The EMBO Journal vol.7 no.13 pp.4141–4150, 1988

Organization of human immunoglobulin heavy chain diversity gene loci

Y.Ichihara, H.Matsuoka[1] and Y.Kurosawa

Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake, Aichi, Japan 470-11 and [1]Department of Pediatrics, Nagoya University School of Medicine, Tsurumai, Showa-ku, Nagoya, Japan 466

Communicated by G.Köhler

The variable region of immunoglobulin heavy chain is encoded by three separate genes on the germline genome: variable (V_H), diversity (D_H) and joining (J_H) genes. Most human D_H genes are encoded in 9-kb repeating sequences. We determined the nucleotide sequence of a 15-kb DNA fragment containing more than one and a half of these repeating units, and identified 12 different D_H genes. Based on the sequence similarities of D_H coding and the surrounding regions, they can be classified into six different D_H gene families (D_{XP}, D_A, D_K, D_N, D_M and D_{LR}). Nucleotide sequences of D_H genes belonging to different families diverge greatly, while those belonging to the same families are well conserved. Since the 9-kb DNA containing the six D_H genes are multiplied at least five times, the total number of D_H genes must be ~30. These D_H genes are sandwiched by 12-nucleotide spacer signals. Most of the somatic D_H sequences found in the published $V_H-D_H-J_H$ structures (the somatic D_H segment being defined as the region which is not encoded either by germline V_H or J_H gene) were assigned to one of the germline D_H genes. Other than these typical D_H genes, however, we found a new kind of D_H gene (which we termed DIR) the spacer lengths of whose neighbouring signals were irregular. The DIR gene appears to be involved in DIR$-D_H$ or D_H-DIR joining by inversion or deletion. Two of the somatic D_H sequences were assigned to the DIR genes. Long N segments might, therefore, originate from DIR genes.

Key words: CDRIII/DIR gene/diversity gene/immunoglobulin

Introduction

The variable region of immunoglobulin (Ig) heavy (H) chain is encoded by three separate genes on the germline genome: variable (V_H), diversity (D_H) and joining (J_H) genes (for review see Tonegawa, 1983). At an early stage of B-cell differentiation, D_H-J_H joinings occur. Later on, V_H-D_H joinings complete active V_H genes (Alt *et al.*, 1984). The D_H portion in $V_H-D_H-J_H$ structure corresponds to the complementarity determining region (CDR) III of H chain (Schilling *et al.*, 1980). In mice, 12 germline D_H genes have been identified and they can be classified into three D_H gene families (D_{Q52}, D_{SP2} and D_{FL16}) based on sequence similarities (Kurosawa and Tonegawa, 1982). In human genome, Siebenlist *et al.* (1981) identified two D_H gene families (D_{HQ52} and D_{LR}). D_{LR} genes are encoded in 9-kb intervals. Recently, we identified five different D_H genes (D_{LR1}, D_{M1}, D_{N1}, D_{XP1} and $D_{XP'1}$) in one of the 9-kb units (Ichihara *et al.*, 1988) and predicted the presence of two new D_H genes. In this study, we determined the complete nucleotide sequence of 15-kb DNA fragment of D_H gene loci and found two new D_H gene families (D_A and D_K). Therefore, one repeating unit contains six different D_H gene families (D_{XP}, D_A, D_K, D_N, D_M and D_{LR}).

D_H-J_H and V_H-D_H joinings are mediated by the recombinase which recognizes a heptamer, CACTGTG or CACAGTG, and a nonamer, GGTTTTTGT or ACAAAAACC (Sakano *et al.*, 1979). The spacer lengths which separate these oligomers are also regular: 23 nucleotides for V_H and J_H genes and 12 nucleotides for D_H genes (Early *et al.*, 1980). The recombinase also recognizes the spacer lengths (Sakano *et al.*, 1981). All of the D_H genes identified so far are sandwiched by two sets of 12-nucleotide spacer signals without an exception (Kurosawa and Tonegawa, 1982; Siebenlist *et al.*, 1981; Ichihara *et al.*, 1988). In this study, however, we found a new kind of D_H gene family (which we termed DIR), the spacer lengths of whose neighbouring signals were irregular. The DIR gene appears to be involved in DIR$-D_H$ or D_H-DIR joining by inversion or deletion.

Results

Nucleotide sequence of human D_H gene loci

Germline D_H genes which belong to the same family are encoded at regular intervals. The interval is 5 kb in mouse (Kurosawa and Tonegawa, 1982) and 9 kb in man (Siebenlist *et al.*, 1981). Heteroduplex analyses of D_H-gene-containing clones indicate that the nucleotide sequence of each repeating unit is highly conserved in mouse and human (Kurosawa and Tonegawa, 1982; Siebenlist *et al.*, 1981). In a previous paper (Ichihara *et al.*, 1988), we identified five different D_H genes (D_{LR1}, D_{M1}, D_{N1}, D_{XP1} and $D_{XP'1}$) in one of the 9-kb units on human genome. To find another kind of new D_H gene, we determined the complete nucleotide sequence of a 15-kb DNA fragment containing more than one and a half of the repeating units (Figures 1 and 2). The D_H gene, by definition, is a D_H-coding region sandwiched by two sets of signal heptamers and nonamers separated by 12 nucleotide spacers. Twelve D_H genes were identified in the 15-kb DNA region (Figures 1 and 2). In the first 9-kb unit, six different D_H genes were identified: 5′$-D_{XP4}-$(1061 bp)$-D_{A4}-$(889 bp)$-$ $D_{K4}-$(1773 bp)$-D_{N4}-$(430 bp)$-$ $D_{M1}-$(2610 bp)$-D_{LR1}-$3′. Between D_{LR1} and D_{XP1} genes, characteristic 16-bp repeating sequences were found to exist. They repeated 21 times, and the consensus sequence was $\mathrm{CCTGG}^{\mathrm{G}}_{\mathrm{A}}\mathrm{C}^{\mathrm{C}}_{\mathrm{T}}\mathrm{TCACCTG}^{\mathrm{A}}_{\mathrm{G}}$. The same 16-bp sequences which repeated 17 times were also present upstream of D_{XP4}. In the second 9-kb unit, the D_{XP}-gene-containing DNA fragment is duplicated, resulting in D_{XP1} and $D_{XP'1}$.

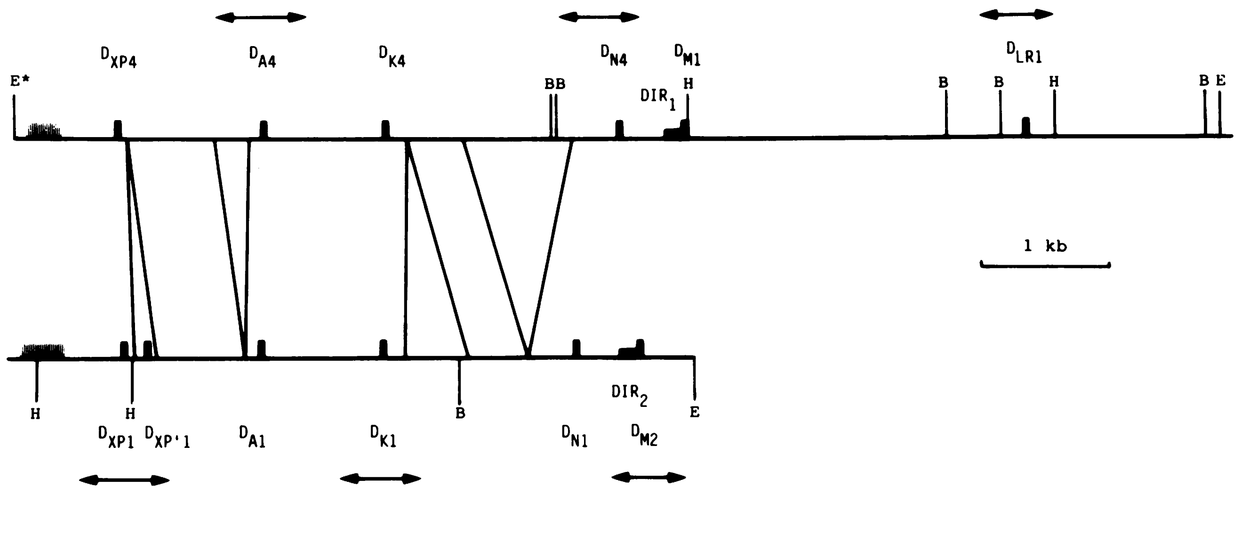


Fig. 1. Organization of 12 D_H genes and two DIR genes on a 15-kb DNA fragment. In human Ig D_H gene loci, ~9-kb DNAs were multiplied. Identical regions were aligned. Solid boxes indicate the position of D_H and DIR genes. D_{LR1} corresponds to D_1 in the paper described by Siebenlist *et al.* (1981). A D_{XP}-gene-containing fragment was duplicated in the second repeat. There are three large deleted (or inserted) portions. Hatched regions indicate 16-bp repetition. Restriction enzymes are: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII. The asterisk indicates the artificial *Eco*RI site. Six arrows indicate the regions corresponding to six D_H gene probes (D_{XP1}, D_{A4}, D_{K1}, D_{N4}, D_{M2} and D_{LR1}).

The order of D_H genes in the second 9-kb unit was the same as in the first unit: D_{XP1}−(97 bp)−$D_{XP'1}$−(804 bp)−D_{A1}−(884 bp)−D_{K1}−(1426 bp)−D_{N1}−(430 bp)−D_{M2}, and D_{LR2} should exist as reported by Siebenlist *et al.* (1981) although our sequence did not reach it. In the intervening regions, there were four large deleted or inserted DNA fragments (I, 256 bp; II, 859 bp; III, 489 bp; IV 16 bp) (Figures 1 and 2). There were slight differences, such as point mutations and deletions and/or insertions of one to three nucleotides, between the first and second 9-kb units. Other than in the neighbouring regions of typical D_H-coding regions, 15 CAC(A/T)GTG sequences exist (Figure 2). One of them located 106 bp upstream of D_M gene is sandwiched by signal nonamer sequences. The biological meaning of this structure is discussed later.

Total number of D_H genes is ~30

Siebenlist *et al.* (1981) reported that four D_{LR} genes were tandemly encoded on human genome at intervals of 9 kb. Recently, Matsuda *et al.* (1988) identified one more D_{LR} gene (D_{LR5}) in the V_H gene-clustered region. Buluwela *et al.* (1988) also characterized the organization of major and minor D_H gene clusters. The 15-kb DNA fragment analysed in this study corresponds to the 3′ side of D_{LR4} to the 5′ side of D_{LR2} in these published papers. In order to estimate the total number of D_H genes, we prepared six probes containing D_H genes belonging to different families as indicated in Figure 1. Figure 3 shows the Southern hybridization of *Bam*HI-, *Eco*RI- and *Hin*dIII-digested human placenta DNA using these probes. D_{LR} probe identified five bands at 18, 7.2, 6.5, 2.2 and 1.9 kb in the *Bam*HI digests, which was almost the same as reported previously (Siebenlist *et al.*, 1981). D_{XP} probe also identified five bands at 20, 7.2, 6.5, 4.4 and 3.7 kb in the *Bam*HI digests. Similarly, D_A, D_K and D_M probes identified five bands in one of the three different enzyme digests. This indicates that all of these five D_H gene families (D_{LR}, D_{XP}, D_A, D_K and D_M) consist of five members. On the other hand, D_N probe identified only three bands in all of the enzyme digests. Since the nucleotide sequence of the 9-kb repeats are very similar among them, one band may contain more than two D_H genes. In addition, large DNA fragments can also contain more than two D_H genes. Although the size of each probe is ~500 nucleotides long, a probe may identify two bands for one D_H gene because of the presence of a restriction site in the region covered by the probe. The estimation of number of D_H genes by Southern hybridization may be disturbed by these factors. However, since each probe identified three to five bands in the three enzyme digests except for D_N probe, it is quite likely that five 9-kb repeats contain six D_H gene families and that the total number of D_H genes is ~30.

