Volume 12 Number 24 1984 Nucleic Acids Research

Immunoglobulin genes of different subgroups are interdigitated within the V_K locus

Michael Pech and Hans G.Zachau

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

Received 14 November 1984; Accepted 23 November 1984

ABSTRACT

The variable regions of immunoglobulins are encoded by multigene families which are rearranged during B-cell differentiation. These families were classified in groups and subgroups based on their amino acid sequences. Genes belonging to a distinct subgroup are believed to occur in the genome within clusters. We are investigating the organization of human variable region genes of the kappa type (V_K genes, ref. 1) in the germline and found now for the first time that V_K sequences of three of the four different subgroups are interdigitated within the V_K locus. We present evidence for the interspersion of two V_{KIII} genes and a V_{KII} pseudogene within an array of five V_{KI} genes. All eight V_K sequences are arranged in the same orientation. An evolutionary model for the generation of this 'mixed cluster' is discussed.

INTRODUCTION

While the organization of the immunoglobulin gene loci is known in outline, data on the arrangement of the gene segments encoding the variable regions of immunoglobulins are still scarce. In the case of the V_H multigene family the available information (2-4) points to a clustering of closely related genes which belong to a defined subgroup. The seven families of mouse V_H genes (5) are believed to be organized in clusters (2).

A close physical linkage was also established for a group of mouse V_K genes (6). The arrangement of the human V_K genes coding for the four subgroups (I-IV; refs. 7-9) is unknown except for a recent report from our laboratory (10) where we described an 80 kb region of the human V_K gene locus with five very similar V_{KI} sequences and a V_{KII} pseudogene. We now located two potentially functional V_{KIII} genes within this region of clustered V_{KI} sequences. The organization may be explained by an evolutionary model which includes, besides amplification

events, transposition-like processes.

MATERIALS AND METHODS

Construction and screening of cosmid libraries

The construction and screening of three cosmid libraries derived from the same genomic DNA is described in refs. 10 and 11.

DNA hybridizations

Cosmid DNA was digested with different restriction nucleases, transferred (12) to nitrocellulose filters or GeneScreenPlus membranes (New England Nuclear, Boston) and hybridized with the nick-translated DNA of M13 subclones at 68° C for 12 hs in 4xSSC, 0.1 % SDS, 1 x Denhardt's solution (13); a final washing step was at 68° C in 1xSSC, 0.1 % SDS.

Sequence analysis and computer programs

Sequences were determined by the chain termination method (14) using M13mp8 and mp9 subclones (15) and the procedures described in refs. 10 and 11. The programs of R. Staden (16) were used for assembling and editing of the sequences.

RESULTS AND DISCUSSION

Two potentially functional V_{KIII} genes are located within a cluster of V_{KI} sequences

In the V_{KI} gene cluster we had previously detected two regions which hybridized weakly with a V_{KI} probe, but it was not clear whether these regions represent remnant minigenes or possibly genes of other subgroups (10). We now mapped (Fig. 1) and sequenced (Fig. 2) the regions and established that they contain V_{KIII} genes. The assignment to subgroup III on the basis of the codons for characteristic amino acids (17) is unequivocal. DNA sequences known to be important for the V-J rearrangement (hepta- and nonanucleotide boxes (18,19)) and for correct transcription (TATA-box (20)); the pentekaideca- and decanucleotide sequences pd and dc (21) conform to the respective consensus sequences. Also the putative splice sites of the intron in the leader region are consistent with the established rules (20), at least for the Vg gene; in the Vh gene there are two possibilities for placing the intron (Fig. 2). Since both, Vg and Vh, have no features indicative of pseudogenes (cf. refs. 10,11,22,

