was used as reference in the comparisons. A dash (—) indicates sequence identity. An asterisk (*) indicates a gap introduced in the shorter sequences, an apostrophe (') indicates a likely identity with the reference sequence. A blank corresponds to an undetermined residue in an mRNA sequence. The boxed areas correspond to CDR1, CDR2 and CDR3. The boxed codons are stop codons. The L8 germ-line gene sequence is from (Hočhtl *et al.*, 1982) clone R1 is indentical to S107B (Kwan *et al.*, 1981) and the other germ-line gene sequences were determined by us. The BFPC61B cDNA sequence, lacking the 14 first codons, is from Bernard *et al.* (1981). The NQ20.14.13 sequence was obtained by C.Berek (unpublished) and the other mRNA sequences are from (Berek *et al.*, 1985; Kaartinen *et al.*, 1983b). The NQ2.45.10.4 sequence is a revised and completed version of that published in Kaartinen *et al.* (1983b).

wise comparisons of the germ-line sequence (Figure 4e) shows a bimodal distribution in which the first group (I) contains the more closely related gene pairs, indicating the presence of one or more closely related subfamilies. The genes involved have been marked in Figure 3 and include H3 through H1 in one subfamily and R1, L8 and R11 in another. In the subfamily of the genes most closely related to V_x-Ox1 (H3 to H1) H3 and R9 are almost identical whereas the other genes share a characteristic pattern of differences in FR2 and CDR2 (Figure 2). This shows that within this subfamily there are two subsets of genes, one containing H3 and R9 and another containing H9, H13, H4, H8 and H1. Likewise, in the other subfamily, R1 and L8 may belong to one subset and R1 to another. H6 has similarities with both subfamilies and is positioned between the two. R13, R2 and H2 show some elements in FR1 that put them closer to the L8-R1 subfamily but they cannot be further classified using the same criteria as for the preceding genes.

Germ-line genes and mRNA sequences of anti-phOx antibodies

The response to phOx, particularly at the early stage, is dominated by antibodies containing light chains homologous to those defined by the H3-V_x-Ox1 germ-line gene (Berek *et al.*, 1985; Griffiths *et al.*, 1984; Kaartinen *et al.*, 1983b). Careful analysis allows a distinction to be made between those sequences most likely derived from germ-line gene H3-V_x-Ox1, and those likely to be derived from other members of the same family. In Figure 4b–d

J.Even *et al.*

	H3	R9	H13	H9	H4	H8	H1	H6	L8	R1	R11	R13	R2	H2	NQ2.45.10.4	NQ2.6.1	NQ10.11.1	NQ10.2.12.8	BFPC61B *	NQ11.1.18.
H3 (Vk-Ox1)																				
R9	6																			
H13 (R4)	12	16																		
H9	14	12	12																	
H4	17	19	10	15																
H8	19	19	14	18	16															
H1	21	21	16	20	15	16														
H6 (R17)	26	31	34	35	35	37	41													
L8	25	29	29	30	31	32	25	34												
R1	28	32	29	31	31	31	28	33	12											
R11	27	31	31	32	33	31	30	40	24	22										
R13	25	27	27	30	26	27	29	36	31	33	36									
R2	27	28	28	28	29	29	31	37	36	38	39	39								
H2 (R3)	34	36	38	39	40	40	42	42	41	38	37	42	46							
NQ2.45.10.4	12	16	0	12	10	14	16	34	29	29	31	27	28	38						
NQ2.6.1	15	17	12	14	15	16	18	37	30	31	33	28	29	39	12					
NQ10.11.1	21	25	17	21	23	19	28	43	38	37	39	37	38	44	17	21				
NQ10.2.12.8	18	22	15	18	20	19	25	40	35	34	36	34	33	41	15	16	8			
BFPC61B *	20	24	15	19	21	21	26	35	31	30	33	32	29	36	15	19	13	10		
NQ11.1.18.	22	26	23	25	22	23	17	37	23	21	23	24	35	34	23	24	28	25	24	
NQ10.4.6.1.	43	44	41	44	42	43	41	49	45	44	47	49	46	49	41	45	49	47	41	43