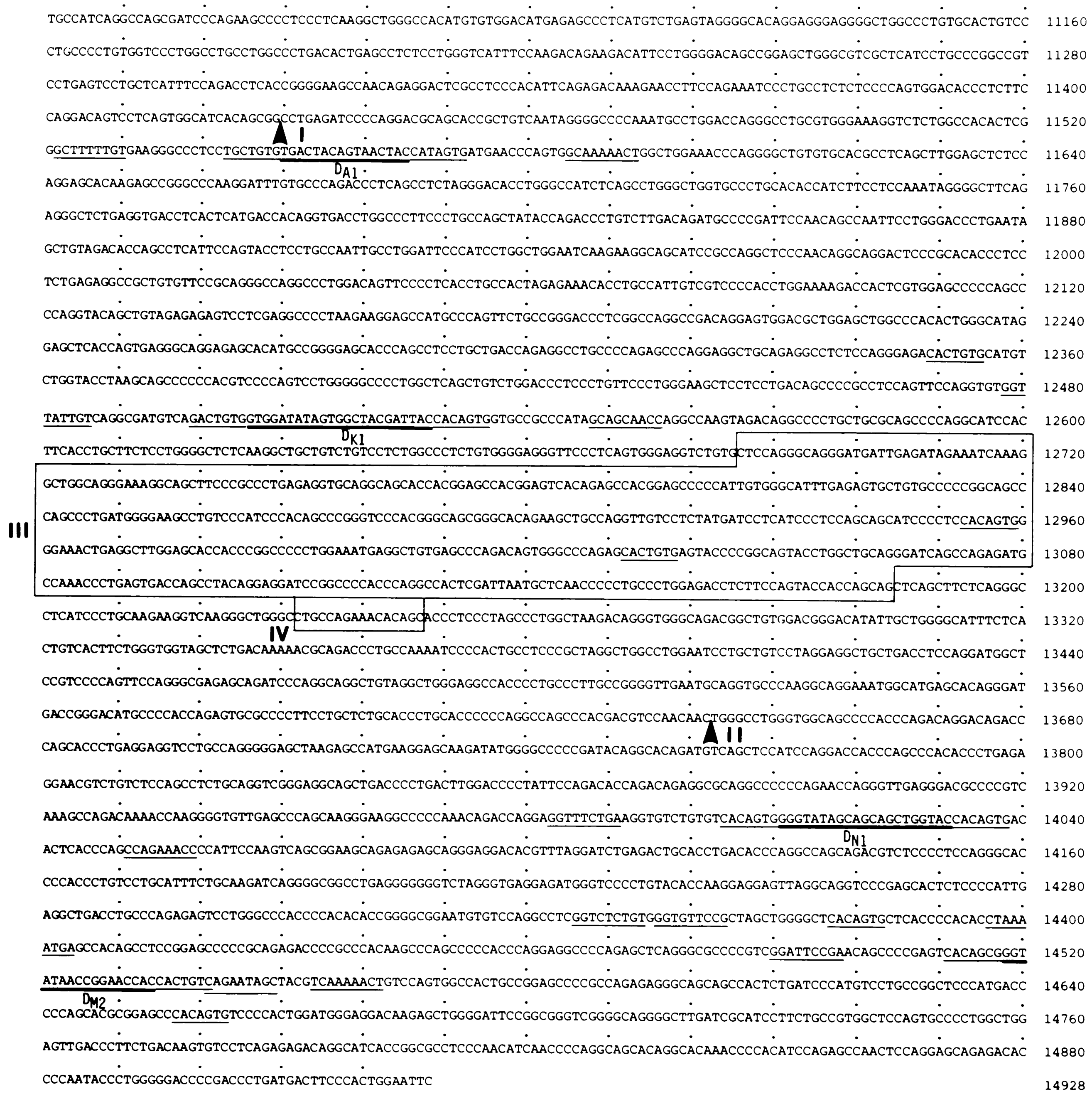
Discussion

Nucleotide sequences of D_H genes belonging to different families diverge greatly, and those belonging to the same families are well conserved