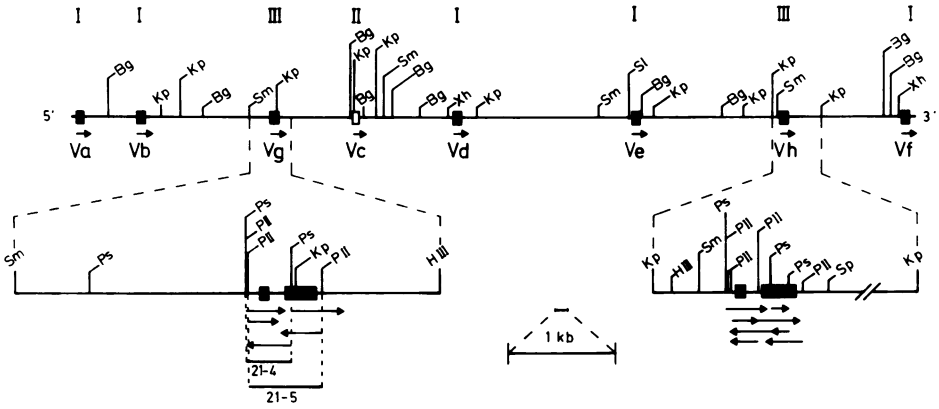

Figure 1. Map showing a section of the human V_K gene locus. The map is based on six overlapping cosmid inserts, five of which are described in Figs. 1 and 2 of ref. 10 where also the details of the cloning and mapping procedures are given. Only a selection of restriction sites is shown. The relative orientation of the two V_{KIII} genes Vg and Vh was confirmed by hybridization experiments with the M13mp8 (15) subclones 21-4 and 21-5. Subclone 21-4 seems to detect only V_{KIII} sequences when hybridized with numerous cosmid inserts (data not shown). The extent and direction of sequencing are symbolized by arrows within the extended maps of the V_{KIII} gene regions. Nucleases are abbreviated: Bg, BglII; HIII, HindIII; Kp, KpnI; PII, PvuII; Ps, PstI; Sl, SalI; Sm, SmaI; Sp, SphI; Xh, XhoI. The Roman numerals refer to subgroups I-III.

23) we consider them to be potentially functional and to contribute to the V_K repertoire. Vg and Vh are the first germline K_{III} sequences known; a cDNA clone of a rearranged K_{III} gene was reported recently (24).

The 80 kb section of the human V_K locus contains only the V_{KI}, V_{KII}, and V_{KIII} genes shown in Fig. 1. The presence of a V_{KIV} gene can be excluded on the basis of hybridization experiments with a V_{KIV} probe (H.-G. Klobeck, B. Straubinger, unpublished).

All V_K sequences within the 80 kb section are arranged in the same orientation

Inversion and deletion models for V-J joining implicate a defined orientation of genes within the V gene locus and relative to J gene segments (e.g. ref. 25). We determined the orientation of both V_{KIII} genes by sequencing across mapped restric-

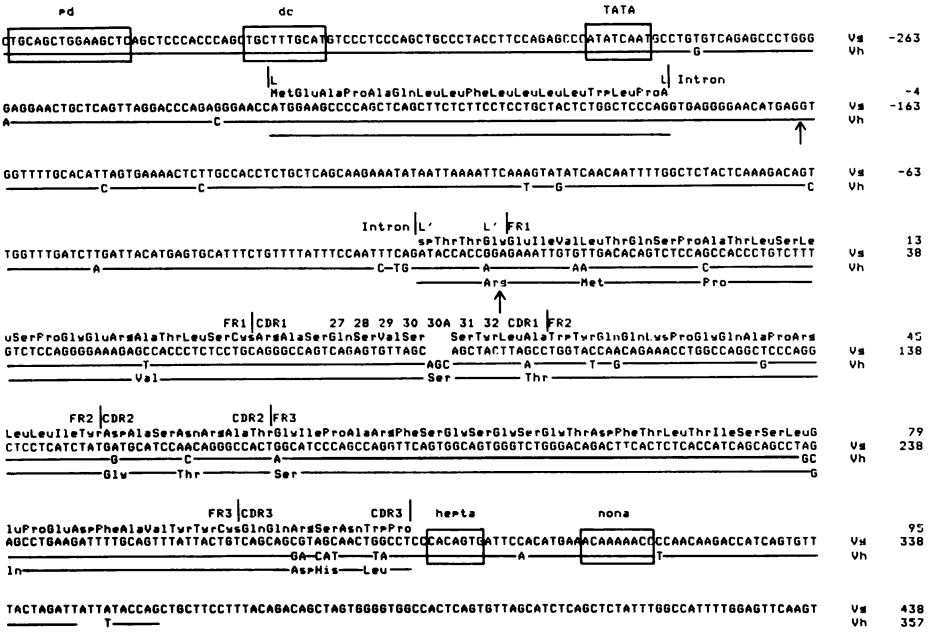

Figure 2. Sequence comparison of the two V_{KIII} genes Vg and Vh. The sequences were determined using the strategy shown in Fig. 1. The Vg and Vh subclones were derived from cos 21 and cos 103 (10), respectively. Vh is identical with Vg in its DNA sequence (first lower line) and amino acid sequence (second lower line) except for the positions where differences are indicated. The sequence of Vh extends only to position 357. Nucleotide and amino acid numbering starts with framework region (FR) 1. The latter numbering deviates from the one of E.A. Kabat et al. (17) at positions 27 to 31. The assignment of the extra codon (see text) to position 30A, and not to position 27A as in ref. 17, is based on a comparison between the DNA sequences of Vg, Vh, and NG9 (cDNA of a rearranged gene; ref. 24, and on re-examination of the published K_{KIII} protein sequences, refs. 17, 29). pd and dc (21), TATA (20), hepta- and nonanucleotide boxes (18,19) are placed according to their respective consensus sequences. The leader peptide (L and L') and the other coding regions (CDR: complementarity determining region) were identified by comparison with published amino acid sequences (17) and according to exon-intron splicing rules (20). The acceptor splice site of Vh at amino acid position -4 violates the GT-AG rule (20). An alternative donor splice site at nucleotide position -165/-164 of Vh (Fig. 2) together with an acceptor splice site at amino acid position -1 conforms to the GT-AG rule and would result in a leader protein of 22 amino acids (vertical arrows).