Fig. 3. Degree of relatedness between V_x-Ox germ-line genes and mRNA sequences. The sequences were compared in pairs using the computer programs TWO and TWOB (R.Staden, unpublished) and the distances between the two sequences in each pair expressed in genetic events were plotted in a double entry table. The boxed areas correspond either to subfamilies of genes or to comparisons mentioned in the Results and Discussion sections. To make possible the comparisons of mRNA sequences the likely identical (') and the undetermined (blanks) residues were assumed to be identical to the corresponding residues in the reference sequence (Figure 2). The mRNA data refer to sequences which were originally described as probably not derived from the V_x-Ox1 but more likely from other related germ-line genes (Berek *et al.*, 1985; Kaartinen *et al.*, 1983b). (*) Does not include the 42 bases at the 5′ end.

we have compared all those belonging to the first category with the germ-line V_x-Ox1 sequence. The comparison of the sequence clearly shows the accumulation of mutations which correlate with the maturation of the response. Even during the secondary response the distance from the germ-line gene H3-V_x-Ox1 to all somatic mutants is lower than the distance separating any pairs of germ-line genes except the pair R9 and H3 who differ only by six genetic events. Comparison of these mRNA sequences to R9 and all other germ-line genes (not shown) showed that in all cases H3-V_x-Ox1 was by far the most closely related sequence. This points to H3 as the germ-line sequence from which all those mRNAs have been derived by somatic mutation. Furthermore, when all mRNA sequences discussed above were compared pair-wise with each other the differences for the primary response at day 14 (Figure 4g) and the secondary response (Figure 4h) shift towards higher diversity, as a single homogeneous population set. The maximum distances remain within the lower values obtained in the germ-line gene comparisons (Figure 4e).

The other set of mRNAs which were considered as related rather than derived from V_x-Ox1 have different properties. Among them, NQ2.45.10.4 (Kaartinen *et al.*, 1983b) mRNA and the H13 germ-line gene seem to have identical sequences, indicating that H13 codes for a functional anti-phOx light chain. Not surprisingly, NQ2.45.10.4 is a hybridoma from the early primary (day 7) response. Recently we have observed another example from a different day 7 response in which this germ-line gene has been used (see NQ20.14.13 in Figure 2). The sequences of other mRNAs (Figure 4i) differ from each other as much as the 14 germ-line genes themselves (Figure 4e). The nearest pair is NQ10.11.1 and NQ10.2.12.8 which differ by eight genetic events. Since these two antibodies are from the secondary response there is a possibility that they derive from the same germ-line gene. Even BFPC61B may derive from the same germ-line gene although the first 42 nucleotides ae missing from its sequence. On the other hand, the 12 differences between H13-NQ2.45.10.4 and NQ2.6.1, although within the range of variation of secondary response hybridomas (Figure 4h), is very high for hybridomas obtained at day 7 in the primary response (Figure 4b). Therefore, NQ2.6.1 is also likely to derive from a germ-line gene different from the 14 genes in Figure 2. NQ2.6.1, NQ10.11.1, NQ10.2.12.8 and BFPC61B belong to the subfamily of genes most closely related to H3 and more precisely to the subset H9 to H1 while NQ11.1.18 belongs to the L8-R1 subfamily. NQ10.4.6.1 is the most distantly related sequence present in our comparisons and like the germ-line sequence H6 does not fit in any of the subfamilies although it contains FR1 and CDR1 regions closer to L8-R1 and its CDR2 shares characteristics of the subset H9-H1.

In summary we conclude that NQ2.6.1, NQ11.1.18, NQ10.4.6.1 and the group NQ10.11.1.-NQ10.2.12.8-BFPC61B represents transcripts of 4–6 germ-line genes hitherto undescribed.