The nucleotide sequences of 17 different D_H-coding regions and their neighbouring signals are summarized in Figure 4. They can be classified into seven families. All the D_H genes are sandwiched by two sets of 12-nucleotide spacer signals. Signal heptamers are well conserved except that, in the 5′ side of D_{M2}, CACAGCG is found instead of CACAGTG and, in the 5′ side of D_{A1}, TGCTGTG is found. On the other hand, the signal nonamers diverge relatively from the consensus nonamer sequence, GGTTTTTGT or ACAAAAACC. This phenomenon is commonly observed in other signal regions of Ig- and T-cell receptor genes (for review see Akira *et al.*, 1987). In mouse, 12 D_H genes are classified into three families (Kurosawa and Tonegawa, 1982). One of the families, D_{Q52} in mouse, located close to the J_H gene cluster, is similar to D_{HQ52} in man (Sakano *et al.*, 1981; Ravetch *et al.*, 1981). Nucleotide sequences of D_H genes belonging to D_{SP2} are highly conserved (Kurosawa and Tonegawa, 1982). Since D_{FL16}-gene-

```
   GAATTCATGCCGCCATCTGGCAGGCACAGAGCATGGGCTGGGAGGAGGGGCAGGGACACCAGGCAGGTTGGCACCAACTGAAAATTACAGAAGTCTCATACATCTACCTCAGCCTTGCCT   120
   GACCTGGGCCTCACCTGACCTGGACCTCACCTGGCCTGGACCTCACCTGGCCTAGACCTCACCTCTGGGCTTCACCTGAGCTCGGCCTCACCTGACTTGGACCTTGCCTGTCCTGAGCTC   240
   ACATGATCTGGGCCTCACCTGACCTGGTTTCACCTGACCTGGGCTTCACCTGACCTGGGCCTCATCTGACCTGGGCCTCACTGGCCTGGACCTCACCTGGCCTGGGCTTCACCTGGCCTC   360
   AGGCCTCATCTGCACCTGCTCCAGGTCTTGCTGGAACCTCAGTAGCACTGAGGCTGCAGGGGCTCATCCAGGGTTGCAGAATGACTCTAGAACCTCCCACATCTCAGCTTTCTGGGTGGA   480
   GGCACTGGTGGCCCAGGGAATATAAAAAGCCTGAATGATGCCTGCGTGATTTGGGGGCAATTTATAAACCCAAAATGGACATGGCCATGCAGCGGGTAGGGACAATACAGACAGATATCA   600
   GCCTGAAATGGAGCCTCAGGGCACAGTGGGCACGGACACTGTCCACCTAAGCCAGGGGCAGACCCGAGTGTCCCCGCAGTAGACCTGAGAGCGCTGGGCCCACAGCCTCCCCTCGGTGCC   720
   CTGCTACCTCCTCAGGTCAGCCCTGGACATCCCGGGTTTCCCCAGGCTGGCGGTAGGTTTGGGGTGAGGTCTGTGTCACTGTGGTATTACGATTTTTGGAGTGGTTATTATACCCACAGT   840
                                                                                                             DXP4
   GTCACAGAGTCCATCAAAAACCCATCCCTGGGAACCTTCTGCCACAGCCCTCCTGTGGGGCACCGCCGCGTGCCATGTTAGGATTTTGACTGAGGACACAGCACCATGGGTATGGTGCTA   960
   CCGCAGCAGTGCAGCCTGTGACCCAAACACACAGGGCAGCAGGCACAACAGACAAGCCCACAAGTGACCACCCTGAGCTCCTGCCTGCCAGCCCTGGAGACCATGAAACAGATGGCCAGG  1080
   ATTATCCCATAGGTCAGCCAGACCTCAGTCCAACAGGTCTGCATCGCTGCTGCCCTCCAATACCAGTCCGGATGGGGACAGGGCCGGCCCACATTACCATTTGCTGCCATCCGGCCAACA  1200
   GTCCCAGAAGCCCCTCCCTCAAGGCTGGGCCACATGTGTGGACCCTGAGAGCCCCCCATGTCTGAGTAGGGGCACCAGGAAGTGGGCTGGCCCTGTGCACTGTCACTGCCCCTGTGGTCC  1320
   CTGGCCTGCCTGGCCCTGACACCTGGGCCTCTCCTGGGTCATTTCCAAGACAGAAGACATTCCCAGGACAGCTGGAGCTGGGAGTCCATCATCCTGCCTGGCCATCCTGAGTCCTGCGCC  1440
   TTTCCAAACCTCACCCGGGAAGCCAACAGAGGAATCACCTCCCACAGGCAGAGACAAAGACCTTCCAGAAATCTCTGTCTCTCTCCCCAGTGGGCACCCTCTTCCAGGGCAGTCCTCAGT  1560
   GATATCACAGTGGGAACCCACATCTGGATCGGGACTGCCCCCAGAACACAAGATGGCCCACAGGGACAGCCCCACAGCCCAGCCCTTCCCAGACCCCTAAAAGGCGTCCCACCCCCTGCA  1680
I  TCTGCCCCAGGGCTCAAACTCCAGGAGGACTGACTCCTGCACACCCTCCTGCCAGACATCACCTCAGCCCCTCCTGGAAGGGACAGGAGCGCGCAAGGGTGAGTCAGACCCTCCTGCCCT  1800
   CGATGGCAGGCGGAGAAGATTCAGAAAGGTCTGAGATCCCCAGGACGCAGCACCACTGTCAATGGGGGCCCCAGACGCCTGGACCAGGGCCTGCGTGGGAAAGGCCTCTGGGCACACTCA  1920
   GGGGCTTTTTGTGAAGGGTCCTCCTACTGTGTGACTACAGTAACTACCACAGTGATGAACCCAGCAGCAAAAACTGACCGGACTCCCAAGGTTTATGCACACTTCTCGCTCAGAGCTCTC  2040
                                                    DA4
   CAGGATCAGAAGAGCCGGGCCCAAGGGTTTCTGCCCAGACCCTCGGCCTCTAGGGACATCTTGGCCATGACAGCCCATGGGCTGTGCCCCACACATCGTCTGCCTTCAAACAAGGGCTTC  2160
   AGAGGGCTCTGAGGTGACCTCACTGATGACCACAGGTGCCCTGGCCCCTTCCCCGCCAGCTGCACCAGACCCCGTCCTGACAGATGCCCCGATTCCAACAGCCAATTCCTGGGGCCAGGA  2280
   ATCGCTGTAGACACCAGCCTCCTTCCAACACCTCTTGCCAATTGCCTGGATTCCCATCCCGGTTGGAATCAAGAGGACAGCATCCCCCAGGCTCCCAACAGGCAGGACTCCCACACCCTC  2400
   CTCTGAGAGGCCGCTGTGTTCCGTAGGGCCAGGCTGCAGACAGTCCCCCTCACCTGCCACTAGACAAATGCCTGCTGTAGATGTCCCCACCTGGAAAAGACCACTCATGGAGCCCCCAGC  2520
   CCCAGGTACAGCCATAGAGAGAGTCTCTGAGGCCCCTAAGAAGTAGCCATGCCCAGTTCTGCCGGGACCCTCGGCCAGGCTGACAGGAGTGGACGCTGGAGCTGGGCCCACACTGGGCCA  2640
   CATAGGAGCTCACCAGTGAGGGCAGGAGAGCACATGCCGGGGAGCACCCAGCCTCCTGCTGACCAGAGGCCCGTCCCAGAGCCCAGGAGGCTGCAGAGGCCTCTCCAGGGGGACACTGTG  2760
   CATGTCTGGTCCCTGAGCAGCCCCCCATGTCCCCAGTCCTGGGGGGCCCCCTGGCACAGCTGTCTGGACCCTCTCTATTCCCTGGGAAGCTCCTCCTGACAGCCCCGCCTCCAGTTCCAG  2880
   GTGTGGTTATTGTCAGGGGGTGTCAGACTGTGGTGGATACAGCTATGGTTACCACAGTGGTGCTGCCCATAGCAGCAACCAGGCCAAGTAGACAGGCCCCTGTGCGCAGCCCCAGGCATC  3000
                                                    DK4
   CACTTCACCTGCTTCTCCTGGGGCTCTCAAGGCTGCTGTTTTCTGCACTCTCCCCTCTGTGGGGAGGGTTCCCTCAGTGGGAGGATCTGTTCTCAACATCCCAGGGCCTCATTCCTGCAA  3120
                                                                                                           ▲ III
   GGAAGGCCAATGATGGGCAACCTCACATGCCGCGGCTAAGATAGGGTGGGCAGCTGGCGGGGACAGGACATCCTGCTGGGGTATCTGTCACTGTGCCTAGTGGGGCACTGGCTCCCAAAC  3240
                     ▲ IV
   AACGCAGTCCTCGCCAAAATCCCCACGGCCTCCCCCGCTAGGGGCTGGCCTGATCTCCTGCAGTCCTAGGAGGCTGCTGACCTCCAGAATGGCTCCGTCCCCAGTTCCAGGGCGAGAGCA  3360
   GATCCCAGGCCGGCTGCAGACTGGGAGGCCACCCCCTCCTTCCCAGGGTTCACTGCAGGTGACCAGGGCAGGAAATGGCCTGAACACAGGGATAACCGGGCCATCCCCCAACAGAGTCCA  3480
   CCCCCTCCTGCTCTGTACCCCGCACCCCCAAGGCCAGCCCATGACATCCGACAACCCCACACCAGAGTCACTGCCCGGTGCTGCCCTAGGGAGGACCCCTCAGCCCCCACCCTGTCTAGA  3600
   GGACTGGGGAGGACAGGACACGCCCTCTCCTTATGGTTCCCCCACCTGGCTCTGGCTGGGACCCTTGGGGTGTGGACAGAAAGGACGCTTGCCTGATTGGCCCCCAGGAGCCCAGAACTT  3720
   CTCTCCAGGGACCCCAGCCCGAGCACCCCCTTACCCAGGACCCAGCCCTGCCCCTCCTCCCATCTGCTCTCCTCTCATCACCCCATGGGAATCCAGAATCCCCAGGAAGCCATCAGGAAG  3840
   GGCTGAGGGAGGAAGTGGGGCCACGTGCACCACCAGGCAGGAGGCTCCGTCTTTGTGAACCCAGGGAGGTGCCAGCCTCCTAGAGGGTATGGTCCACCCTGCCTATGGCTCCCACAGTGG  3960
II CAGGCTGCAGGGAAGGACCAGGGACGGTGTGGGGGAGGGCTCAGGGCCGCGCGGGTGCTCCATCTTGGATGAGCCCATCTCTCTCACCCACGGACTCACCCACCTCCTCTCCACCCTGGT  4080
   CACACGTCGTCCACACCATCCTAAGTCCCACCTACACCAGAGCCGGCACAGCCAGTGCAGACAGAGGCTGGGGTGCAGGGGGGCCGCCAGGGCAGCTTTGGGGAGGGAGGAATGGAGGAA  4200
   GGGGAGTTCAGTGAAGAGGCCCCCCTCCCCTGGGTCCAGGATCCTCCTCTGGGACCCCCGGATCCCATCCCCTCCAGGCTCTGGGAGGAGAAGCAGGATGGGAGAATCTGTGCGGGACCC  4320
   TCTCACAGTGGAATACCTCCACAGCGGCTCAGGCAAGACCCAAAAGCCCCTCAGTGAGCCCTCCACTGCAGTCCTGGGCCTGGGTAGCAGCCCCTCCCACAGAGGATGAACCCAGCACCC  4440
   CGAGGATGTCCTGCCAGGGGGAGCTCAGAGCCATGAAGGAGCAGGATATGGGACCCCCGATACAGGCACAGACCTCAGCTCCATTCAGGACTGCCACGTCCTGCCCTGGGAGGAACCCCT  4560
   TTCTCTAGTCCCTGCAGGCCAGGAGGCAGCTGACTCCTGACTTGGACGCCTATTCCAGACACCAGACAGAGGGGCAGGCCCCCCAGAACCAGGGATGAGGACGCCCCGTCAAGGCCAGAA  4680
   AAGACCAAGTTGTGCTGAGCCCAGCAAGGGAAGGTCCCCAAACAAACCAGGAACGTTTCTGAAGGTGTCTGTGTCACAGTGGAGTATAGCAGCTCGTCCCACAGTGACACTCGCCAGGCC  4800
                                                                                          DN4
   AGAAACCCCATCCCAAGTCAGCGGAATGCAGAGAGAGCAGGGAGGACATGTTTAGGATCTGAGGCCGCACCTGACACCCAGGCCAGCAGACGTCTCCTGTCCATGGCACCCTGCCATGTC  4920
   CTGCATTTCTGGAAGAACAAGGGCAGGCTGAAGGGGGTCCAGGACCAGGAGATGGGTCCCCTCTACCCAGAGAAGGAGCCAGGCAGGACACAAGCCCCCTCCCCATTGAGGCTGACCTGC  5040
   CCAGAGGGTCCTGGCCCACCCCACACACCGGGGCGGAATGTGTGCAGGCCTCGGTCTCTGTGGGTGTTCCGCTAGCTGGGGCTCACAGTGCTCACCCCACACCTAAAACGAGCCACAGC  5160
   CTCAGAGCCCCTGAAGGAGACCCCGCCCACAAGCCCAGCCCCCACCCAGGAGGCCCCAGAGCACAGGGCGCCCCGTCGGATTCTGAACAGCCCCGAGTCACAGTGGGTATAACTGGAACT  5280
                                                                                                                 DM1
   ACCACTGTGAGAAAAGCTTCGTCCAAAACGGTCTCCTGGCCACAGTCGGAGGCCCCGCCAGAGAGGGGAGCAGCCACCCCAAACCCATGTTCTGCCGGCTCCCATGACCCCGTGCACCTG  5400
   GAGCCCCACAGTGTCCCCACTGGATGGGAGGACAAGGGCCGGGGGCTCCGGCGGGTCGGGGCAGGGGCTTGATGGCTTCCTTCTGCCGTGGCTCCAGTGCCCCTGGCTGGAGTTGACCCT  5520
```

```
TCTGACAAGTGTCCTCAGAGAGTCAGGGATCAGTGGCACCTCCCAACATCAACCCCACGCAGCCCAGGCACAAACCCCACATCCAGGGCCAACTCCAGGAACAGAGACACCCCAATACCC  5640
TGGGGGACCCCAACCCTGATGACTCCCGTCCCATCTCTGTCCCTCACTTGGGGCCTGCTGCGGGGCGAGCACTTGGAGCAAACTCAGGCTTAGGGGACACCACTGTGGGCCTGACCTCG  5760
AGCAGGCCACAGACCCTTCCCTCCTGCCCTGGTGCAGCACAGACTTTGGGGTCTGGGCAGGGAGGAACTTCTGGCAGGTCACCAAGCACAGAGCCCCCAGGCTGAGGTGGCCCCAGGGGG  5880
AACCCCAGCAGGTGGCCCACTACCCTTCCTCCCAGCTGGACCCCATGTCTTCCCCAAGATAGGGGTGCCATCCAAGGCAGGTCCTCCATGGAGCCCCCTTCAGGCTCCTCTCCAGACCCC  6000
ACTGGGCCTCAGTCCCCACTCTAGGAATGCAGCCACCACGGGCACACCAGGCAGCCCAGGCCCAGCCACCCTGCAGTGCCCAAGCCCACACCCTGGAGGAGAGCAGGGTGCGTCTGGGAG  6120
GGGCTGGGCTCCCCACCCCCACCCCCACCTGCACACCCCACCCACCCTTGCCCGGGCCCCCTGCAGGAGGGTCAGAGCCCCCATGGGATATGGACTTAGGGTCTCACTCACGCACCTCCC  6240
CTCCTGGGAGAAGGGGTCTCATGCCCAGATCCCCCCAGCAGCGCTGGTCACAGGTAGAGGCAGTGGCCCCAGGGCCACCCTGACCTGGCCCCTCAGGCATCCTCTAGCCCTGGCTGCCCT  6360
GCTGTCCCTGGGAGGCCTGGGCTCCACCAGACCACAGGTCTAGGGCACCGCCCACACTGGGGCCGCCCACACACAGCTCACAGGAAGAAGATAAGCTCCAGACCCCCAGGCCCGGGACCT  6480
GCCTTGCTGCTACGACTTCCTGCCCCAGACCTCGTTGCCCTCCCCCGTCCACTTACACACAGGCCAGGAAGACTGTTCCCACACAGACCAACCCCAGACGGGGACCACCTGGCACTCAGG  6600
TCACTGCCATTTCCTTCTCCATTCACTTCCAATGCCTCTGTGCTTCCTCCTCCTCCTTCCTTCGGGGGAGCACCCTGTGCAGCTCCTCCCTGCAGTCCACACCCTGGGGAGACCCGACCC  6720
TGCAGCCCACACCCTGGGGAGACCTGACCCTCCTCCAGCCCTTTCTCCCCCGCTGCTCTTGCCACCCACCAAGACAGCCCTGGGGTCCTGTCCCTACAGCCCCCACCCAGTTCTCTACCT  6840