tion sites and confirmed the results by hybridization experiments with the subclones 21-4 and 21-5 (Fig. 1). The identical orientation of all eight V_K genes makes multiple inversion events for the rearrangement of these genes unlikely.

The two V_{KIII} genes may belong to different 'sub-subgroups'

The known amino acid sequences of mouse and human K proteins have been assigned to distinct groups and subgroups (17). While in the mouse system a large number of groups and subgroups have been defined (26,27) and the DNA hybridization data cannot always be correlated with the protein groups (4), the situation in humans appears to be rather clear: all known V_K protein and gene sequences can be classified in four 'groups' which historically have been called subgroups (7-9). Without such a comparatively simple grouping of the human K sequences our considerations of V_K gene evolution (see below) would not be possible.

The K_{III} proteins have been divided into 'sub-subgroup' IIIa and IIIb proteins on the basis of serological data and of sequence comparisons (28,29). Vg has to be assigned to IIIa because of the Ala codon at position 9. IIIb proteins have been associated with a Gly residue at this position (28,29). Although Vh has a Pro codon at position 9, it seems to be related to sub-subgroup IIIb because it has, as many of the IIIb proteins, an extra codon between positions 30 and 31 (cf. legend of Fig. 2). The difference between the Vg and Vh sequences in this region (AGCAGC, codons 30, 31 vs. AGCAGCAGC, codons 30, 30A, 31) might have originated by slippage replication (30,31).

The evolution of a 'mixed V_K gene cluster'

Studies of human (e.g. ref. 3) and mouse (e.g. ref. 4) V_H gene organization resulted in the suggestion that a V_H gene subgroup constitutes a physically linked multigene family separated from another V_H subgroup. A detailed investigation of the V_K-21 group in the mouse germline also pointed to a linkage arrangement of the very closely related V_K-21 genes.

As other multigene families the present day human V_K gene locus is likely to be the result of a series of amplification steps. Originating from a primordial V gene the ancestor genes for the four human subgroups arose; further amplification steps and divergence led to clustered genes belonging to distinct sets

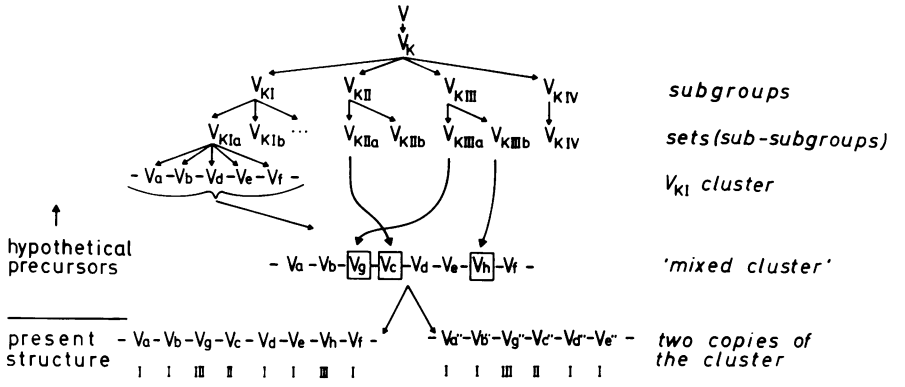

Figure 3. Model for the evolution of part of the human V_K gene locus. A cascade of amplification and transposition steps is shown which contains only the minimal number of branch points. Before a final duplication of the gene cluster transposition-like events are assumed which dispersed the V_{KIII} and V_{KII} sequences within the set of clustered V_{KI} sequences (see text). The present structure of cluster Va-Vf is shown in detail in Fig. 1.

(10,11) which may be equivalent to the serologically defined sub-subgroups. At this stage and before a final duplication of the gene cluster (ref. 10 and M. Pech et al., manuscript in preparation) we have to assume transposition-like events which dispersed the members of other subgroups within the set of closely related V_{KI} genes (Fig. 3). Whether V_K genes really behaved like transposons can, of course, not be decided on the basis of the present experimental evidence. The occurrence of isolated V_{KI} flanking sequences (32) and of imperfect inverted repeats on both sides of V_{KI} genes (10), however, supports the notion that V_K genes once had features of transposable elements. Since the eight V_K genes of the 'mixed cluster' are oriented in the same direction, one may speculate that V_K genes which became integrated in the opposite orientation diverged rather rapidly and are not detected by our probes. The remnant minigenes found within this cluster (10) may constitute such evolutionary relics. If V_K genes were transposed to positions outside the V_K locus we also would expect them to diverge rapidly.