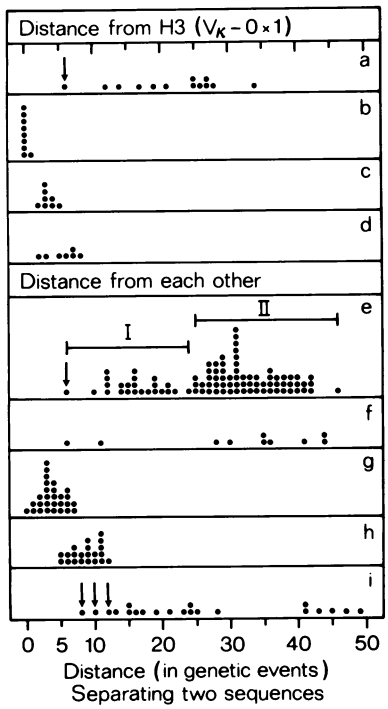

Fig. 4. Comparisons of germ-line genes and mRNA sequences. The coding regions of germ-line genes and mRNAs were compared using the computer program TWOB (R.Staden, unpublished) and the results expressed in distances in genetic events were plotted; each dot (●) corresponds to one comparison. In (**a**)–(**d**) sequences were compared with H3-V_x-Ox1. The sequences compared were in: (**a**) the 13 germ-line genes (R9-H2) from Figure 2, (**b**) the day 7 mRNAs (Kaartinen *et al.*, 1983b), and (**c**) the day 14 mRNAs (Griffiths *et al.*, 1984) from the primary response, (**d**) the mRNAs from the secondary (Griffiths *et al.*, 1984; Berek *et al.*, 1985) response to PhOx. In (**e**)–(**i**) the sequences were all compared with each other. The sequences compared were in: (**e**) the 14 germ-line genes from Figure 2, (**f**) V_x-21 germ-line genes (Heinrich *et al.*, 1984), (**g**) the day 14 primary response mRNAs, (**h**) the secondary response mRNAs, (**i**) the V_x-Ox1 related mRNAs shown in Figure 2. (↓) refers to values being mentioned in the text.

Number of germ-line V regions which hybridize to V_x-Ox1

Cloning and sequencing experiments identified 14 different germ-line V regions (Figure 2) and mRNA sequences gave strong evidence of the existence of 4–6 additional V regions in the V_x-Ox1 family bringing the minimal number of genes up to 18–20.

To assess the likely number of V_x-Ox1 related germ-like genes we did a statistical analysis of the data based on the probabilities of finding pre-specified sequences and repeats of the same sequence in a sample of 16 clones. To carry out these calculations we assumed that all the sequences found had the same probability of detection and took into account that among the 16 different clones there were three pairs containing the same germ-line gene and that among the three pre-specified genes H3-V_x-Ox1, R1-S107B and L8, only H3 and R1 were each found once. The curve in Figure 5a shows the 'posterior probability' for each family size. The area under this curve is normalised to 1. Figure 5b corresponding to its integral shows that there is a probability of 0.5 that there are <35 and a probability of 0.8 that there are <53 V_x-Ox1 related V-regions in the BALB/c mouse germ-line.

In conclusion, the number of V_x germ-line genes capable of cross-hybridization to V_x-Ox1 under our conditions is >20 and probably <50.