AGACCCGTCTTCCTCCCTCTAAACACCTCTCCCAGGCCAACCCTACACCTGCAGGCCCTCCCCTCCACTGCCAAAGACCCTCAGTTTCTCCTGCCTGTGCCCACCCCCGTGCTCCTCCTG  6960
CCCACAGCTCGAGCTCTTCCTCTCCTAGGGCCCCTGAGGGATGGCATTGACCGTGCCCTCGCACCCACACACTGCCCATGCCCTCACATTCCTCCTGGCCACTCCAGCCCCACTCCCCTC  7080
TCAGGCCTGGCTCTGGTATTTCTGGGACAAAGCCTTACCCAAGTCTTTCCCATGCAGGCCTGGGCCCTTACCCTCACTGCCCGGTTACAGGGCAGCCTCCTGTGCACAGAAGCAGGGAGC  7200
TCAGCCCTTCCACAGGCAGAAGGCACTGAAAGAAATCGGCCTCCAGCGCCTTGACACACGTCTGCCTGTGTCTCTCACTGCCCGCACCTGCAGGGAGGCTCGGCACTCCCTCTAAAGACG  7320
AGGGATCCAGGCAGCAGCATCACAGGAGAATGCAGGGCTACCAGACATCCCAGTCCTCTCACAGGCCTCTCCTGGGAAGAGACCTGAAGACGCCCAGTCAACGGAGTCTAACACCAAACC  7440
TCCCTGGAGGCCGATGGGTAGTAACGGAGTCATTGCCAGACCTGGAGGCAGGGGAGCAGTGAGCCCGAGCCCACACCATAGGGCCAGAGGACAGCCACTGACATCCCAAGCCACTCACTG  7560
GTGGTCCCACAACACCCCATGGAAAGAGGACAGACCCACAGTCCCACCTGGACCAGGGCAGAGACTGCTGAGACCCAGCACCAGAACCAACCAAGAAACACCAGGCAACAGCATCAGAGG  7680
GGGCTCTGGCAGAACAGAGGAGGGGAGGTCTCCTTCACCAGCAGGCGCTTCCCTTGACCGAAGACAGGATCCATGCAACTCCCCCAGGACAAAGGAGGAGCCCCTTGTTCAGCACTGGGC  7800
TCAGAGTCCTCTCCAAGACACCCAGAGTTTCAGACAAAAACCCCCTGGAATGCACAGTCTCAGCAGGAGAGCCAGCCAGAGCCAGCAAGATGGGGCTCAGTGACACCCGCAGGGACAGGA  7920
GGATTTTGTGGGGGCTCGTGTCACTGTGAGGATATTGTACTAATGGTGTATGCTATACCCACAGTGACACAGCCCCATTCCCAAAGCCCTACTGCAAACGCATTCCACTTCTGGGGCTGA  8040
                                            DLR1
GGGGCTGGGGGAGCGTCTGGGAAATAGGGCTCAGGGGTGTCCATCAATGCCCAAAACGCACCAGACTCCCCTCCATACATCACACCCACCAGCCAGCGAGCAGAGTAAACAGAAAATGAG  8160
AAGCAAGCTGGGGAAGCTTGCACAGGCCCCAAGGAAAGAGCTTTGGCGGGTGTGTAAGAGGGGATGCGGGCAGAGCCTGAGCAGGGCCTTTTGCTGTTTCTGCTTTCCTGTGCAGAGAGT  8280
TCCATAAACTGGTGTTCGAGATCAATGGCTGGGAGTGAGCCCAGGAGGACAGCGTGGGAAGAGCACAGGGAAGGAGGAGCAGCCGCTATCCTACACTGTCATCTTTCGAAAGTTTGCCTT  8400
GTGCCCACACTGCTGCATCATGGGATGCTTAACAGCTGATGTAGACACAGCTAAAGAGAGAATCAGTGAGATGGATTTGCAGCACAGATCTGAATAAATTCTCCAGAATGTGGAGCAGCA  8520
CAGAAGCAAGCACACAGAAAGTGCCTGATGCAAGGACAAAGTTCAGTGGGCACCTTCAGGCATTGCTGCTGGGCACAGACACTCTGAAAAGCCCTGGCAGGAACTCCCTGTGACAAAGCA  8640
GAACCCTCAGGCAATGCCAGCCCCAGAGCCCTCCCTGAGAGCCTCATGGGCAAAGATGTGCACAACAGGTGTTTCTCATAGCCCCAAACTGAGAGCAAAGCAAACGTCCATCTGAAGGAG  8760
AACAGGCAAATAAACGATGGCAGGTTCATGAAATGCAAACCCAGACAGCCACAAGCACAAAAGTACAGGGTTATAAGCGACTCTGGTTGAGTTCATGACAATGCTGAGTAATTGGAGTAA  8880
CAAAGTAAACTCCAAAAAATACTTTCAATGTGATTTCTTCTAAATAAAATTTACACCCTGCAAAATGAACTGTCTTCTTAAGGGATACATTTCCCAGTTAGAAAACCATAAAGAAAACCA  9000
AGAAAAGGATGATCACATAAACACAGTGGTGGTTACTTCTGCTGGGGAAGGAAGAGGGTATGAACTGAGATACACAGGGTGGGCAAGTCTCCTAACAAGAACAGAACGAATACATTACAG  9120
TACCTTGAAAACAGCAGTTAAACTTCTAAATTGCAAGAAGAGGAAAATGGACACAGTTGTGTTTAGAAAATTCTCAGTCCAGCACTGTTCATAATAGCAAAGACATTAACCCAGGTCGGA  9240
TAAATAAGCGATGACACAGGCAATTGCACAATGATACAGACATATATTTAGTATATGAGACATCGATGATGTATCCCCAAATAAACGACTTTAAAGAGATAAAGGGCTGATGTGTGGTGG  9360
CATTCACCTCCCTGGGATCCCCGGACAGGTTGCAGGCTCACTGTGCAGCAGGGCAGGCGGGTACCTGCTGGCAGTTCCTGGGGCCTGATGTGGAGCAAGCGCAGGGCCATATATCCCGGA  9480
GACGGCACAGTCAGTGAATTCCAGAGAGAAGCAACTCAGCCACACTCCCCAGGCAGAGCCCGAGAGGGACGCCCACGCACAGGGAGGCAGAGCCCAGCACCTCCGCAGCCAGCACCACCT  9600
GTGCACGGGCCACCACCTTGCAGGCACAGAGTGGGTGCTGAGAGGAGGGGCAGGGACACCAGGCAGGTGAGCACCCAGAGAAAACTGCAGACGCCTCACACATCCACCTCAGCCTCCCCT  9720
GACCTGGACCTCACTGCCTGGGCCTCACTTAACCTGGGCTTCACCTGACCTTGGCCTCACCTGACTTGGACCTGCCTGTCCCAAGCTTTACCTGACCTGGGCCTCAACTCACCTGAACGT  9840
CTCCTGACCTGGGTTTAACCTGTCCTGGAACTCACCTGGCCTTGGCTTCCCCTGACCTGGACCTCATCTGGCCTGGGCTTCACCTGGCCTGGGCCTCACCTGACCTGGACCTCATCTGGC  9960
CTGGACCTCACCTGGCCTGGACTTCACCTGGCCTGGGCTTCACCTGACCTGGACCTCACCTGGCCTCGGCCTCACCTGCACCTGCTCCAGGTCTTGCTGGAGCCTGAGTAGCACTGAGG  10080
GTGCAGAAGCTCATCCAGGGTTGGGGAATGACTCTAGAAGTCTCCCACATCTGACCTTTCTGGGTGGAGGCAGCTGGTGGCCCTGGGAATATAAAAATCTCCAGAATGATGACTCTGTGA  10200
TTTGTGGGCAACTTATGAACCCGAAAGGACATGGCCATGGGGTGGGTAGGGACATAGGGACAGATGCCAGCCTGAGGTGGAGCCTCAGGACACAGGTGGGCACGGACACTATCCACATAA  10320
GCGAGGGATAGACCCGAGTGTCCCCACAGCAGACCTGAGAGCGCTGGGCCCACAGCCTCCCCTCAGAGCCCTGCTGCCTCCTCCGGTCAGCCCTGGACATCCCAGGTTTCCCCAGGCCTG  10440
CCGGTAGGTTTAGAATGAGGTCTGTGTCACTGTGGTATTACGATATTTTGACTGGTTATTATAACCACAGTGTCACAGAGTCCATCAAAAACCCATGCCTGGAAGCTTCCCGCCACAGCC  10560
                                                   DXP1
CTCCCCATGGGGCCCTGCTGCCTCCTCAGGTCAGCCCCGGACATCCCGGGTTTCCCCAGGCTGGGCGGTAGGTTTGGGGTGAGGTCTGTGTCACTGTGGTATTACTATGGTTCGGGGAGT  10680
                                                                                                                 DXP'1
TATTATAACCACAGTGTCACAGAGTCCATCAAAAACCCATCCCTGGGAGCCTCCCGCCACAGCCCTCCCTGCAGGGACCGTACGTGCCATGTTAGGATTTTGATCGAGGAGACAGCACCA  10800
TGGGTATGGTGGCTACCACAGCAGTGCAGCCTGTGACCCAAACCCACAGGGCAGCAGGCACGATGGACAGGCCCGTGACTGACCACGCTGGGCTCCAGCCTGCCAGCCCTGGAGATCATG  10920
AAACAGATGGCCAAGGTCACCCTACAGGTCATCCAGATCTGGCTCCGAGGGGTCTGCATCGCTGCTGCCCTCCCAACGCCAGTCCAAATGGGACAGGGACGGCCTCACAGCACCATCTGC  11040
```



```
TGCCATCAGGCCAGCGATCCCAGAAGCCCCTCCCTCAAGGCTGGGCCACATGTGTGGACATGAGAGCCCTCATGTCTGAGTAGGGGCACAGGAGGGAGGGGCTGGCCCTGTGCACTGTCC  11160
CTGCCCCTGTGGTCCCTGGCCTGCCTGGCCCTGACACTGAGCCTCTCCTGGGTCATTTCCAAGACAGAAGACATTCCTGGGGACAGCCGGAGCTGGGCGTCGCTCATCCTGCCCGGCCGT  11280
CCTGAGTCCTGCTCATTTCCAGACCTCACCGGGGAAGCCAACAGAGGACTCGCCTCCCACATTCAGAGACAAAGAACCTTCCAGAAATCCCTGCCTCTCTCCCCAGTGGACACCCTCTTC  11400
CAGGACAGTCCTCAGTGGCATCACAGCGGCCTGAGATCCCCAGGACGCAGCACCGCTGTCAATAGGGGCCCCAAATGCCTGGACCAGGGCCTGCGTGGGAAAGGTCTCTGGCCACACTCG  11520
GGCTTTTTGTGAAGGGCCCTCCTGCTGTGTGACTACAGTAACTACCATAGTGATGAACCCAGTGGCAAAAACTGGCTGGAAACCCAGGGGCTGTGTGCACGCCTCAGCTTGGAGCTCTCC  11640
AGGAGCACAAGAGCCGGGCCCAAGGATTTGTGCCCAGACCCTCAGCCTCTAGGGACACCTGGGCCATCTCAGCCTGGGCTGGTGCCCTGCACACCATCTTCCTCCAAATAGGGGCTTCAG  11760
AGGGCTCTGAGGTGACCTCACTCATGACCACAGGTGACCTGGCCCTTCCCTGCCAGCTATACCAGACCCTGTCTTGACAGATGCCCCGATTCCAACAGCCAATTCCTGGGACCCTGAATA  11880
GCTGTAGACACCAGCCTCATTCCAGTACCTCCTGCCAATTGCCTGGATTCCCATCCTGGCTGGAATCAAGAAGGCAGCATCCGCCAGGCTCCCAACAGGCAGGACTCCCGCACACCCTCC  12000
TCTGAGAGGCCGCTGTGTTCCGCAGGGCCAGGCCCTGGACAGTTCCCCTCACCTGCCACTAGAGAAACACCTGCCATTGTCGTCCCCACCTGGAAAAGACCACTCGTGGAGCCCCCAGCC  12120
CCAGGTACAGCTGTAGAGAGAGTCCTCGAGGCCCCTAAGAAGGAGCCATGCCCAGTTCTGCCGGGACCCTCGGCCAGGCCGACAGGAGTGGACGCTGGAGCTGGCCCACACTGGGCATAG  12240
GAGCTCACCAGTGAGGGCAGGAGAGCACATGCCGGGGAGCACCCAGCCTCCTGCTGACCAGAGGCCTGCCCCAGAGCCCAGGAGGCTGCAGAGGCCTCTCCAGGGAGACACTGTGCATGT  12360
CTGGTACCTAAGCAGCCCCCCACGTCCCCAGTCCTGGGGGCCCCTGGCTCAGCTGTCTGGACCCTCCCTGTTCCCTGGGAAGCTCCTCCTGACAGCCCCGCCTCCAGTTCCAGGTGTGGT  12480
TATTGTCAGGCGATGTCAGACTGTGGTGGATATAGTGGCTACGATTACCACAGTGGTGCCGCCCATAGCAGCAACCAGGCCAAGTAGACAGGCCCCTGCTGCGCAGCCCCAGGCATCCAC  12600
TTCACCTGCTTCTCCTGGGGCTCTCAAGGCTGCTGTCTGTCCTCTGGCCCTCTGTGGGGAGGGTTCCCTCAGTGGGAGGTCTGTGCTCCAGGGCAGGGATGATTGAGATAGAAATCAAAG  12720
GCTGGCAGGGAAAGGCAGCTTCCCGCCCTGAGAGGTGCAGGCAGCACCACGGAGCCACGGAGTCACAGAGCCACGGAGCCCCCATTGTGGGCATTTGAGAGTGCTGTGCCCCCGGCAGCC  12840
CAGCCCTGATGGGGAAGCCTGTCCCATCCCACAGCCCGGGTCCCACGGGCAGCGGGCACAGAAGCTGCCAGGTTGTCCTCTATGATCCTCATCCCTCCAGCAGCATCCCCTCCACAGTGG  12960
GGAAACTGAGGCTTGGAGCACCACCCGGCCCCCTGGAAATGAGGCTGTGAGCCCAGACAGTGGGCCCAGAGCACTGTGAGTACCCCGGCAGTACCTGGCTGCAGGGATCAGCCAGAGATG  13080
CCAAACCCTGAGTGACCAGCCTACAGGAGGATCCGGCCCCACCCAGGCCACTCGATTAATGCTCAACCCCCTGCCCTGGAGACCTCTTCCAGTACCACCAGCAGCTCAGCTTCTCAGGGC  13200
CTCATCCCTGCAAGAAGGTCAAGGGCTGGGCCTGCCAGAAACACAGCACCCTCCCTAGCCCTGGCTAAGACAGGGTGGGCAGACGGCTGTGGACGGGACATATTGCTGGGGCATTTCTCA  13320
CTGTCACTTCTGGGTGGTAGCTCTGACAAAAACGCAGACCCTGCCAAAATCCCCACTGCCTCCCGCTAGGCTGGCCTGGAATCCTGCTGTCCTAGGAGGCTGCTGACCTCCAGGATGGCT  13440
CCGTCCCCAGTTCCAGGGCGAGAGCAGATCCCAGGCAGGCTGTAGGCTGGGAGGCCACCCCTGCCCTTGCCGGGGTTGAATGCAGGTGCCCAAGGCAGGAAATGGCATGAGCACAGGGAT  13560
GACCGGGACATGCCCCACCAGAGTGCGCCCCTTCCTGCTCTGCACCCTGCACCCCCCAGGCCAGCCCACGACGTCCAACAACTGGGCCTGGGTGGCAGCCCCACCCAGACAGGACAGACC  13680
CAGCACCCTGAGGAGGTCCTGCCAGGGGGAGCTAAGAGCCATGAAGGAGCAAGATATGGGGCCCCCGATACAGGCACAGATGTCAGCTCCATCCAGGACCACCCAGCCCACACCCTGAGA  13800
GGAACGTCTGTCTCCAGCCTCTGCAGGTCGGGAGGCAGCTGACCCCTGACTTGGACCCCTATTCCAGACACCAGACAGAGGCGCAGGCCCCCCAGAACCAGGGTTGAGGGACGCCCCGTC  13920
AAAGCCAGACAAAACCAAGGGGTGTTGAGCCCAGCAAGGGAAGGCCCCCAAACAGACCAGGAGGTTTCTGAAGGTGTCTGTGTCACAGTGGGGTATAGCAGCAGCTGGTACCACAGTGAC  14040
ACTCACCCAGCCAGAAACCCCATTCCAAGTCAGCGGAAGCAGAGAGAGCAGGGAGGACACGTTTAGGATCTGAGACTGCACCTGACACCCAGGCCAGCAGACGTCTCCCCTCCAGGGCAC  14160
CCCACCCTGTCCTGCATTTCTGCAAGATCAGGGGCGGCCTGAGGGGGGGTCTAGGGTGAGGAGATGGGTCCCCTGTACACCAAGGAGGAGTTAGGCAGGTCCCGAGCACTCTCCCCATTG  14280
AGGCTGACCTGCCCAGAGAGTCCTGGCCCACCCCACACACCGGGGCGGAATGTGTCCAGGCCTCGGTCTCTGTGGGTGTTCCGCTAGCTGGGGCTCACAGTGCTCACCCCACACCTAAA  14400
ATGAGCCACAGCCTCCGGAGCCCCCGCAGAGACCCCGCCCACAAGCCCAGCCCCCACCCAGGAGGCCCCAGAGCTCAGGGCGCCCCGTCGGATTCCGAACAGCCCCGAGTCACAGCGGGT  14520
ATAACCGGAACCACCACTGTCAGAATAGCTACGTCAAAAACTGTCCAGTGGCCACTGCCGGAGCCCCGCCAGAGAGGGCAGCAGCCACTCTGATCCCATGTCCTGCCGGCTCCCATGACC  14640
CCCAGCACGCGGAGCCCACAGTGTCCCCACTGGATGGGAGGACAAGAGCTGGGGATTCCGGCGGGTCGGGGCAGGGGCTTGATCGCATCCTTCTGCCGTGGCTCCAGTGCCCCTGGCTGG  14760
AGTTGACCCTTCTGACAAGTGTCCTCAGAGAGACAGGCATCACCGGCGCCTCCCAACATCAACCCCAGGCAGCACAGGCACAAACCCCACATCCAGAGCCAACTCCAGGAGCAGAGACAC  14880
CCCAATACCCTGGGGGACCCCGACCCTGATGACTTCCCACTGGAATTC                                                                         14928
```