V_K genes of different subgroups are interdigitated also in other parts of the human V_K locus, but no distinct V_{KII} or V_{KIII}

gene clusters have been found yet into which V_K genes of the other subgroups have been inserted (H.-D. Pohlenz, B. Straubinger, E. Lötscher, unpublished data). The scrambled organization described in this paper is perhaps not restricted to the human V_K gene locus but may occur also in other multigene families.

ACKNOWLEDGMENT

We thank R. Gerl for assistance and P.S. Neumaier for help in the preparation of Fig. 2. The work was supported by Bundesministerium für Forschung und Technologie and Fonds der Chemischen Industrie.

REFERENCES

1. Abbreviations used: V_H, variable region of immunoglobulin heavy chains; V_{KI}, V_{KII}, V_{KIII}, V_{KIV}, variable region of immunoglobulin light chains of the kappa type belonging to subgroups I, II, III, or IV; J, joining segment; kb, 10^3 base pairs.
2. Yancopoulos, G.D., Desiderio, S.V., Paskind, M., Kearney, J.F., Baltimore, D., and Alt, F.W. (1984) Nature 311, 727-733.
3. Rechavi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J.B., Zakut, R., and Givol, D. (1982) Proc. Natl. Acad. Sci. USA 79, 4405-4409.
4. Kemp, D.J., Tyler, B., Bernard, O., Gough, N., Gerondakis, S., Adams, J.M., and Cory, S. (1981) J. molec. appl. Genet. 1, 245-261.
5. Brodeur, P.H. and Riblet, R. (1984) Eur. J. Immunol. 14, 922-930.
6. Heinrich, G., Traunecker, A., and Tonegawa, S. (1984) J. exp. Med. 159, 417-435.
7. Milstein, C. (1967) Nature 216, 330-332.
8. Hood, L. and Talmage, D.W. (1970) Science 168, 325-334.
9. Wang, A.-C., Fudenberg, H.H., Wells, J.V., and Roelcke, D. (1973) Nature new. Biol. 243, 126-128.
10. Pech, M., Jaenichen, H.-R., Pohlenz, H.-D., Neumaier, P.S., Klobeck, H.-G., and Zachau, H.G. (1984) J. Mol. Biol. 176, 189-204.
11. Jaenichen, H.-R., Pech, M., Lindenmaier, W., Wildgruber, N., and Zachau, H.G. (1984) Nucl. Acids Res. 12, 5249-5263.
12. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
13. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.
14. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
15. Messing, J. (1983) Meth. Enzym. 101, 20-78.
16. Staden, R. (1980) Nucl. Acids Res. 8, 3673-3694.
17. Kabat, E.A., Wu, T.T., Bilofsky, H., Reid-Miller, M., and Perry, H. (1983) Sequences of Proteins of Immunological Interest, National Institutes of Health, Bethesda.

18. Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S. (1979) Nature 280, 288-294.
19. Max, E.E., Seidman, J.G., and Leder, P. (1979) Proc. Natl. Acad. Sci. USA 76, 3450-3454.
20. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
21. Falkner, F.G. and Zachau, H.G. (1984) Nature 310, 71-74.
22. Bentley, D.L. and Rabbitts, T.H. (1980) Nature 288, 730-733.
23. Cohen, J.B. and Givol, D. (1983) EMBO J. 2, 1795-1800.
24. Bentley, D.L. (1984) Nature 307, 77-80.
25. Lewis, S., Rosenberg, N., Alt, F., and Baltimore, D. (1982) Cell 30, 807-816.
26. Potter, M. (1977) Adv. Immun. 25, 141-211.
27. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., and Hood, L. (1978) Nature 276, 785-790.
28. McLaughlin, C. and Solomon, A. (1972) J. Biol. Chem. 247, 5017-5025.
29. Ledford, D.K., Goni, F., Pizzolato, M., Franklin, E.C., Solomon, A., and Frangione, B. (1983) J. Immun. 131, 1322-1325.
30. Farabaugh, P., Schmeissner, U., Hofer, M., and Miller, J.H. (1978) J. Mol. Biol. 126, 847-857.
31. Efstratiadis, A. et al. (1980) Cell 21, 653-668.
32. Straubinger, B., Pech, M., Mühlebach, K., Jaenichen, H.-R., Bauer, H.-G., and Zachau, H.G. (1984) Nucl. Acids Res. 12, 5265-5275.