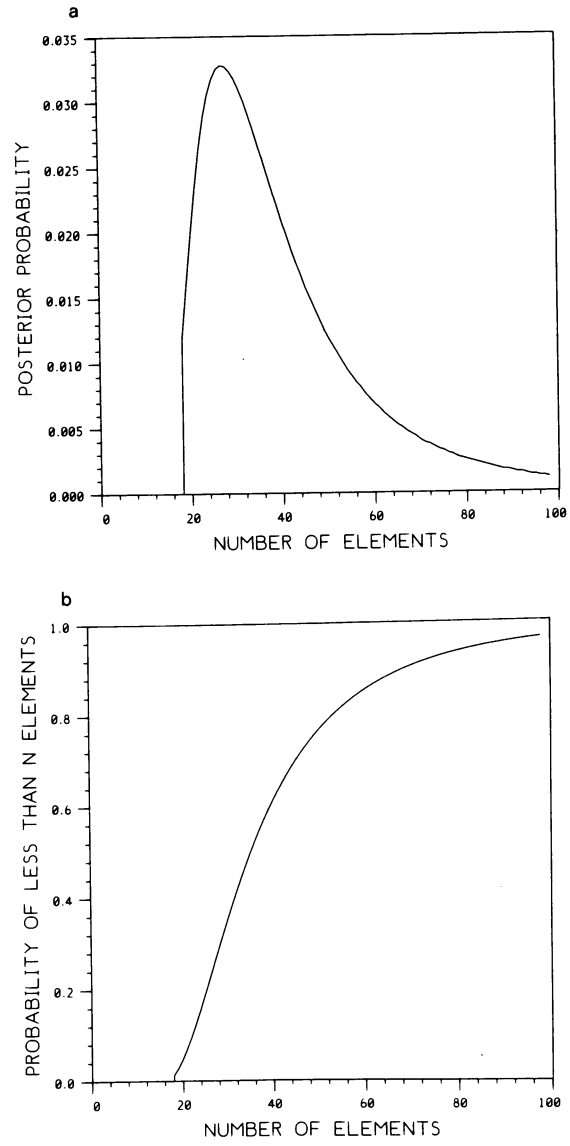

Fig. 5. Size of the V_x-Ox1 related germ-line family. (**a**) Represents the 'posterior probability' (Box and Tiao, 1983) of a given number of elements in the family considering that among 16 different clones we found 10 singletons and three pairs of identical sequences. The curve in (**a**) is further defined by the fact that three elements (H3-V_x-Ox1, R1-S107B and L8) were pre-specified and that the first two were found and the last one not found in our experiments. The area under the curve equals 1. (**b**) Is the integral of (**a**) and shows the probability that the size of the V_x-Ox2 family is less than a given number of elements. For both curves the minimum number of elements was assumed to be 18.

Discussion

The response to phOx has been shown to be dominated in the early stages by antibodies which express an identical V_x-segment. It was proposed that this represents the structure of a germ-line gene which was referred to as V_x-Ox1. The other antibodies included some closely related sequences presumably expressing alternative germ-line genes (Kaartinen *et al.*, 1983b).

As the response matured in the late primary and early secondary responses the unmutated form of V_x-Ox1 was no longer found and instead there were a number of structures closely related to V_x-Ox1 (Berek *et al.*, 1985; Griffiths *et al.*, 1984). From the pattern of the differences it was proposed that some of them were somatic mutants of V_x-Ox1 while others originated

from similar but distinct genes. In this paper we have analyzed 16 related V_χ-Ox1 germ-line genes isolated from two genomic libraries. Among them, two were the germ-line genes corresponding to mRNAs from the primary response. The most important was H3 since it was the predicted V_χ-Ox1. The other (H13) justifies our previous suggestion that the mRNA from NQ2.45.10, in spite of its similarity to V_χ-Ox1, originates from a different gene (Kaartinen *et al.*, 1983b). Therefore, other anti-oxazolone early primary antibodies seem to express an unmutated germ-line gene in addition to V_χ-Ox1. The early primary response to phOx is thus shown to be dominated by unmutated V-region genes even more than originally described (Kaartinen *et al.*, 1983b).

An interesting feature of the sequence of the germ-line gene H3 is that the last base of the V_χ segment is A followed by C as predicted (Griffiths *et al.*, 1984; Kaartinen *et al.*, 1985). Thus, the difference between the mRNA sequences at the boundary of V_χ and J_χ is due to the use of a nucleotide from either the germ-line V gene or the J_χ segment. Out of 25 informative examples of recombination between germ-line genes H3 and $J_{\chi 5}$ we found nine with the A from the V_χ-Ox1 gene and 16 with G from $J_{\chi 5}$. The recombination is clearly optimized for alternative use of either base, although a minor bias in favour of the J segment residue is likely.

To isolate all the genes which hybridize to a given probe is a daunting task as illustrated by this study. A statistical analysis of the data was in our case possible because we had analysed sufficient clones (16) to find three independent (since they originated from different libraries) identical repeats. Small differences between the mRNA of a hybridoma and a given germ-line gene can arise by somatic mutation but also from different germ-line genes. For example, the 14 germ-line genes in Figure 2 disclosed differences ranging from 6 to 46 genetic events (Figure 4e) and five germ-line genes from the V_χ21 family (Heinrich *et al.*, 1984) had a similar range (Figure 4f). This compares with differences of as many as 120 genetic events between totally unrelated genes [i.e., V_χ-Ox1 and V_χ24 gene family (Joho *et al.*, 1984)]. There have been examples of as little as three bases separating closely related germ-line V genes (Schiff *et al.*, 1985). Six differences between the germ-line genes H3 and R9 are within range of the phOx secondary response mRNAs differences (see Figure 4) but a comparison of all mRNAs with V_χ-Ox1 and R9 shows that none is likely to derive from R9. In addition to the number of base changes our study provides the all important correlation between the accumulation of mutants and the maturation of the response which starts with a major repetitive mRNA sequence originating from its predicted germ-line gene. Thus the importance of somatic mutation is emphasised more by the comparison of Figure 4a–d than by any of the individual results.