Fig. 2. Nucleotide sequence of a 15-kb DNA fragment containing 12 D_H genes. D_H-coding regions are marked by thick lines. Signal heptamers and nonamers are underlined. Other than in the neighbouring regions of typical D_H-coding regions, 15 CAC(A/T)GTG sequences exist, and are also underlined. The heptamers located upstream of D_M genes are sandwiched by signal nonamers. The deleted (or inserted) DNA portions (I, II, III and IV) are boxed and marked by arrowheads. D_{XP1} and $D_{XP'1}$ were created by local duplication of a D_{XP1}-containing fragment. In a previous paper (Ichihara *et al.*, 1988), we proposed that it had happened by unequal crossing-over in the homologous sequences $\mathrm{CCACAG}^{C}_{*}\mathrm{CCTCCC}^{C}_{T}$. This hypothesis was confirmed by this study since the homologous sequences were also found both sides of D_{XP4}. The consensus sequence of putative sites of unequal crossing-over is $\mathrm{CCACAGCC}^{C}_{*}\mathrm{TCC}^{C}_{T}$ by the comparison of five sequences (hatched under lines). The 16-bp repeating regions are boxed by broken lines. The consensus sequence of the 16-bp repeat is $\mathrm{CCTGG}^{G}_{A}\mathrm{C}^{C}_{T}\mathrm{TCACCTG}^{A}_{G}$.

containing fragments can be identified by cross-hybridization with D_{SP2} probe, sequences of D_H genes belonging to D_{SP2} and D_{FL16} are relatively similar to each other (Kurosawa and Tonegawa, 1982). In humans, nucleotide sequences of D_H genes belonging to the same families are also well conserved. However, those belonging to different families presented in this study very much diverge. Southern hybridization experiments with each D_H-gene-containing probe detected only the D_H genes belonging to the same family as the probe (Figure 3).

Most of the somatic D_H sequences are assigned to one of the germline D_H genes

In mouse, most CDRIII of H chains derived from myelomas or hybridomas sequenced so far can be assigned to one of the 12 already identified germline D_H genes by the homology of nucleotide sequences of D_H-coding regions, although N segments are found at the boundaries between D_H and J_H as well as V_H and D_H (Kurosawa and Tonegawa, 1982). The somatic D_H segments of human Ig show a much higher degree of diversity than those of mouse.

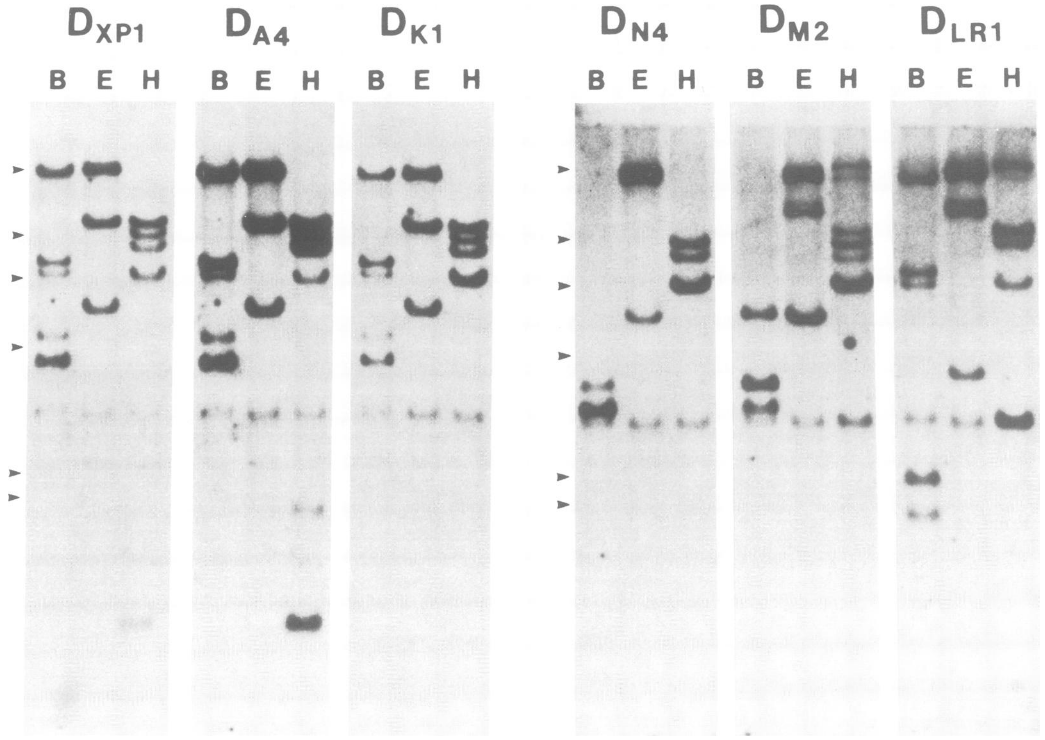


Fig. 3. Southern hybridization of germline DNA with six D_H-gene-containing clones as probes. Human placental DNA was digested with the restriction enzymes *Bam*HI, *Eco*RI and *Hin*dIII respectively. DNA samples (5 μg) were electrophoresed to 0.8% agarose gel. Southern hybridization was performed as described in Materials and methods. The D_{XP1} probe contained both D_{XP1} and $D_{XP'1}$ genes. D_{M2} probe contained not only D_{M2} gene but DIR2 gene. Arrowheads indicate the position of *Hin*dIII-digested λ phage DNA fragments.

```
DXP4    GGTTTGGGG TGAGGTCTGTGT  CACTGTG GTATTACGATTTTTGGAGTGGTTATTATACC CACAGTG TCACAGAGTCCA TCAAAAACC
DXP1    GGTTTAGAA TGAGGTCTGTGT  CACTGTG GTATTACGATATTTTGACTGGTTATTATAAC CACAGTG TCACAGAGTCCA TCAAAAACC
DXP'1   GGTTTGGGG TGAGGTCTGTGT  CACTGTG GTATTACTATGGTTCGGGGAGTTATTATAAC CACAGTG TCACAGAGTCCA TCAAAAACC

DA4     GCTTTTTGT GAAGGGTCCTCC  TACTGTG TGACTACAGTAACTAC                CACAGTG ATGAACCCAGCA GCAAAAACT
DA1     GCTTTTTGT GAAGGGCCCTCC  TGCTGTG TGACTACAGTAACTAC                CATAGTG ATGAACCCAGTG GCAAAAACT

DK4     GGTTATTGT CAGGGGGTGTCA  GACTGTG GTGGATACAG   CTATGGTTAC         CACAGTG GTGCTGCCCATA GCAGCAACC
DK1     GGTTATTGT CAGGCGATGTCA  GACTGTG GTGGATATAGTGGCTACGATTAC         CACAGTG GTGCCGCCCATA GCAGCAACC

DN4     CGTTTCTGA AGGTGTCTGTGT  CACAGTG GAGTATAGCAGC   TCGTCC           CACAGTG ACACTCGCCAGG CCAGAAACC
DN1     GGTTTCTGA AGGTGTCTGTGT  CACAGTG GGGTATAGCAGCAGCTGGTAC           CACAGTG ACACTCACCCAG CCAGAAACC

DM1     GGATTCTGA ACAGCCCCGAGT  CACAGTG GGTATAACTGGAACTAC               CACTGTG AGAAAAGCTTCG TCCAAAACG
DM2     GGATTCCGA ACAGCCCCGAGT  CACAGCG GGTATAACCGGAACCAC               CACTGTC AGAATAGCTACG TCAAAAACT

DLR5    GGATTTTGT GGGGGCTTGTGT  CACTGTG AGAATATTGTAATAGTACTACTTTCTATGCC CACAGTG ACACAGCCCCAG TCCCAAAGC
DLR4    GGATTTTGT GGGGGCTCGTGT  CACTGTG AGGATATTGTAGTAGTACCAGCTGCTATGCC CACAGTG ACACAGCCCCAT TCCCAAAGC
DLR1    GGATTTTGT GGGGGCTCGTGT  CACTGTG AGGATATTGTACTAATGGTGTATGCTATACC CACAGTG ACACAGCCCCAT TCCCAAAGC
                                                    GG
DLR2    GGATTTTGT GGGGGCTCGTGT  CACTGTG AGGATATTGTAGTGGTGGTAGCTGCTACTCC CACAGTG ACACAGACCCAT TCCCAAAGC
DLR3    GGATTTTGT GGGGGCTCGTGT  CACTGTG AGCATATTGTGGTGGTGAT   TGCTATTCC CACAGTG ACACAACCCCAT TCCTAAAGC

DHQ52   GGTTTTTGG CTGAGCTGAGAAC CACTGTG CTAACTGGGGA                     CACAGTG ATTGGCAGCTCT ACAAAAACC
```

Fig. 4. Nucleotide sequences of 17 human germline D_H genes. D_H-coding sequences are sandwiched by signal heptamers and nonamers separated by 12 nucleotide spacers. They can be classified into seven families. D_{LR1} to D_{LR4} were published by Siebenlist *et al.* (1981). Our sequence data of D_{LR1} has a discrepancy at two positions compared with their data: AA with GG. D_{LR5} was reported by Zong *et al.* (1988). D_{HQ52} was published by Ravetch *et al.* (1981). Signal heptamers and nonamers are underlined.

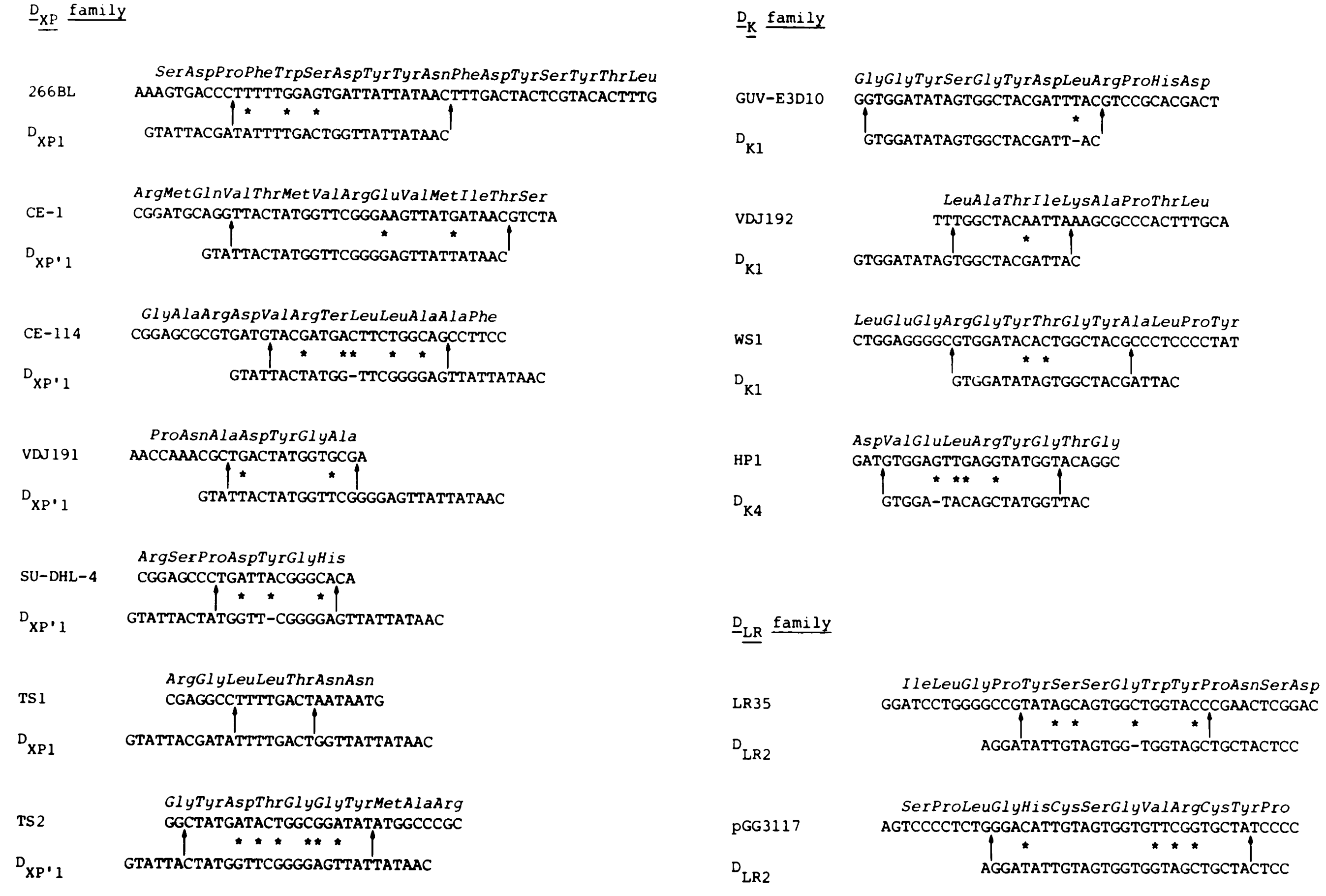

D_XP family
266BL
SerAspProPheTrpSerAspTyrTyrAsnPheAspTyrSerTyrThrLeu
AAAGTGACCCTTTTTGGAGTGATTATTATAACTTTGACTACTCGTACACTTTG
D_XP1
GTATTACGATATTTTGACTGGTTATTATAAC
CE-1
ArgMetGlnValThrMetValArgGluValMetIleThrSer
CGGATGCAGGTTACTATGGTTCGGGAAGTTATGATAACGTCTA
D_XP'1
GTATTACTATGGTTCGGGGAGTTATTATAAC
CE-114
GlyAlaArgAspValArgTerLeuLeuAlaAlaPhe
CGGAGCGCGTGATGTACGATGACTTCTGGCAGCCTTCC
D_XP'1
GTATTACTATGG-TTCGGGGAGTTATTATAAC
VDJ191
ProAsnAlaAspTyrGlyAla
AACCAAACGCTGACTATGGTGCGA
D_XP'1
GTATTACTATGGTTCGGGGAGTTATTATAAC
SU-DHL-4
ArgSerProAspTyrGlyHis
CGGAGCCCTGATTACGGGCACA
D_XP'1
GTATTACTATGGTT-CGGGGAGTTATTATAAC
TS1
ArgGlyLeuLeuThrAsnAsn
CGAGGCCTTTTGACTAATAATG
D_XP1
GTATTACGATATTTTGACTGGTTATTATAAC
TS2
GlyTyrAspThrGlyGlyTyrMetAlaArg
GGCTATGATACTGGCGGATATATGGCCCGC
D_XP'1
GTATTACTATGGTTCGGGGAGTTATTATAAC
D_K family
GUV-E3D10
GlyGlyTyrSerGlyTyrAspLeuArgProHisAsp
GGTGGATATAGTGGCTACGATTTACGTCCGCACGACT
D_K1
GTGGATATAGTGGCTACGATT-AC
VDJ192
LeuAlaThrIleLysAlaProThrLeu
TTTGGCTACAATTAAAGCGCCCACTTTGCA
D_K1
GTGGATATAGTGGCTACGATTAC
WS1
LeuGluGlyArgGlyTyrThrGlyTyrAlaLeuProTyr
CTGGAGGGGCGTGGATACACTGGCTACGCCCTCCCCTAT
D_K1
GTGGATATAGTGGCTACGATTAC
HP1
AspValGluLeuArgTyrGlyThrGly
GATGTGGAGTTGAGGTATGGTACAGGC
D_K4
GTGGA-TACAGCTATGGTTAC
D_LR family
LR35
IleLeuGlyProTyrSerSerGlyTrpTyrProAsnSerAsp
GGATCCTGGGGCCGTATAGCAGTGGCTGGTACCCGAACTCGGAC
D_LR2
AGGATATTGTAGTGG-TGGTAGCTGCTACTCC
pGG3117
SerProLeuGlyHisCysSerGlyValArgCysTyrPro
AGTCCCCTCTGGGACATTGTAGTGGTGTTCGGTGCTATCCCC
D_LR2
AGGATATTGTAGTGGTGGTAGCTGCTACTCC

Fig. 5. Comparison of the nucleotide sequences between germline D_H segments and somatic D_H segments. Fifteen somatic D_H sequences have already been published. In a previous paper (Ichihara *et al.*, 1988), we reported that seven somatic D_H sequences were assigned to one of the germline D_H segments. In this study, six more somatic D_H sequences were also assigned to one of the germline sequences. For the assignment, we tried to find maximum homology instead of perfect match between germline and somatic sequences, since there are five members in each D_H gene family, and the nucleotide sequences of the D_H genes belonging to each family differed slightly among the members. Moreover, somatic mutations may contribute to increased diversity in somatic sequences. The somatic sequence data were referred from the following: 266BL, Kenter *et al.* (1982); CE1 and CE114, Takahashi *et al.* (1984); VDJ191 and VDJ192, Mensink *et al.* (1986); TS1, TS2, WS1 and HP1, Shen *et al.* (1987); GUV-E3D10, Noma *et al.* (1984); LR35, Ravetch *et al.* (1981); pGG3117, Y.Ohshita (personal communication); SU-DHL-4, Cleary *et al.* (1986). Apparent breaking points are shown by vertical arrows. The asterisks indicate the mismatched nucleotides.