Southern blots of BALB/c mouse germ-line DNA with a V_χ-Ox1 cDNA probe reveal ~10 bands if digested with *Hind*III or with *Eco*RI (Figure 1) and 18 bands with *Xba*I (not shown). Identical patterns of bands were detected in BALB/c and A/J DNAs, which have a common ancestor but AKR mice, derived from a different ancestor (Green, 1966), share only some of the bands with BALB/c and A/J which agrees with previous studies using different V region probes (Cory *et al.*, 1981). AKR does not contain the 3.3-kb *Hind*III band corresponding to the V_χ-Ox1 germ-line gene nor does it contain the 3.6-kb *Hind*III band corresponding to four different genes (H9, H13, H4 and H1) closely related to H3 (Figure 1). R9, which is the only other isolated gene closely related to H3, is located in the 4.3-kb *Hind*III band of BALB/c. Its presence in AKR is unknown because, given the intensity of the band, it very likely contains several genes. These data suggest either that AKR does not contain V_χ-Ox1 and some closely related genes, or that a common mutation has changed their restriction sites. In either case, the expansion of the V_χ-Ox1 family seems to have taken place after the separation of the AKR and BALB/c populations.

It is striking that these closely related genes are present on fragments of identical size and with identical localisation of the V region within the fragments. Other studies have shown that adjacent sequences of related genes can be highly conserved (Cohen *et al.*, 1982; Heinrich *et al.*, 1984; Joho *et al.*, 1984). Our results suggest that those V_χ-Ox1 related germ-line genes have arisen by recent duplication of at least 3.6 or 4.0 kb long blocks of DNA not folowed by any noticeable rearrangements.

The number of bands we detect with a V_χ-Ox1 probe in *Hind*III- and *Eco*RI-digested DNA (Figure 1) are less numerous than those detected after restriction with *Xba*I (18 bands, not shown). Multiple genes were detected in some bands and are probable in others. The minimum of 20 genes in the V_χ-Ox1 family was deduced from the sequences of germ-line genes and mRNA studies. The statistical analysis of the most complete cloning and sequencing data available in this system yields a value of <35 genes with a probability of 0.5 and <53 genes with a probability of 0.8. Better accuracy requires characterization of a larger number of clones and mapping the locus containing the related genes with the help of cosmid clones as attempted for a human kappa V region family (Pech *et al.*, 1985).

The largest previously defined V region germ-line gene family is mouse heavy chain V_HII with 24 genes sequenced. There have been no repeats reported, suggesting a population size of well over 60 germ-line genes (Schiff *et al.*, 1985). However, ~50% of them are pseudogenes (Loh *et al.*, 1983; Rechavi *et al.*, 1983; Schiff *et al.*, 1985). This high value contrasts with two out of 14 (14%) reported in this study, one out of four in V_χ24 (Joho *et al.*, 1984) and none out of five in the V_χ21 family (Heinrich *et al.*, 1983). It is possible that there are less pseudogenes in the V_χ than in the V_H germ-line genes of the mouse. A more attractive alternative is that the proportion of pseudogenes increases with the size of the family. If true, this would increase selective pressure against larger gene families.

Materials and methods

DNA techniques

All DNA techniques were according to Maniatis *et al.* (1982) unless stated otherwise. Experiments with bacteriophage M13 vectors and sequencing reactions were according to Bankier and Barrell (1983).