In a previous paper (Ichihara *et al.*, 1988), however, we showed that seven of the 11 published somatic D_H sequences could be assigned to one of the identified germline D_H sequences. Moreover, we predicted the presence of two more germline D_H families, based on the sequence similarities between the somatic D_H sequences. In fact, the D_K gene family newly identified in this paper corresponds to one of the predicted D_H genes (Figure 5). After we had submitted a previous paper (Ichihara *et al.*, 1988), four somatic D_H sequences were reported (Shen *et al.*, 1987). All of them were assigned to one of the germline D_H sequences shown in Figure 5. These results indicate that the somatic D_H sequences of Ig H chain are created by essentially the same mechanism in mouse and man, and that a central part of the individual somatic D_H region in $V_H-D_H-J_H$ structure is derived from only one D_H gene.

Identification of a new kind of D_H gene

In a previous paper (Ichihara *et al.*, 1988), we predicted one more germline D_H gene corresponding to the somatic D_H sequences in HIG1 and 333 cells, which are rich in G and C residues. However, the already identified 17 germline D_H genes have no sequence homology to them. Also in mouse, some of the somatic D_H sequences rich in G and C residues are not homologous to any of the 12 germline D_H segments (Kurosawa and Tonegawa, 1982). In the case of these GC-rich somatic D_H sequences, the regions encoded by germline D_H genes were presumably removed by exonuclease activity. This argument is supported by the proposition of the involvement of terminal deoxynucleotidyl-transferase (TdT) in N-region diversification by Alt and Baltimore (1982), in which they emphasized high frequencies of GC-rich sequences in N regions and the preference of TdT for dG residues. How about the D_H sequences in HIG1 and 333? Alternatively, are there more germline D_H gene families rich in G and C residues? We propose a third possibility. As shown in Figure 6(a), upstream of D_M gene, we found a DNA region whose sequence is complementary to somatic D_H gene of HIG1 (18/21 nucleotides) and homologous to that of 333 (16/23 nucleotides). The region surrounding this DNA has several signal heptamers and nonamers. Figure 6(b) schematically shows the location and spacer lengths. Although the distance between the heptamer at the 5′ end and the heptamers located in D_M gene are

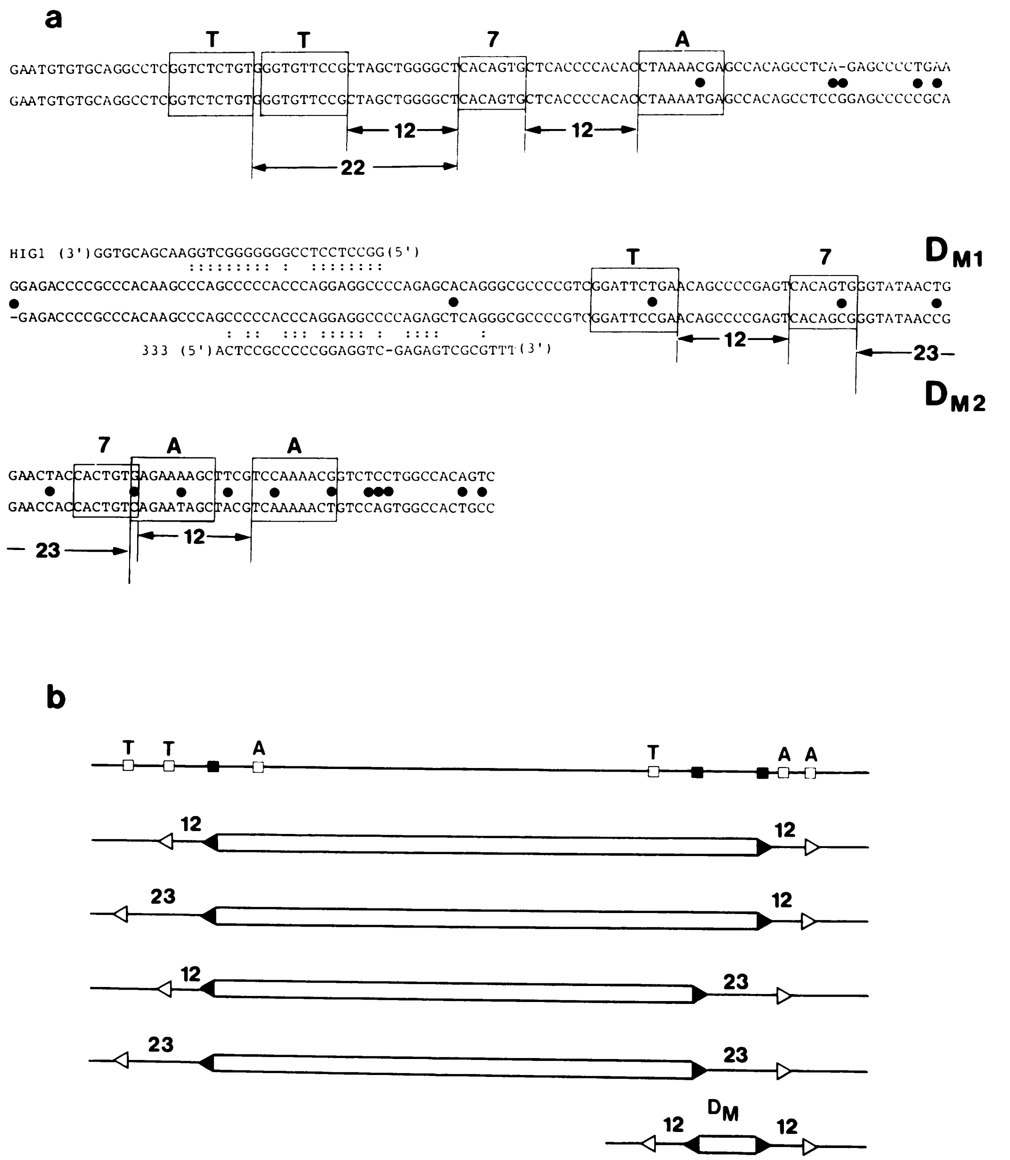

a
T
T
7
A
GAATGTGTGCAGGCCTCGGTCTCTGTGGGTGTTCCGCTAGCTGGGGCTCACAGTGCTCACCCCACACCTAAAACGAGCCACAGCCTCA-GAGCCCCTGAA
GAATGTGTGCAGGCCTCGGTCTCTGTGGGTGTTCCGCTAGCTGGGGCTCACAGTGCTCACCCCACACCTAAAATGAGCCACAGCCTCCGGAGCCCCCGCA
12
12
22
HIG1 (3')GGTGCAGCAAGGTCGGGGGGGCCTCCTCCGG(5')
GGAGACCCCGCCCACAAGCCCAGCCCCCACCCAGGAGGCCCCAGAGCACAGGGCGCCCCGTCGGATTCTGAACAGCCCCGAGTCACAGTGGGTATAACTG
-GAGACCCCGCCCACAAGCCCAGCCCCCACCCAGGAGGCCCCAGAGCTCAGGGCGCCCCGTCGGATTCCGAACAGCCCCGAGTCACAGCGGGTATAACCG
333 (5')ACTCCGCCCCCGGAGGTC-GAGAGTCGCGTTT(3')
T
7
D_{M1}
12
23
D_{M2}
7
A
A
GAACTACCACTGTGAGAAAAGCTTCGTCCAAAACGGTCTCCTGGCCACAGTC
GAACCACCACTGTCAGAATAGCTACGTCAAAAACTGTCCAGTGGCCACTGCC
23
12
b
T
T
A
T
A
A
12
12
23
12
12
23
23
23
D_M
12
12

Fig. 6. Structure of a new kind of D_H gene family (DIR gene). (**a**) There is a CACAGTG sequence upstream of D_M gene. The heptamers are sandwiched by signal nonamer-like sequences. A and T indicate GGTTTTTGT- and ACAAAAACC-like sequences respectively. The upper sequence (DIR1-D_{M1}) is in the first 9-kb repeat and the lower sequence (DIR2-D_{M2}) is in the second repeat. Mismatched nucleotides between two DIR gene regions are dotted. Two somatic D_H sequences (HIG1 and 333) which could not be assigned to any of the germline D_H sequences have sequence homology to these regions. The somatic D_H sequence of HIG1 has complementary sequence and that of 333 shows homology in the same orientation to DIR gene regions. Colons indicate matched or complementary nucleotides between somatic D_H genes and germline DIR genes. The somatic D_H sequence data were referred from the following: HIG1, Kudo *et al.* (1985); 333, Cleary *et al.* (1986). (**b**) Location of signal heptamers and nonamers in the DIR region. The DIR region is sandwiched by signal heptamers and nonamers separated by 12 and 23 nucleotide spacers on both sides. Open squares indicate GGTTTTTGT- and ACAAAAACC-like sequences. Closed squares indicate signal heptamer sequences. Putative D_H genes are shown by open boxes sandwiched by signal heptamers (closed triangles) and nonamers (open triangles). The numbers between heptamers and nonamers indicate spacer length.

rather long (127 or 151 bp), these regions are sandwiched by both 12- and 23-nucleotide spacer signals at both ends. We propose to refer to this region as DIR gene (D_H gene containing irregular spacer signals). DIR region may be associated with putative DIR$-D_H$ or D_H-DIR joining by either deletion or inversion. Interestingly, when we aligned the homologous sequences of somatic D_H gene of HIG1 and DIR gene, the polarities of the two segments were opposite and those of 333 and DIR were the same. Since a putative DIR$-D_H$ joint should have 12- and 23-nucleotide spacer signals at the 5′ side of DIR$-D_H$ joint and a 12-nucleotide spacer signal at the 3′ side, it can be a substrate for forming either V_H-DIR$-D_H-J_H$, V_H-D_H-DIR$-J_H$ or V_H-D_H-DIR$-D_H-J_H$ structure. If this is the case, long N segments might originate from DIR genes.

Enormous diversity in the CDRIII regions of human IgH chains is created by a limited number of germline D_H genes and by somatic mutations

In the mouse, there are three D_H genes: D_{SP2}, D_{FL16} and D_{Q52} (Kurosawa and Tonegawa, 1982). The largest family, D_{SP2}, contains the TACTA$\underline{T}$GGT sequence in their central regions. The most frequently used D_H gene, $D_{FL16.1}$, contains the TACTA$\underline{C}$GGT sequence. The D_{SP2} family can encode Tyr-Tyr-Gly, Thr-Met and Leu-Trp, and $D_{FL16.1}$ can encode Tyr-Tyr-Gly, Thr-Thr and Leu-Arg. However, the majority of the somatic D_H sequences contain Tyr-Tyr-Gly (Kabat *et al.*, 1987). It suggests that one of the three coding frames is predominantly used in the mouse. On the other hand, all of the three coding frames are equally used in the human, as shown in Figure 5. Other than the nucleotide sequences summarized in Figure 5, amino acid sequences of many human somatic D_H regions have been reported (Kabat *et al.*, 1987). They completely diverged except for two combinations, POM and LAY, MCE and NZU, as shown in Figure 7. The amino acid sequences of POM and LAY are identical. MCE and NZU contain the same sequence RPPWRFT. The other sequences do not have sequence homology if they are longer than three amino acids. We compared the amino acid sequences of the somatic D_H with the germline D_H deduced from the nucleotide sequences in Figure 4. We assumed that all of the three coding frames could be read, and searched for the examples matching three amino acids in four, or matching more than four amino acids. The reason why we adopted these criteria for assignment of somatic D_H to germline D_H is that a three-amino-acid match is hardly due to accidental coincidence. Although the apparent amino acid sequences are so different among them, 19 somatic D_H sequences were assigned to one of the germline D_H genes as indicated in Figure 7. Presence of the same sequences in different somatic D_H sequences observed in MCE and NZU indicates that they are not the products of insertion of random nucleotides by TdT, but that they are encoded in the germline sequences. The predicted nucleotide sequences encoding RPPWR are rich in GC residues. They might be encoded in DIR regions. Since all of the 15 published somatic D_H sequences have been assigned to either germline D_H genes or DIR genes at DNA sequence level (Figure 5), there might not remain many D_H genes other than the seven D_H gene families described in this paper. The somatic D_H sequences which could not be assigned to the already determined germline D_H genes in Figure 7 might originate from other D_H genes belonging to the six D_H gene families, and/or the diversification of the sequences could be amplified by somatic mutations. It is reasonable that the enormous diversity in the CDRIII regions of human IgH chains is created by a limited number of D_H genes and somatic mutations.