Enzymes

Restriction enzymes were from New England Biolabs, reverse transcriptase was from Life Sciences Inc. and Klenow from Boehringer Mannheim.

cDNA cloning

Poly(A)-containing RNA from the hybridoma NQ2.48.2.2 (Kaartinen *et al.*, 1983b) was mixed with the C_χ17 kappa constant region primer (Kaartinen *et al.*, 1983b) and the first and second strands were synthesized as in Maniatis *et al.* (1982). The double-stranded DNA was then digested with *Alu*I, ligated into *Sma*I cut M13mp[9] transfected into *Escherichia coli* JM101 (Messing, 1983) and the positive clones were detected with ^{32}P-labelled cDNA synthesized like the first strand. The V_χ-Ox1 cDNA clone used here contains the 199 bases of the 3′ end of the V region and the 36 adjacent bases of the $J_{\chi 5}$ segment. As a probe we used purified nick-translated insert cut out of the vector with *Bam*HI and *Eco*RI.

DNA hybridizations

Pre-hybridizations were in 5 × Denhardt's solution at 65°C for 2 h, hybridizations were in 5 × SET, 0.5% SDS, 1 × Denhardt's and 10% dextran sulfate for 24 h at 65°C and the final wash was in 2 × SET and 0.5% SDS twice for 1 h at 65°C (20 × SET contains 3 M NaCl, 600 mM Tris-HCl at pH 8.0,

250 mM EDTA). The J region probe was a nick-translated *Hind*III-*Bgl*II fragment containing $J_{\kappa 1-5}$ (Max *et al.*, 1979).

Germ-line gene cloning and sequencing

The BALB/c germ-line DNA libraries in lambda Charon 4A were obtained from Dr. L.R.Hood (Early *et al.*, 1979) and Drs. R.Ollo and F.Rougeon (Ollo *et al.*, 1981). V_κ-Ox1 related fragments were recloned into the plasmid pAT153 (Twig and Sherratt, 1980) and then subcloned into M13mp18 and mp19 sequencing vectors (Yanisch-Perron *et al.*, 1985) as outlined in Figure 1C.

Comparisons of V region sequences

All the comparisons were done on a VAX 11/780(VMS) computer using the programs TWO and TWOB (Rodger Staden, unpublished). TWO and TWOB are computer programs for comparing exactly aligned sequence families. TWO will give counts of (i) the number of base differences; (ii) the number of genetic events required to change one sequence into another; (iii) the number of codon differences between two sequences (distinguishing between silent and expressed changes). It will also list aligned sequences and their translations. TWOB performs all the counting operations but can handle whole batches of sequences. It compares each pair of sequences in the batch using any of the above measures of similarity and then ranks them accordingly.

Statistical analysis

Reasonable belief about the value of N, which denotes the number of distinct elements in a library, can be presented by a 'posterior probability' value for each N (where 'posterior' refers to our opinion following the experiment in contrast to our 'prior' opinion before the experiment). If we represent a nominal position of prior ignorance concerning the number of elements by assigning equal prior probabilities to any value of N, then the posterior probability of N is simply proportional to the probability of observing the data obtained were N in fact true (Box and Tiao, 1983). Thus we need only to calculate the probability of the particular pattern observed, plot this against N for N greater than any known minimum, and normalise the graph to have total area of unity.

If it is assumed that in every experiment each element has a probability of $1/N$ of being sampled then the probability for each of 13 specified elements to be observed $(n_1, \ldots, n_{13})$ times in $(n_1 + \ldots + n_{13})$ samples is the multinominal probability (Feller, 1957).

$$\left(\frac{(n_1+\ldots+n_{13})!}{n_1!\,n_2!\ldots n_{13}!}\right)\left(\frac{1}{N}\right)^{n_1+\ldots+n_{13}}$$

Thus, for three pairs and 10 singletons the probability = $16!/(2!2!2!N^{16})$. To obtain the overall probability for observing the data, this figure must be multiplied by the number of ways of selecting the 13 observed elements from N. This will depend on restrictions made on pre-specified elements, which in our case is that three elements are pre-specified and only two are found. Thus the 13 observed elements have 11 unspecified members but these are now selected from only $N-3$ unspecified members in the library. Hence there are

$$\frac{N-3}{11} = (N-3)!/(N-14)!11!$$

different ways of selecting the observed elements. Multiplying by the multinomial probability shown above gives a function in N proportional to $(N-3)!/(N-14)!N^{16}$, which when normalized to have unit area provides Figure 5a.