Materials and methods

Clones

The D_H-gene-containing germline clone (HUD-3) was previously described (Ichihara *et al.*, 1988). All the probes used in this study were subcloned into Bluescript KS plus vector (Stratagene, San Diego, CA). The restriction enzyme sites of germline D_H probes are described below. The number in parentheses indicates the nucleotide position of each enzyme site. The asterisks indicate the enzyme sites which were broken to construct the probe. The poly-linker sites of the vector (restriction enzyme sites without parentheses) can be used to isolate D_H gene inserts: D_{XP1} (containing $D_{XP'1}$) probe, *Bam*HI–*Bal*I*(10232)–*Pst*I(10749); D_{A4} probe, *Bam*HI–*Eco*RV*(1561)–*Acc*I(2287)–*Xho*I; D_{K1} probe, *Sac*I(12241)–*Sma*I(12875); D_{N4} probe, *Bam*HI(4260)–*Nco*I*(4901)–*Sal*I; D_{M2} (containing DIR2) probe, *Xba*I–*Stu*I*(14339)–*Eco*RI(14923); and D_{LR1}, *Xba*I–*Nco*I(7577)–*Hin*dIII(8174).

Southern blotting

Human placental DNA was extracted by the method of Gross-Bellard *et al.* (1973). The samples of extracted high molecular DNA were digested with the restriction enzymes *Bam*HI, *Eco*RI and *Hin*dIII respectively. DNA samples (5 μg) were electrophoresed to 0.8% agarose gel. The gels were treated sequentially with (i) 0.25 M HCl for 15 min, (ii) 0.5 M NaOH, 1.5 M NaCl for 30 min, and (iii) 1.0 M Tris–HCl (pH 7.5) buffer containing 1.5 M NaCl for 30 min. The DNA samples were transferred to nylon membrane (Hybond N, Amersham) by the method of Southern (1975) with modification (Olszewska and Jones, 1988) using LKB 2016 VacuGene vacuum blotting system (Pharmacia LKB Biotechnology AB, Bromma, Sweden). The transfer solution used was 20 × SSC. The DNA samples were fixed onto the membrane by UV irradiation for 5 min following baking (80°C, 2 h).

MCE	RPPWRFTGNLGG	WAS	FRQPFVQF
NZU	RPPWRFTSDLGSFSP	TEI	VTPAAASLTFSA VVPAA (D_{LR4})
POM,LAY	DAGPYVSPTF	BRO	SPVSLVDGWL
EU	GYGIYSPEEYN YYYGSGSYYN ($D_{XP'1}$)	GRA	HIYVTL HIVV (D_{LR3})
SIE	EWKGQVNVNP	ZAP	TRPGGYFS GYSS (D_{N1})
WOL	EYGFDTSD	JON	VVVSTS VVV ($D_{LR2,3}$)
NDCL	SDPFWSDYYNFDYSYTL FWSGYY (D_{XP4})	BUT	DLAAARLFGK AAR (D_{N4}) AAA (D_{N1})
MOT	GAHYSDTDDSGTSLGP SGGS (D_{LR2})	DOB	GYIWNG YNWN (D_{M1})
HUS	BRBBYGBF	WEA	GWLLN WLL ($D_{XP4,1}$)
COR	ITVIPAPAGYMDV IVVVPA (D_{LR4})	NIE	IRDTAMF DTAM (D_{K4})
DAW	SCGSQ	CAM	DRPLYGBYRA RILY (D_{LR1})
OU	VVNSVMAG	GAL	GWGGG
HE	RHPRTL	TRO	TNNFNWSTFSL
NEWM	NLIAGCI	KOL	DGGHGFCSSASCFGP GYCSSTSC (D_{LR4})
WAH	GNPPPYYDIGTGSDD YYDILTG (D_{XP1})	HIL	DPDILTAFS DILT (D_{XP1})
TUR	LSVTAV	BUR	LIAVAGTR IAAAGT (D_{N1})
TIL	GKVSAYY SGYY (D_{XP4})	GA	SGIALGSVAGT GIA (D_{N1})

Fig. 7. Assignment of the somatic D_H sequences to the germline D_H genes at amino acid sequence level. Amino acid sequences of human somatic D_H regions have been reported (Kabat *et al.*, 1987). The amino acid sequences of the somatic D_H were compared with the germline D_H deduced from the nucleotide sequences. The germline D_H genes corresponding to each CDRIII region are in parentheses. The solid lines between CDRIII regions and germline D_H genes indicate the coincident amino acids.

Filter hybridization

The membranes were immersed in boiled washing solution (0.1 × SSC, 0.1% SDS) for 15 min and were subjected to prehybridization. The condition of prehybridization was 5 × SSPE [20 × SSPE: 0.2 M NaH_2PO_4 buffer (pH 7.4) containing 3.6 M NaCl, 20 mM EDTA], 0.1 mg/ml heat-denatured salmon sperm DNA, 50% formamide (Merck, Darmstadt, FRG), 5% Irish cream liqueur (Baileys Co. Ltd, Dublin, Ireland) and 0.1% SDS at 42°C overnight. Conditions of hybridization were the same as that of prehybridization except for the addition of heat-denatured ^{32}P-labelled probe by random oligonucleotide labelling (Feinberg and Vogelstein, 1983) using multiprimer labelling kit (Amersham). The hybridization was performed at 42°C overnight. Washing conditions were at 42°C for 30 min in 4 × SSC following 0.5 × SSC twice. Autoradiography was performed for 2 days at −80°C.

Nucleotide sequence determination and sequence analysis

The germline DNA fragments were subcloned into Bluescript KS vector. The unidirectionally deleted insert-containing clones were obtained using exonuclease III and Mung Bean nuclease (Henikoff, 1984; Yanisch-Perron *et al.*, 1985). DNA sequencing was performed by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using deoxy-7-deazaguanosine triphosphate in place of dGTP (Mizusawa *et al.*, 1986). The sequencing primers used were KS, SK, T3 and T7 primers purchased from Stratagene. Sequences were analysed with HIBIO DNASIS software (Hitachi, Japan).

Acknowledgements

We thank Dr S.Ohno for a valuable suggestion in the preparation of the manuscript; Dr Y.Ohshita for communicating an unpublished somatic D_H sequence; and Drs Y.Takagi, I.Ishiguro and K.Fujita for their encouragement. We are grateful to Y.Sugawara, M.Ueda and Y.Maeda for technical assistance, and to A.Nagata for excellent preparation of the manuscript. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan, Fujita-Gakuen Health University, Japan Private School Promotion Foundation (to Y.K.) and the Uehara Memorial Foundation (to Y.I.).

References

Akira,S., Okazaki,K. and Sakano,H. (1987) *Science*, **238**, 1134–1138.

Alt,F.W. and Baltimore,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4118–4122.

Alt,F.W., Yancopoulos,G.D., Blackwell,T.K., Wood,C., Thomas,E., Boss,M., Coffman,R., Rosenberg,N., Tonegawa,S. and Baltimore,D. (1984) *EMBO J.*, **3**, 1209–1219.

Buluwela,L., Albertson,D.G., Sherrington,P., Rabbitts,P.H., Spurr,N. and Rabbitts,T.H. (1988) *EMBO J.*, **7**, 2003–2010.

Cleary,M.L., Smith,S.D. and Sklar,J. (1986) *Cell*, **47**, 19–28.

Early,P., Huang,H., Davis,M., Calame,K. and Hood,L. (1980) *Cell*, **19**, 981–992.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Gross-Bellard,M., Oudet,P. and Chambon,P. (1973) *Eur. J. Biochem.*, **36**, 32–38.

Henikoff,S. (1984) *Gene*, **28**, 351–359.

Ichihara,Y., Abe,M., Yasui,H., Matsuoka,H. and Kurosawa,Y. (1988) *Eur. J. Immunol.*, **18**, 649–652.

Kabat,E.A., Wu,T.T., Reid-Miller,M., Perry,H.M. and Gottesman,K.S. (1987) *Sequences of Proteins of Immunological Interest.* US Dept Health and Human Services, Washington, DC.

Kenter,J.H., Molgaard,H.V., Houghton,M., Derbyshire,R.B., Viney,J., Bell,L.O. and Gould,H.J. (1982) *Proc. Natl. Acad. Sci. USA*, **74**, 6661–6665.

Kudo,A., Ishihara,T., Nishimura,Y. and Watanabe,T. (1985) *Gene*, **33**, 181–189.

Kurosawa,Y. and Tonegawa,S. (1982) *J. Exp. Med.*, **155**, 201–218.

Matsuda,F., Lee,K.H., Nakai,S., Sato,T., Kodaira,M., Zong,S.Q., Ohno,H., Fukuhara,S. and Honjo,T. (1988) *EMBO J.*, **7**, 1047–1051.

Mensink,E.J.B.M., Schuurman,R.K.B., Schot,J.D.L., Thompson,A. and Alt,F.W. (1986) *Eur. J. Immunol.*, **16**, 963–967.

Mizusawa,S., Nishimura,S. and Seela,F. (1986) *Nucleic Acids Res.*, **14**, 1319–1324.

Noma,Y., Yaoita,Y., Matsunami,N., Rosen,A., Klein,G. and Honjo,T. (1984) *Mol. Biol. Med.*, **2**, 337–350.

Olszewska,E. and Jones,K. (1988) *Trends Genet.*, **4**, 92–94.

Ravetch,J.V., Siebenlist,U., Korsmeyer,S., Waldmann,T. and Leder,P. (1981) *Cell*, **27**, 583–591.

Sakano,H., Huppi,K., Heinrich,G. and Tonegawa,S. (1979) *Nature*, **280**, 288–294.

Sakano,H., Kurosawa,Y., Weigert,M. and Tonegawa,S. (1981) *Nature*, **290**, 562–565.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schilling,J., Clevinger,B., Davie,J.M. and Hood,L. (1980) *Nature*, **283**, 35–40.

Shen,A., Humpries,C., Tucker,P. and Blatter,F. (1987) *Proc. Natl. Acad. Sci. USA*, **86**, 8563–8567.

Siebenlist,U., Ravetch,J.V., Korsmeyer,S., Waldmann,T. and Leder,P. (1981) *Nature*, **294**, 631–635.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.

Takahashi,N., Noma,T. and Honjo,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5194–5198.

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

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

Zong,S.Q., Nakai,S., Matsuda,F., Lee,K.H. and Honjo,T. (1988) *Immunol. Lett.*, **17**, 329–334.

Received on September 8, 1988