Acknowledgements

We thank Dr. M.Neuberger for help with cDNA cloning, Dr. D.Spiegelhalter (MRC Biostatistics Unit) for performing the statistical analysis and Dr. R.Staden for writing new computer programs; Dr L.Hood and Drs. R.Ollo and F.Rougeon for the lambda libraries, and Miss J.Firth for typing the manuscript. J.Even was supported by an EMBO long-term fellowship.

References

Bankier,A.T. and Barrell,B.G. (1983) in Flavell,R.A. (ed.), *Techniques in Life Sciences: Techniques in Nucleic Acid Biochemistry* (B508), Elsevier, Ireland, pp. 1-34.

Berek,C., Griffiths,G.M. and Milstein,C. (1985) *Nature*, **316**, 412-418.

Bernard,O., Gough,N.M. and Adams,J.M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5812-5816.

Box,G.E.P. and Tiao,G.C., eds. (1973) *Bayesian Inference in Statistical Analysis*, published by Addisson and Wesley, Reading, MA.

Cohen,J.B., Effron,K., Rechavi,G., Ben-Neriah,Y., Zakut,R. and Givol,D. (1982) *Nucleic Acids Res.*, **10**, 3353-3370.

Cory,S., Tyler,B.M. and Adams,J.M. (1981) *J. Mol. Appl. Genet.*, **1**, 103-116.

Early,P.W., Davis,M.M., Kaback,D.B., Davidson,N. and Hood,L. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 857-861.

Feller,W., ed. (1957) *An Introduction to Probability Theory and its Applications, Vol. 1 (2nd Ed.)*, published by Wiley, NY.

Galfre,G. and Milstein,C. (1981) *Methods Enzymol.*, **73**, 3-64.

Green,E.L., ed. (1966) *Biology of the Laboratory Mouse*, published by McGraw Hill, NY.

Griffiths,G.M., Berek,C., Kaartinen,M. and Milstein,C. (1984) *Nature*, **312**, 271-275.

Heinrich,G., Traunecker,A. and Tonegawa,S. (1984) *J. Exp. Med.*, **159**, 417-435.

Höchtl,J., Müller,C.R. and Zachau,H.G. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1383-1387.

Joho,R., Gershenfeld,H. and Weissman,I.L. (1984) *EMBO J.*, **3**, 185-191.

Kaartinen,M., Griffiths,G.M., Hamlyn,P.H., Markham,A.F., Karjalainen,K., Pelkonen,J.L.T., Makela,O. and Milstein,C. (1983a) *J. Immunol.*, **130**, 937-945.

Kaartinen,M., Griffiths,G.M., Markham,A.F. and Milstein,C. (1983b) *Nature*, **304**, 320-324.

Kabat,E.A., Wu,T.T., Bilofsky,H., Reid-Miller,M. and Perry,H., eds. (1983) *Sequences of Proteins of Immunological Interest*, published by N.I.H. Bethesda, MD.

Kwan,S.P., Max,E.M., Seidman,J.G., Leder,P. and Scharff,M.D. (1981) *Cell*, **26**, 57-66.

Loh,D.Y., Bothwell,A.L.M., White-Scharf,M.E., Imanishi-Kari,T. and Baltimore,D. (1983) *Cell*, **33**, 85-93.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Max,E.E., Seidman,J.G. and Leder,P. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3450-3454.

Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.

Ollo,R., Auffray,C., Sikorav,J.L. and Rougeon,F. (1981) *Nucleic Acids Res.*, **9**, 4099-4109.

Pawlita,M., Potter,M. and Rudikoff,S. (1982) *J. Immunol.*, **129**, 615-618.

Pech,M., Smola,H., Pohlenz,H.D., Straubinger,B., Gerl,R. and Zachau,H.G. (1985) *J. Mol. Biol.*, **183**, 291-299.

Potter,M., Newell,J.B., Rudikoff,S. and Haber,E. (1982) *Mol. Immunol.*, **19**, 1619-1630.

Rechavi,G., Ram,D., Glazer,L., Zakut,R. and Givol,D. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 855-859.

Schiff,C., Milili,M. and Fougereau,M. (1985) *EMBO J.*, **4**, 1225-1230.

Tonegawa,S. (1983) *Nature*, **302**, 575-581.

Twigg,A.J. and Sherratt,D. (1980) *Nature*, **283**, 216-218.

Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103-119.

Received on 6 September 1985